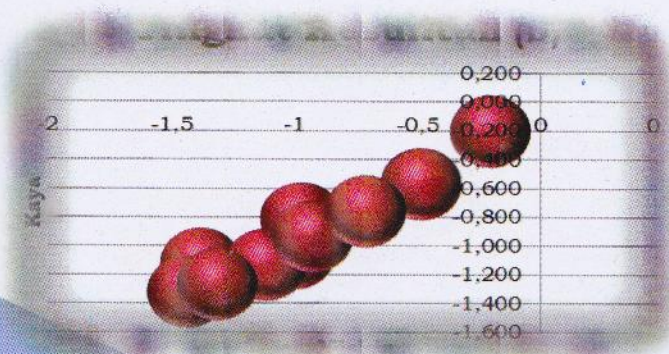


Teori Respons Butir (IRT) Dikotomus Dengan

Dr. Muhammad Ali Gunawan, S.Pd., M.Pd.
Prof. Dr. Heri Retnawati, S.Pd., M.Pd.
Dr. Ari Setiawan, S.Sos.I., M.Pd.
Amalia Fitri, M.Pd.



Teori Respons Butir (IRT) Dikotomus Dengan



**Dr. Muhammad Ali Gunawan, S.Pd.,M.Pd.
Prof. Dr. Heri Retnawati, S.Pd.,M.Pd.
Dr. Ari Setiawan, S.Sos.I.,M.Pd.
Amalia Fitri, M.Pd.**

Teori Respons Butir (IRT) Dikotomus Dengan R

Nuta Media, Yogyakarta
Ukuran. 15,5 x 23
Halaman: 157 + x

Cetakan Februari 2023

ISBN : 978-623-8126-28-6 (EPUB)

Penulis : Dr. Muhammad Ali Gunawan, S.Pd.,M.Pd.
Prof. Dr. Heri Retnawati, S.Pd.,M.Pd.,
Dr. Ari Setiawan, S.Sos.I.,M.Pd.
Amalia Fitri, M.Pd

Editor : Dedek Andrian
Sampul : team nuta
Layout : team nuta

Diterbitkan oleh :

NUTA MEDIA

Anggota IKAPI: No. 135/DIY/2021

Jl. Nyi Wiji Adhisoro, Prenggan Kotagede Yogyakarta
nutamediajogja@gmail.com; 081228153789

© 2023, Hak Cipta dilindungi undang-undang, dilarang keras menterjemahkan, memfotokopi atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari penerbit

dicetak oleh : Nuta Media

PENGANTAR PENULIS

Puji syukur kehadirat Alloh SWT. Atas berkat rahmat dan hidayah-Nya buku “IRT Dikotomus Dengan R ” ini bisa hadir di hadapan pembaca. Buku sederhana ini ditulis guna mendukung pelaksanaan program/kebijakan Pemerintah yang ditetapkan melalui regulasi tentang Penilaian hasil belajar oleh pendidik dan satuan pendidikan pada pendidikan dasar dan menengah.

Kebijakan tersebut akan sangat sulit bisa dilaksanakan oleh guru maupun pihak sekolah, mengingat penyusunan dan pengembangan tes hasil belajar (kognitif, afektif dan psikomotor) memiliki asumsi dan teori yang ketat sehingga skor hasil tes dapat diinterpretasikan sama di setiap sekolah, kelas/tingkat. Selain itu, informasi yang dihasilkan melalui tes harus mampu menjadi dasar pengambilan keputusan atau merancang program perbaikan kegiatan belajar mengajar, baik guru dan siswa serta seluruh civitas akademika yang ada, agar terjadi peningkatan/perbaikan mutu pendidikan secara maksimal dan professional serta adil bagi peserta didik.

Sampai saat ini, hampir semua sekolah/guru masih menggunakan teori tes klasik yang notabenehnya sudah tidak lagi memadai untuk menilai hasil belajar siswa dengan segala keterbatasannya. Pengembangan bank soal juga membutuhkan analisis statistik dengan aplikasi/program yang tersedia begitu banyak, seperti BILOG-MG, Parscale, Multilog, Winsteps dan lainnya. Namun, aplikasi tersebut tidak bisa diakses oleh guru atau mahasiswa calon guru karena lisensinya berbayar dan mahal. Hal ini menjadi alasan kuat bagi penulis untuk menghadirkan buku yang membahas analisis instrument penelitian/penilaian dengan program/aplikasi yang *open source* (gratis) yaitu Program R.

Tidak ada gading yang tak retak, begitu juga dengan buku ini, masih banyak kekurangan dan kelemahan yang oleh karenanya kritik dan saran yang konstruktif sangat diharapkan. Semoga bermanfaat dan mampu menjadi salah satu solusi perbaikan mutu/kualitas pendidikan di tanah air.

Pekalongan, 10 November 2022

Penulis

DAFTAR ISI

Pengantar Penulis.....	i
Daftar Isi.....	ii
Daftar Tabel.....	iii
Daftar Gambar.....	iv
BAB 1: Teori Tes Modern (<i>Item Respons Theory</i>).....	1
A. Teori Respons Butir (Item Respons Theory).....	1
B. Asumsi Teori Respons Butir (IRT).....	4
1. Unidimensi	4
2. Independensi Lokal.....	5
3. Invariansi	5
C. Model Teori Respons Butir.....	5
D. Estimasi Parameter Butir dan Orang (Sifat Laten)....	7
E. Menguji Kecocokan Model.....	8
F. Jumlah Sampel dan Panjang Tes.....	9
BAB 2 : Program R (Pengantar).....	13
A. Sejarah Singkat R.....	13
B. Definisi dan Konsep R.....	14
C. Menginstal R.....	15
D. Dasar-dasar R.....	23
E. Simulasi (Membangkitkan Data) Dengan R.....	31
F. Paket Perintah (<i>Packages</i>) R.....	35
BAB 3. Uji Asumsi IRT dengan R.....	36
A. Unidimensi	36
B. Independensi Lokal.....	43
C. Invariansi	43
BAB 4. Analisis Karakteristik Butir Dikotomus.....	45
A. Definisi dan Konsep.....	45
B. Model Rasch dan IRT 1 Parameter Logistik.....	45
C. Model 2 Parameter Logistik (2PL).....	52
D. Model 3 Parameter Logistik (3PL).....	61
E. Model 4 Parameter Logistik (4PL).....	66
BAB 5. Analisis Butir Tes (Skor Dikotomus).....	72
A. Analisis Butir Tes Model Rasch.....	72
B. Analisis Butir Tes Model 1PL.....	94
C. Analisis Butir Tes Model 2PL.....	99
D. Analisis Butir Tes Model 3PL.....	115

E. Analisis Butir Tes Model 4PL.....	126
DAFTAR PUSTAKA.....	134
LAMPIRAN.....	139
INDEKS.....	147
GLOSARIUM.....	148

DAFTAR TABEL

Tabel 1.	Tokoh ilmu pengukuran dan karyanya.....	3
Tabel 2.	Ikhtisar Kemungkinan pembedaan untuk ukuran sampel.....	11
Tabel 3.	Simbol operasi matematika di R	24
Tabel 4.	Fungsi Matematika di R	26
Tabel 5.	Rentang nilai daya beda butir tes.....	52
Tabel 6.	Hasil analisis estimasi parameter butir model 4PL.....	67
Tabel 7.	Hasil analisis estimasi kemampuan orang (peserta tes).....	67
Tabel 8.	Output hasil analisis kecocokan butir dengan model.....	69
Tabel 9.	Output hasil kecocokan orang (peserta tes) dengan model.....	70

DAFTAR GAMBAR

Gambar 1.	Tampilan situs R.....	16
Gambar 2.	Kotak dialog instalasi R.....	16
Gambar 3.	Kotak dialog Setup_Informasi Instalasi R...	17
Gambar 4.	Kotak dialog Setup_pemilihan lokasi tujuan instalasi R.....	17
Gambar 5.	Kotak dialog Setup_pemilihan komponen instalasi R.....	18
Gambar 6.	Kotak dialog Setup_opsi instalasi R.....	18
Gambar 7.	Kotak dialog Setup_pemilihan folder instalasi R.....	19
Gambar 8.	Kotak dialog proses instalasi R.....	19
Gambar 9.	Layar Console ketika R pertama kali di buka.....	20
Gambar 10.	Situs Rstudio download.....	21
Gambar 11.	Penjelasan layar Rstudio.....	21
Gambar 12.	ICC butir tes dengan data simulasi ($a = 1$; $b = 0$).....	35
Gambar 13.	Tampilan data frame pada layar Console.....	37
Gambar 14.	Grafik dimensi data menggunakan PCA.....	41
Gambar 15.	Scree plot.....	42
Gambar 16.	Scatter plot invariansisi.....	43
Gambar 17.	Hubungan antara tingkat kesulitan (b_i) dan tingkat kemampuan (θ) pada model Rasch..	48
Gambar 18.	Kurva Karakteristik Butir Model 2PL.....	56
Gambar 19.	Kurva fungsi informasi daya beda butir 7 dan butir 11.....	58
Gambar 20.	ICC model 2PL ($a=1,876$ dan $b = 0,275$).....	58
Gambar 21.	Butir tes dengan $b = 0$ dan $a = -0,75$	59
Gambar 22.	Kurva Karakteristik Butir Model 3PL.....	62
Gambar 23.	Item Characteristic Curves (ICC) IRT model 3 PL.....	62
Gambar 24.	Perilaku menebak di butir 7.....	63
Gambar 25.	Kurva fungsi informasi.....	64
Gambar 26.	Kurva empiris item nomor 20.....	65
Gambar 27.	Kurva karakteristik butir nomor 1.....	68
Gambar 28.	Trace line butir soal atau ICC secara	

	keseluruhan.....	68
Gambar 29.	Kurva informasi tes.....	69
Gambar 30.	Trace line butir soal atau ICC secara keseluruhan.....	71
Gambar 31.	Layar script editor pada Rstudio.....	72
Gambar 32.	Grafik ICC berbasis model.....	76
Gambar 33.	Grafik ICC berbasis estimasi empiris.....	76
Gambar 34.	Grafik parameter kemampuan orang (peserta tes).....	78
Gambar 35.	Sebaran kemampuan peserta tes pada semua butir soal.....	79
Gambar 36.	Grafik analisis kecocokan model.....	81
Gambar 37.	Informasi Tes dan Informasi Butir.....	84
Gambar 38.	ICC model Rasch butir nomor 1.....	88
Gambar 39.	ICC model Rasch seluruh butir.....	88
Gambar 40.	TCC model Rasch.....	89
Gambar 41.	Kurva Informasi Tes model Rasch.....	89
Gambar 42.	Kurva Informasi Butir model Rasch pada butir nomor 1.....	90
Gambar 43.	ICC Empiris model Rasch butir nomor 3.....	92
Gambar 44.	ICC model IRT 1PL.....	95
Gambar 45.	IIC dan TIF model IRT 1PL.....	96
Gambar 46.	ICC model IRT 2PL.....	102
Gambar 47.	ICC model IRT 2PL butir nomor 1 dan 2.....	102
Gambar 48.	IIC model IRF model IRT 2PL.....	103
Gambar 49.	ICC model IRT 2PL butir nomor 3 dengan paket “mirt”.....	110
Gambar 50.	ICC model IRT 2PL dengan paket “mirt”.....	111
Gambar 51.	ETS dan Tes Information model IRT 2PL dengan paket “mirt”.....	112
Gambar 52.	ICC empirical model IRT 2PL dengan paket “mirt”.....	113
Gambar 53.	ICC model IRT 3PL dengan paket “lrm”.....	116
Gambar 54.	IIC model IRT 3PL dengan paket “lrm”.....	117
Gambar 55.	ICC model IRT 3PL butir nomor 3 dengan paket “mirt”.....	121
Gambar 56.	ICC model IRT 3PL dengan paket “mirt”.....	122
Gambar 57.	ICC Total model IRT 3PL dengan paket “mirt”.....	122
Gambar 58.	Kurva Informasi Tes model IRT 3PL dengan paket “mirt”.....	123

Gambar 59.	ICC Empiris model IRT 3PL butir nomor 15 dengan paket “mirt”	124
Gambar 60.	ICC model IRT 4PL butir nomor 1 dengan paket “mirt”	129
Gambar 61.	ICC model IRT 4PL dengan paket “mirt”	130
Gambar 62.	TCC model IRT 4PL dengan paket “mirt”	130
Gambar 63.	ICC Empiris model IRT 4PL butir nomor 15 dengan paket “mirt”	132

1

Teori Tes Modern (Item Respons Theory)

“Sebagus-bagus manusia adalah yang menerima karunia panjang umur dan sepanjang umurnya ia selalu mengerjakan kebaikan”

-TGKH.M. Zaenuddin Abdul Madjid-

A. Teori Respons Butir (*Item Response Theory*)

Seperti yang Anda baca di banyak buku referensi tentang teori tes, bahwa teori tes klasik memiliki kelemahan-kelemahan, untuk itu para ahli pengukuran berusaha untuk mencari alternatif. Model yang diharapkan memiliki karakteristik: (1) karakteristik butir tidak tergantung kepada kelompok peserta tes yang dikenai butir soal tersebut, (2) skor yang menyatakan kemampuan peserta tes tidak tergantung pada tes, (3) model dinyatakan dalam tingkatan (level) butir soal, tidak dalam tingkatan tes, (4) model tingkat tidak memerlukan tes paralel untuk menghitung koefisien reliabilitas, dan (5) model menyediakan ukuran yang tepat untuk setiap skor kemampuan (Hambleton, Swaminathan, & Rogers, 1991).

Model alternatif yang mendekati ciri-ciri itu adalah model pengukuran yang disebut teori respons butir (*Item Response Theory*) atau biasa juga dikenal dengan *Latent Trait Theories* (Weiss, 1983). Karakteristik ini dimungkinkan karena model IRT menyediakan invariansi parameter item dan kemampuan untuk item tes dan orang (de Ayala, 2022), ketika model IRT yang dikaji benar-benar sesuai dengan data tes yang tersedia. Dengan kata lain, item yang sama yang digunakan dalam sampel yang berbeda akan mempertahankan sifat statistiknya (misalnya, kesulitan dan diskriminasi), dan skor orang yang mewakili kemampuan atau sifat laten pada konstruksi tertentu tidak akan bergantung pada item tes tertentu yang diberikan (Zanon, et.al., 2016)

Item response theory (IRT) dengan demikian merupakan paradigma untuk menyelidiki/mengkaji hubungan antara respons individu terhadap satu item tes dan kinerjanya pada ukuran keseluruhan kemampuan atau sifat yang ingin diukur oleh item

tersebut (Embretson & Reise, 2000). Banyak model yang ada di bidang IRT untuk mengevaluasi seberapa baik suatu item menangkap sifat laten yang mendasarinya, tetapi beberapa model IRT yang paling populer adalah model IRT logistic (Wood, 2017). Sebagaimana postulat dasar dari teori respons butir, yaitu: (1) hasil kerja seorang peserta tes pada suatu butir soal dapat diprediksikan (atau dijelaskan) dari suatu jenis faktor-faktor yang disebut sifat-sifat, sifat-sifat laten, atau kemampuan; (2) hubungan antara hasil kerja peserta tes pada suatu butir tes dengan sifat-sifat yang mendasarinya dapat dideskripsikan oleh fungsi yang meningkat yang bersifat *monotonic* yang disebut dengan fungsi karakteristik butir (*item characteristic function* atau *item characteristic curve-ICC*) (de Gruijter & van der Kamp, 2008). Fungsi ini menjelaskan bahwa peluang menjawab benar akan meningkat, jika kemampuan peserta tes meningkat (Hambelton, et.al, 1991).

Para ahli psikometrika mengasumsikan bahwa seperangkat butir tes hanya mengukur satu kemampuan laten atau laten trait, yang dinamakan dengan prinsip unidimensi dan respons terhadap setiap butir tidak bergantung pada respons butir yang lainnya (*local independent*) (de Gruijter & van der Kamp, 2008). Teori respons butir (IRT) dikembangkan berdasarkan karya fundamental Lord, Novick dan Birnbaum (1968, dalam Reise & Revicki, 2015) dan banyak dimanfaatkan pada semua bidang keilmuan terutama pendidikan, psikologis, sosial, ekonomi dan kesehatan.

IRT menggunakan model probabilitas untuk mengkarakterisasi interaksi antara peserta tes dan butir tes. Hal ini mengasumsikan bahwa kinerja peserta tes pada tes dapat diprediksi dan/atau dijelaskan oleh sifat laten (atau kemampuan) yaitu, karakteristik yang menjelaskan kinerja peserta tes pada butir soal. Secara khusus, setelah nilai-nilai sifat laten yang diberikan dan diperhitungkan, jawaban butir secara statistik independen. Inilah yang disebut dengan lokal independensi butir atau syarat independensi butir. Asumsi yang ketiga adalah asumsi invariansi parameter menyatakan bahwa karakteristik butir soal tidak berubah meskipun subjek yang menjawab butir tersebut berbeda-beda termasuk level kemampuannya.

Postulat sebagaimana disebutkan oleh Hambelton, et.al (1991) di atas, dapat diperinci ke dalam beberapa asumsi-asumsi yang mendasari teori respons butir, yaitu: (a) Unidimensi, (b) Independensi lokal, dan (c) Fungsi karakteristik butir menyatakan hubungan yang sebenarnya antara variabel yang tak terobservasi

(yaitu kemampuan) dengan variabel terobservasi (yaitu respons butir) (Hambleton, et.al, 1991). Asumsi unidimensi menyatakan bahwa hanya satu kemampuan yang diukur oleh sekumpulan butir-butir soal dalam suatu tes. Asumsi ini pada praktiknya sukar dipenuhi, sebab terdapat banyak faktor yang dapat memengaruhi hasil suatu tes. Faktor-faktor tersebut antara lain tingkat motivasi, kecemasan, kecepatan, dan keterampilan kognitif lain di luar kemampuan yang diukur oleh sekumpulan butir soal dalam suatu tes. Unidimensi dalam hal ini adalah adanya faktor-faktor dominan yang memengaruhi hasil suatu tes. Faktor-faktor dominan itulah yang disebut kemampuan yang diukur oleh suatu tes.

Sejarah perkembangan teori respons butir (IRT) dapat dirunut dari karya Lord pada model normal ogive untuk satu dan dua parameter, perkembangan selanjutnya dapat dilihat pada tabel 1 berikut.

Tabel 1. Tokoh ilmu pengukuran dan karyanya.

Tahun	Karya	Tokoh yang berperan
1952	One and two parameter normal-ogive models	Lord.
1957-1958	Two and three parameter logistic models	Birnbaum
1960	One parameter logistic model	Rasch
1961	Rasch rating scale model	Rasch
1967	Normal ogive multidimensional model	McDonald
1969	Two parameter normal ogive and logistic graded response model	Samejima
1969	Multidimensional model	Samejima
1972	Continuous (free) response model	Samejima
1972	Nominal response model	Bock
1973	Rasch rating scale model	Andersen
1976	Linear logistic Rasch model	Fischer
1978	Rasch rating scale model	Andrich
1980	Multi component response models	Embrestson
1981	Four parameter logistic model	Barton & Lord
1982	Partial Credit Model	Masters

1985	Linear logistic multidimensional model	Reckase
1988	Unfolding model	Andrich
1990	Sequential step model	Tutz
1991	Non parametric models	Ramsey
1992	Generalized partial credit model	Muraki
1992	Full information item bifactor model	Gibbons & Hedeker
1993	Steps model	Verhelst & Glas

Hambleton, van der Linden, & Wells (2010)

B. Asumsi Teori Respons Butir (IRT)

Teori respons butir didasarkan pada asumsi-asumsi: (a) Unidimensi, (b) independensi lokal, dan (c) fungsi karakteristik butir menyatakan hubungan yang sebenarnya antara variabel laten (kemampuan) dengan variabel manifes (terobservasi) (Hambleton, et.al, 1991).

1. Unidimensi

Setiap butir tes hanya mengukur satu kemampuan, misalnya pada tes UAS mata pelajaran matematika, butir-butir yang termuat di dalamnya hanya mengukur aspek kemampuan matematika siswa, bukan mengukur aspek yang lainnya (seperti kemampuan bahasa dan lainnya). Pada praktiknya, asumsi unidimensi tidak dapat dipenuhi secara ketat karena adanya faktor-faktor kognitif, kepribadian dan faktor-faktor pelaksanaan tes, seperti kecemasan, kelelahan, motivasi, dan tendensi untuk menebak. Hampir secara universal disepakati bahwa data respons item jarang benar-benar unidimensional (Reise & Revicki, 2015).

Oleh karena itu, asumsi unidimensi dapat ditunjukkan hanya jika tes mengandung satu komponen yang dominan saja, yang mengukur kemampuan subjek (sifat laten). Sebagaimana diungkap Reise & Revicki (2015) bahwa masalah kritis dalam aplikasi IRT bukanlah apakah datanya “cukup unidimensional”, melainkan sejauh mana multidimensi berdampak atau mendistorsi estimasi parameter butir.

2. Independensi Lokal

Menurut Kolen & Brennan (2004) tes untuk memenuhi asumsi independensi lokal dapat dilakukan dengan membuktikan peluang dari pola jawaban setiap peserta tes sama dengan hasil kali peluang jawaban benar dan salah peserta tes untuk setiap butir soal. Dengan kata lain, asumsi independensi lokal menyatakan bahwa tidak ada korelasi antara respons peserta tes pada butir soal yang berbeda. Hal ini juga berarti bahwa kemampuan yang dinyatakan dalam model adalah satu-satunya faktor yang memengaruhi respons peserta tes pada butir-butir soal. Bukti yang lain adalah peluang dari pola jawaban setiap peserta tes sama dengan hasil kali peluang jawaban peserta tes pada setiap butir soal.

3. Invariansi

Parameter yang menggambarkan peserta tes tidak tergantung pada distribusi parameter kemampuan mereka, dan fitur item tidak tergantung pada parameter yang mencirikan peserta tes. Spesifikasi soal tes tidak akan berubah hanya karena diberikan kepada kelompok peserta tes dengan tingkat kemampuan yang berbeda-beda, dan kemampuan seseorang tidak akan berubah hanya karena dia mengikuti tes dengan tingkat kesulitan yang berbeda (Retnawati, 2014)

Sebagaimana diungkap Zanon, et.al.(2016) bahwa responden dengan sifat laten yang sama akan memperoleh skor yang sama (terlepas dari kesalahan pengukuran) pada tes yang sama atau sama sekali berbeda yang dibangun untuk mengevaluasi konstruk yang sama. Selain itu, orang dengan sifat laten yang berbeda akan mempertahankan perbedaan yang sama di antara skor mereka, jika mereka menjawab bentuk tes yang paralel.

C. Model Teori Respons Butir

Terdapat beberapa model dalam teori respons butir, model-model yang ada dalam IRT sangat bergantung kepada jenis respons peserta tes (jenis data) seperti soal dengan data dikotomi dan politomi. Pada jenis data respons dikotomi, terdapat beberapa model, diantaranya yaitu: 1) model 1 parameter logistik (1 PLM), 2) model 2 parameter logistik (2 PLM), 3) model 3 parameter logistik (3PLM), dan model 4 parameter logistik (4PLM). Model logistik satu parameter sering disebut juga dengan model Rasch. Fungsi karakteristik butir untuk model logistik satu parameter ditentukan dengan persamaan (Hambleton, et.al, 1991).

Istilah IRT dalam literatur dapat ditemukan sebagai teori distribusi laten, kurva karakteristik butir (*Item Characteristic Curve_ICC*) dan fungsi karakteristik butir (*Item Characteristics Function_ICF*). Kurva karakteristik butir ini disajikan dalam kurva hubungan antara karakteristik butir dengan karakteristik peserta tes (*ability/latent trait*) yang ditunjukkan pada absis sedangkan ordinat menunjukkan probabilitas jawaban peserta pada butir tes. Menurut Hambleton, et.al. (1991) IRT memiliki karakteristik yaitu: (1) Karakteristik butir soal tidak bergantung pada kelompok peserta tes yang mengambil butir soal, (2) Skor yang menunjukkan kemampuan peserta tes tidak bergantung pada tes, (3) Model dinyatakan dalam level (level) item, tidak dalam tes level, (4) Model level tidak memerlukan tes paralel untuk menghitung koefisien reliabilitas, dan (5) Model memberikan ukuran yang sesuai untuk setiap skor kemampuan.

Pada teori tes klasik, karakteristik butir soal akan tergantung pada tingkat kemampuan peserta tes, jika butir soal diselesaikan oleh peserta yang berkemampuan tinggi maka butir soal tersebut menunjukkan tingkat kesulitan yang rendah, sebaliknya untuk peserta yang berkemampuan rendah atau sedang butir tersebut akan menunjukkan tingkat kesulitan yang tinggi. Sebaliknya, teori respons butir (IRT) memprediksi kemampuan peserta dari kemampuan mereka dalam menjawab item tes dengan benar, semakin tinggi kemampuan mereka, semakin tinggi peluang menjawab benar. Begitu juga dengan semakin tinggi tingkat kesukaran soal, maka kemampuan peserta tes menjawab soal dengan benar harus lebih tinggi. Dengan demikian, konsep dasar teori respons butir dianggap sebagai teori yang lebih kuat dibandingkan dengan teori tes klasik.

Ada dua postulat dasar dari teori respons butir (Hambleton, et.al, 1991), yaitu: (1) Sifat, sifat laten, atau kemampuan dapat digunakan untuk memprediksi (atau menjelaskan) hasil pekerjaan peserta tes pada suatu item. (2) Fungsi karakteristik butir soal dapat digunakan untuk menggambarkan hubungan antara hasil kerja peserta tes pada suatu butir soal dengan sifat-sifat yang mendasarinya (fungsi karakteristik butir atau kurva karakteristik butir soal-ICC). Fungsi ini menjelaskan bahwa peluang menjawab benar akan meningkat jika kemampuan peserta tes meningkat.

Lebih lanjut Hambleton, et.al. (1991) menjelaskan beberapa asumsi yang mendasari teori respons butir, yaitu: (a) unidimensi, (b) independensi lokal, (c) fungsi karakteristik butir soal secara akurat

menangkap hubungan nyata antara variabel tak teramati berbasis kemampuan dan variabel teramati (respons butir). Menurut asumsi unidimensi, elemen tes dirancang untuk mengevaluasi hanya satu keterampilan. Jika kemampuan atau sikap tes tetap konsisten, independensi lokal menegaskan reaksi setiap peserta tes terhadap setiap pasangan item adalah independen secara statistik. (Kolen & Brennan, 2014). Dengan kata lain, premis kemandirian lokal (*independensi local*) menyiratkan bahwa jawaban terhadap berbagai item tes oleh peserta tes tidak berkorelasi. Ini juga menyiratkan bahwa satu-satunya variabel yang memengaruhi pengambil reaksi tes terhadap item adalah kemampuan yang ditentukan dalam model.

D. Estimasi Parameter Butir dan Orang (sifat laten)

Penerapan IRT dalam merancang tes dan pengukuran serta menganalisis parameter butir terdiri dari model tertentu yang dipilih. Efektif atau tidaknya penggunaan salah satu model IRT) tergantung pada ketersediaan model yang sesuai dan prosedur yang sesuai untuk memperkirakan parameter yang diestimasi, dan pada program komputer yang sesuai (Al-Tarawnah & Al-Qahtani, 2022).

Secara umum terdapat dua metode estimasi parameter butir dan kemampuan orang, yaitu estimasi kemungkinan maksimum (*Maximum Likelihood Estimation*) dan estimasi Bayesian (*Bayesian Estimation*). MLE sendiri terdiri dari: *Joint Maximum Likelihood Estimation* (JML), *Conditional Maximum Likelihood Estimation* (CML), *Marginal Maximum Likelihood Estimation* (MML). Untuk penjelasan lebih rinci mengenai metode estimasi parameter, Anda bisa membaca buku karya Baker & Kim (2004).

Swaminathan (1985) menjelaskan empat langkah utama untuk menentukan nilai kemampuan peserta tes (*testee*) yang diukur menggunakan tes tertentu, yaitu:

1. Mengompilasi atau menyusun data jawaban peserta tes (*testee*) dari satu butir termasuk kelompok yang cukup besar (jawaban seluruh peserta tes).
2. Memilih satu model respons butir setelah memverifikasi bahwa data cocok dengan model (misalnya, 1PLM, 2PLM, 3PLM atau 4PLM).
3. Menggunakan program berbasis komputer dalam mengestimasi parameter butir dan orang (kemampuan laten) (misalnya BILOG-MG, PRSCALE, MULTILOG dan lain sebagainya).
4. Mengonversi skor kemampuan dengan ukuran yang tepat.

E. Menguji Kecocokan Model

Sudah menjadi hal yang lazim dalam ilmu pengukuran untuk membandingkan kinerja yang diamati pada subkelompok kemampuan yang berbeda dengan kinerja yang diantisipasi di bawah model yang dipilih untuk mengevaluasi model-data yang cocok dengan model IRT pada tugas-tugas tertentu (Maydeu-Olivares, 2013). Menguji kecocokan data dengan model dilakukan dengan beberapa langkah berikut (Stone & Zhang, 2003).

1. Estimasi parameter butir dan kemampuan orang (peserta tes θ)
2. Membentuk sejumlah kecil subkelompok kemampuan (misalnya, $K = 10$) untuk mendekati distribusi kontinu θ .
3. Buat distribusi skor respons teramati untuk setiap subkelompok kemampuan dengan mengklasifikasikan peserta ujian menggunakan estimasi θ dan respons skor mereka.
4. Membentuk distribusi skor respons yang diharapkan di seluruh kategori skor untuk item menggunakan model IRT, estimasi parameter item, dan tingkat θ yang mewakili subkelompok kemampuan diskrit (misalnya, titik tengah subkelompok); dan
5. Membandingkan hasil estimasi dengan distribusi respons skor yang diamati. Perbandingan ini umumnya melibatkan komputasi statistik chi-kuadrat untuk setiap item individual.

Pertanyaan pentingnya adalah apakah model IRT yang dipilih sesuai dengan data. Jika model tidak cocok, kita harus menemukan model yang kurang restriktif (ketat) yang sesuai dengan data atau kita harus membuang item yang tidak sesuai (Gruijter & van der Kamp, 2008). Penyelidikan model fit memiliki dua aspek: 1) Uji statistik model fit; 2) Analisis residual, untuk mendapatkan gambaran keseriusan pelanggaran model.

Beberapa teknik statistic untuk menguji kecocokan model sebagaimana ditulis Gruijter & van der Kamp (2008), yaitu: 1) Statistik Q_1 , (Yen's, 1981), 2) Likelihood Ratio Statistik atau G statistic (McKinley and Mills, 1985, 3) Netherlands Differentiation Test (NDT), 4) Pendekatan CML (Andersen, 1973; Kelderman, 1984; Molenaar, 1983), 5) Pendekatan MML (Reise, Widaman, and Pugh, 1993).

F. Jumlah Sampel dan Panjang Tes

Jumlah sampel dan panjang tes sangat berpengaruh terhadap keakuratan hasil estimasi parameter butir tes dan orang (kemampuan) dengan teori respons butir (IRT) (Gunawan, 2022). Berapa jumlah sampel dan banyak butir tes yang ideal pada penerapan teori respons butir? Belum ditemukan konsensus yang menunjukkan hasil yang konstan mengenai berapa jumlah sampel dan panjang tes yang ideal dalam kerangka estimasi parameter IRT.

Untuk mencari panjang tes (jumlah butir) minimum menggunakan rumusan de Gruijter & van der Kamp (2008) sebagai berikut.

$$n = z_{\alpha}^2 \frac{(\sigma_{\zeta_h} + \sigma_{\zeta_l})^2}{(\zeta_h - \zeta_l)^2}$$

Keterangan:

n : Panjang tes minimum

z_{α} : skor-z yang sesuai dengan probabilitas kumulatif dalam distribusi normal.

$\frac{\sigma_{\zeta_h}}{\zeta_h}$: Simpangan baku peserta tes dengan kemampuan tinggi

$\frac{\sigma_{\zeta_l}}{\zeta_l}$: Simpangan baku peserta tes dengan kemampuan rendah.

ζ_h : Peserta tes dengan level kemampuan tinggi

ζ_l : Peserta tes dengan level kemampuan rendah

Pada model 1PL, ukuran sampel 500 dengan panjang tes 10 masih dapat menghasilkan estimasi parameter kesulitan item yang dapat diterima, karena RMSE pada titik-titik ini masih kurang dari 0,6 (Uyigue, 2019). Temuan ini bertentangan dengan temuan Sahin dan Anil (2017) yang menyimpulkan bahwa ukuran sampel 150 cukup baik dikombinasikan dengan panjang tes 10 butir ke atas. Namun tampaknya ada titik kesepakatan sehubungan dengan panjang tes yaitu trade-off antara panjang tes dan ukuran sampel. Temuan dari penelitian ini sesuai dengan temuan Stone (2003) yang melaporkan konvergensi atau estimasi parameter stabil hanya setelah ukuran sampel sebanyak 500 atau lebih.

Temuan Custer (2015) yang menyimpulkan bahwa ukuran sampel 500 dengan kombinasi panjang tes 40 diperlukan untuk estimasi parameter kesulitan item yang akurat. Meskipun Custer memperoleh sampel 500 sebelum sampai pada kesimpulan ini, kesimpulan dari 500 sampel yang cukup baik tidak jauh dari fakta bahwa ada trade-off antara panjang tes dan ukuran sampel, ketika panjang tes yang lebih kecil diperlukan ukuran sampel akan meningkat untuk mendapatkan perkiraan yang akurat. Implikasi dari temuan tersebut adalah bahwa untuk kerangka pengukuran IRT yang akan digunakan dalam pengembangan tes harus ada hingga 1000 peserta ujian dan hingga 10 item, jumlah sampel sebanyak 1000 peserta ujian jauh dari situasi lapangan, di mana peserta ujian mungkin tidak sampai sebanyak 1000 orang. Anggaphlah penelitian dilakukan di Indonesia, dengan jumlah siswa rata-rata 30 sampai 40 orang per kelas, jika di suatu sekolah terdapat 8 kelas dalam satu tingkat (sebut saja kelas 7 SMP/MTs), maka total jumlah siswa dalam satu tingkat adalah 320 orang. Selain itu, dari penelitian yang dilakukan Gunawan (2022) ditemukan bahwa jumlah sampel 155 dengan kombinasi butir 10 masih konvergen pada model Rasch atau IRT 1 PL.

Lakens (2022) melalui hasil penelitian yang dilakukan terhadap ukuran sampel yang ideal menuliskan:

The awareness that sample sizes in past studies were often too small to meet any realistic inferential goals is growing among psychologists (Button et al., 2013; Fraley & Vazire, 2014; Lindsay, 2015; Sedlmeier & Gigerenzer, 1989). As an increasing number of journals start to require sample size justifications, some researchers will realize they need to collect larger samples than they were used to. This means researchers will need to request more money for participant payment in grant proposals, or that researchers will need to increasingly collaborate (Moshontz et al., 2018)

Secara garis besarnya, Lakens menilai bahwa penggunaan sampel besar dikarenakan tuntutan journal ilmiah bukan karena kebutuhan murni penelitian yang dilakukan. Hal ini mendasarinya untuk membuat ikhtisar kemungkinan pembenaran untuk ukuran sampel dalam suatu penelitian, sebagaimana ditunjukkan pada tabel berikut.

Tabel 2. Ikhtisar Kemungkinan pembenaran untuk ukuran sampel

Jenis Pembeneran	Kapan pembenaran ini berlaku
Mengukur seluruh populasi	Seorang peneliti dapat menentukan seluruh populasi (terbatas), dan dimungkinkan untuk mengukur (hampir) setiap entitas dalam populasi.
Kendala Sumber Daya	Sumber daya yang terbatas adalah alasan utama pemilihan ukuran sampel yang dapat dikumpulkan oleh peneliti
Keakuratan	Pertanyaan penelitian berfokus pada ukuran parameter, dan peneliti mengumpulkan data yang cukup untuk memiliki perkiraan dengan tingkat akurasi yang diinginkan.
A-Priori Power Analysis	Pertanyaan penelitian memiliki tujuan untuk menguji apakah ukuran efek tertentu dapat ditolak secara statistik dengan kekuatan statistik yang diinginkan
Heuristics	Seorang peneliti memutuskan ukuran sampel berdasarkan heuristik, aturan umum atau norma yang dijelaskan dalam literatur, atau dikomunikasikan secara lisan.
Tidak ada alasan pbenar	Seorang peneliti tidak memiliki alasan untuk memilih ukuran sampel tertentu, atau tidak memiliki tujuan inferensial yang ditentukan dengan jelas dan ingin mengomunikasikannya dengan jujur.

Penelitian yang dilakukan Al-Tarawnah & Al-Qahtani (2022) menunjukkan bahwa dengan panjang tes (60 item) dan panjang rata-rata (30 item) lebih efektif menggunakan metode kemungkinan maksimum pada semua kondisi ukuran sampel (500, 1000, 1500), sedangkan tes singkat (10 item) metode estimasi Bayesian mode sebelumnya (prior) lebih unggul dibanding metode kemungkinan maksimum pada keakuratan hasil estimasi untuk semua kondisi ukuran sampel. Selain itu, hasil penelitian menunjukkan bahwa peningkatan akurasi estimasi parameter kemampuan dalam model logistik tiga parameter, menurut kemungkinan maksimum dan metode Bayesian, meningkat dengan bertambahnya jumlah item tes. Untuk membandingkan kedua metode sehubungan dengan kisaran

akurasi estimasi parameter kemampuan, hasil menunjukkan bahwa dalam kasus tes pendek (10 item) dan panjang rata-rata (30 item), metode Bayesian mengungguli sehubungan dengan akurasi estimasi pada semua kondisi ukuran sampel. Namun, dengan menggunakan tes panjang (60 item), metode kemungkinan maksimum mengungguli semua kondisi ukuran sampel yang berbeda

Berdasarkan beberapa hasil penelitian tersebut di atas dapat disimpulkan bahwa prosedur IRT masih terbatas dan mengalami kendala serius untuk penerapannya dalam pengembangan tes di sekolah. Perlu kehati-hatian dan pertimbangan yang cermat dalam menggunakan software (program komputer), karena berkaitan dengan metode estimasi yang digunakan dalam menganalisis parameter butir dan orang (sifat laten).

2 Program R (Pengantar)

“Sudah tidak ada lagi Setan bodoh yang menentang Tuhan & Al-Qur’an, yang banyak sekarang itu adalah Setan yang berpenampilan Nabi & bermulut Al-Qur’an”

-M.H. Ainun Najib-

A. Sejarah singkat R

Ross Ihaka dan Robert Gentleman mengembangkan R sebagai perangkat lunak gratis yang digunakan untuk mengajar ketika mereka menjadi rekan kerja di University of Auckland di Selandia Baru. Karena mereka berdua akrab dengan program S yaitu bahasa pemrograman komersial untuk statistic yang dikembangkan pada tahun 1960an-1970an oleh para peneliti di Labaoratorium Bell di Kota New Jersey (pembaca bisa mencari informasi lebih lanjut mengenai program S di buku-buku tentang program S yang banyak beredar). Ihaka dan Gentleman menggunakan sintaks serupa pada pekerjaan mereka sendiri dan mengumumkan perangkat lunak tersebut di milis S-news, software gratis (open source) karya mereka kemudian diberi nama R (diambil dari inisial Ross dan Robert) (Davies, 2016), beberapa orang menjadi tertarik dan mulai berkolaborasi dengan mereka, terutama Martin Mächler.

Saat ini, terdapat 18 orang memiliki hak untuk mengubah arsip pusat kode sumber. Kelompok ini disebut sebagai tim inti (*core team*) pengembangan R. Selain itu, banyak orang lain telah menyumbangkan kode baru dan perbaikan *bug* untuk proyek tersebut. Berikut adalah beberapa tonggak pengembangan R sebagaimana ditulis de Vries & Meys (2012):

- **Awal 1990-an:** Pengembangan R dimulai.
- **Agustus 1993:** Perangkat lunak diumumkan di milis S-news. Sejak saat itu, satu set milis R aktif telah dibuat. Halaman web di www.r-project.org/mail.html memberikan deskripsi daftar ini dan instruksi untuk berlangganan.

- **Juni 1995:** Setelah beberapa argumen persuasif oleh Martin Mächler untuk membuat kode yang tersedia sebagai "perangkat lunak gratis," kode tersebut dibuat tersedia di bawah GNU General Public License (GPL) Free Software Foundation, Versi 2.
- **Pertengahan 1997:** Tim Inti Pengembangan R awal dibentuk (walaupun, pada saat itu, mereka hanya dikenal sebagai kelompok inti).
- **Februari 2000:** Versi pertama R, versi 1.0.0, dirilis.

Selanjutnya, Ross Ihaka menulis tinjauan komprehensif tentang pengembangan R. Halaman web <http://cran.r-project.org/doc/html/interface98-paper/paper.html> memberikan sejarah yang menarik bagi pembaca untuk mengetahui detail sejarah R.

B. Definisi dan Konsep R

R merupakan bahasa pemrograman, lingkungan pemrosesan statistik, cara untuk memecahkan masalah, dan kumpulan alat yang berguna untuk membuat hidup Anda lebih mudah. R bukanlah aplikasi, sehingga pengguna memiliki kebebasan untuk memilih alat pengeditan sendiri untuk berinteraksi dengan R.

Sebagai bahasa pemrograman dan lingkungan untuk komputasi statistik dan grafik, mirip dengan bahasa S yang awalnya dikembangkan di Bell Labs. Hal Ini sangat mendukung terwujudnya suatu solusi open source analisis data yang didukung oleh komunitas riset besar dan aktif di seluruh dunia.

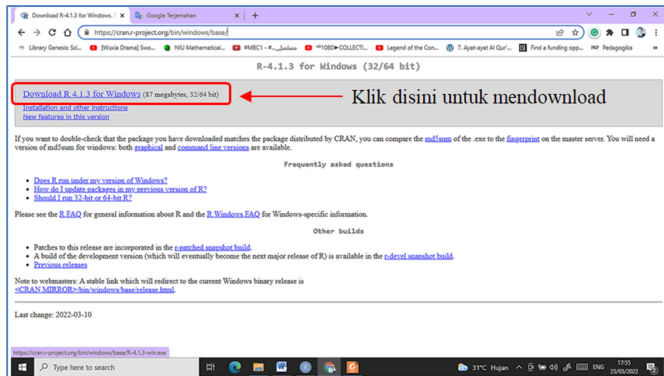
R memiliki beberapa karakteristik khusus dibandingkan dengan aplikasi statistic lainnya:

- ❖ Kebanyakan platform perangkat lunak statistik komersial berharga ribuan bahkan puluhan ribu dollar. Untuk SPSS versi terbaru SPSS 28 harganya Rp. 4.500.000. Sedangkan R gratis, sangat cocok buat seorang guru atau siswa/mahasiswa dengan manfaat yang jelas.
- ❖ R adalah platform statistik yang komprehensif, menawarkan segala macam teknik analisis data. Hampir semua jenis analisis data dapat dilakukan di R.
- ❖ R berisi rutinitas statistik lanjutan yang belum tersedia di paket lain. Bahkan, metode baru tersedia untuk diunduh setiap minggu. Jika Anda pengguna SAS, bayangkan mendapatkan PROC SAS baru setiap beberapa hari.

- ❖ R memiliki kemampuan grafis canggih. Jika Anda ingin memvisualisasikan data yang kompleks, R memiliki kumpulan fitur paling komprehensif dan kuat yang tersedia.
- ❖ R adalah platform yang kuat untuk analisis dan eksplorasi data interaktif. Sejak awal, itu dirancang untuk mendukung pendekatan yang diuraikan dalam Gambar 1.1. Misalnya, hasil dari setiap langkah analitik dapat dengan mudah disimpan, dimanipulasi, dan digunakan sebagai masukan untuk analisis tambahan.
- ❖ Mendapatkan data ke dalam bentuk yang dapat digunakan dari berbagai sumber bisa menjadi proposisi yang menantang. R dapat dengan mudah mengimpor data dari berbagai sumber, termasuk file teks, sistem manajemen basis data, paket statistik, dan penyimpanan data khusus. Itu dapat menulis data ke sistem ini juga. R juga dapat mengakses data langsung dari halaman web, situs media sosial, dan berbagai layanan data online.
- ❖ R menyediakan platform yang tak tertandingi untuk memprogram metode statistik baru dengan cara yang mudah dan lugas. Ini mudah diperluas dan menyediakan bahasa alami untuk memprogram dengan cepat metode yang baru diterbitkan.
- ❖ Secara fungsional R dapat diintegrasikan ke dalam aplikasi yang ditulis dalam bahasa lain, termasuk C++, Java, Python, PHP, Pentaho, SAS, dan SPSS. Ini memungkinkan Anda untuk terus bekerja dalam bahasa yang mungkin Anda kenal, sambil menambahkan kemampuan R ke aplikasi Anda.
- ❖ R beroperasi di berbagai platform, termasuk Windows, Unix, dan Mac OS X, bahkan, kita juga bisa menemukan panduan untuk menginstal R pada iPhone, yang mengesankan tetapi mungkin bukan ide yang bagus.)
- ❖ Jika Anda tidak ingin mempelajari bahasa baru, tersedia berbagai antarmuka pengguna grafis (GUI), menawarkan kekuatan R melalui menu dan dialog.

C. Menginstal R

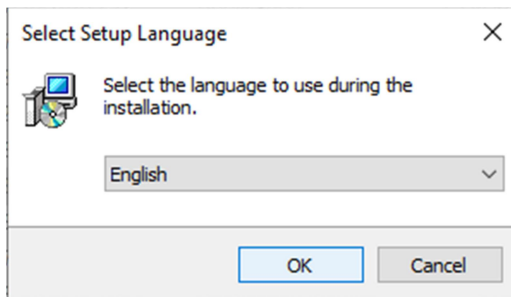
Program R bisa didapatkan (*di-download*) secara gratis melalui situs <https://cran.r-project.org/bin/windows>.



Gambar 1. Tampilan situs R

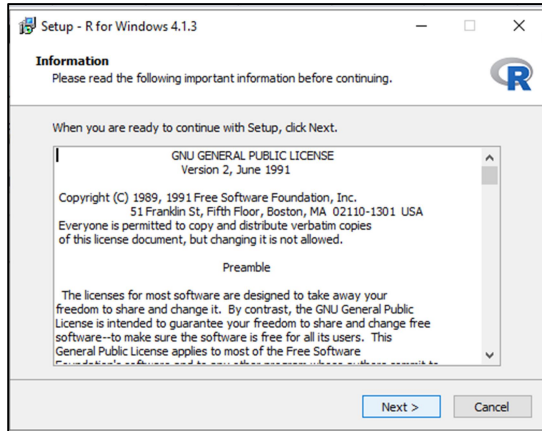
Adapun langkah-langkah instalasi R di komputer/laptop sebagai berikut.

1. Pilih aplikasi R yang sudah didownload, sehingga muncul kotak dialog yang meminta Anda untuk memilih bahasa yang akan digunakan, berhubung bahasa Indonesia tidak tersedia, pilih English dan tekan tombol **OK**.



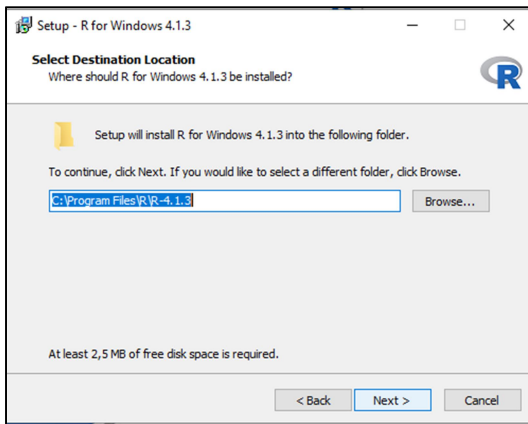
Gambar 2. Kotak dialog instalasi R

2. Setelah menekan tombol **OK**, instalasi R akan menyuguhkan kotak dialog informasi yang penting bagi user (ada baiknya dibaca) sebagaimana tampak pada gambar berikut, jika sudah silahkan tekan tombol **Next**.



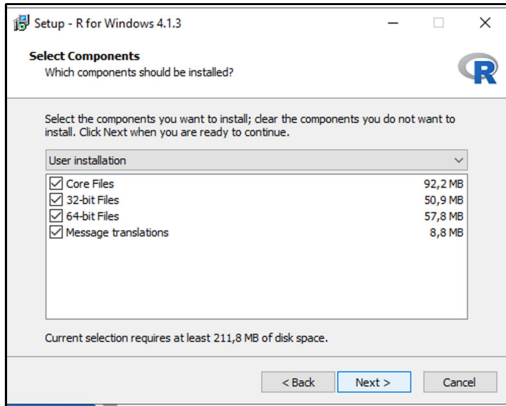
Gambar 3. Kotak dialog Setup_Informasi Instalasi R

3. Kotak dialog berikutnya akan memunculkan dan meminta Anda untuk memilih drive dan folder tempat menyimpan instalasi (biasanya di C), klik **Next**.



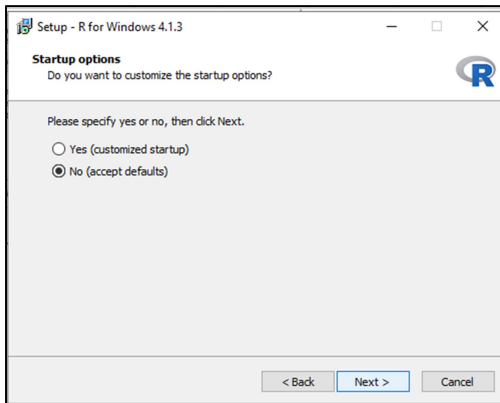
Gambar 4. Kotak dialog Setup_pemilihan lokasi tujuan instalasi R

4. Pilih komponen apa saja yang akan disertakan dalam instalasi, seperti tampak pada gambar berikut. Setelah itu tekan tombol **Next**.



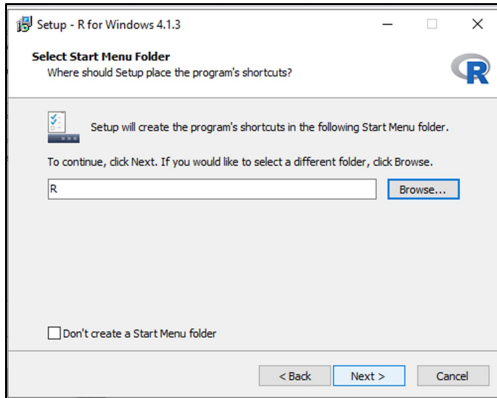
Gambar 5. Kotak dialog Setup_pemilihan komponen instalasi R

5. Pilih pengaturan mulai (Startup) defaults seperti gambar berikut, lalu tekan tombol **Next**.



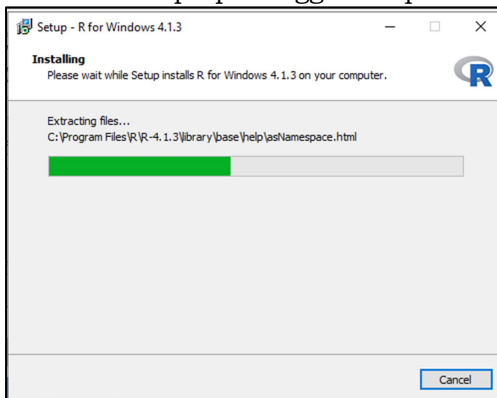
Gambar 6. Kotak dialog Setup_opsi instalasi R

6. Pilihan untuk menempatkan instalasi (program) sesuai kebutuhan, pilih **Next**



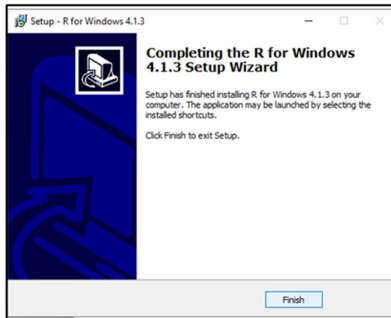
Gambar 7. Kotak dialog Setup_pemilihan folder instalasi R

7. Instalasi berjalan dan jangan kemana-mana, apalagi sampai mematikan laptop. tunggu sampai selesai

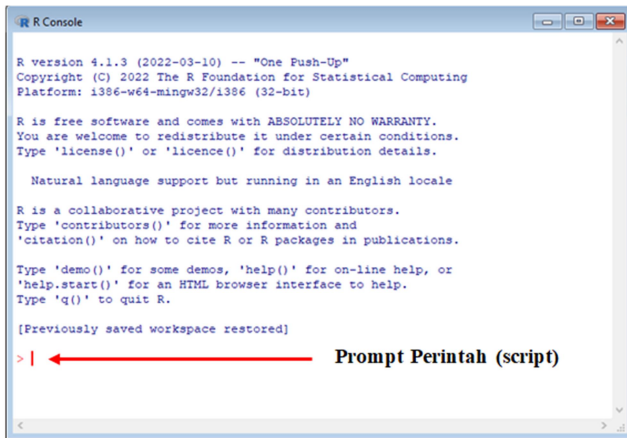


Gambar 8. Kotak dialog proses instalasi R

8. Jika instalasi sudah selesai, akan muncul kotak dialog sebagaimana gambar berikut. Tekan tombol **Finish**.



Tampilan logo R di layar monitor

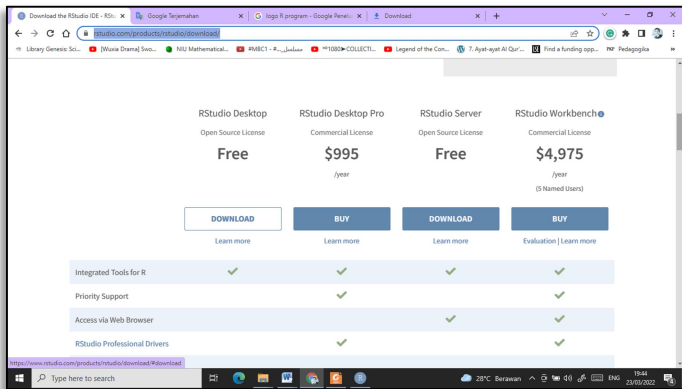


Gambar 9. Layar Console ketika R pertama kali di buka.

Jangan lupa download juga Rstudio melalui:

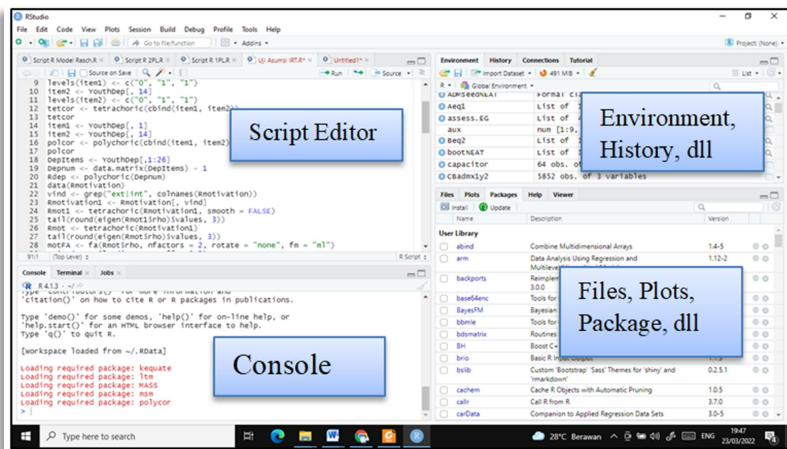
<https://www.rstudio.com/products/rstudio/download/>

RStudio adalah seperangkat alat terintegrasi yang dirancang untuk membantu kerja kita menjadi lebih produktif dengan R dan Python. RStudio terdiri dari beberapa jendela yaitu: jendela konsol, editor sintaks yang mendukung eksekusi kode langsung, dan berbagai alat canggih untuk merencanakan, melihat riwayat, men-debug, dan mengelola ruang kerja.



Gambar 10. Situs Rstudio download

Langkah instalasi Rstudio hampir sama dengan instalasi R. Tampilan layar Rstudio setelah diinstalasi sebagai berikut.



Gambar 11. Penjelasan layar Rstudio.

Layar Rstudio terdiri dari layar *script editor* (untuk menuliskan/mengedit perintah), layar *console*, jendela *environment* dan *files* sebagaimana ditunjukkan pada gambar 11 di atas.

Source : Sudut kiri atas layar berisi editor teks yang memungkinkan Anda bekerja dengan file script sumber. Di sini, Anda dapat memasukkan beberapa baris kode, menyimpan file script Anda ke disk, dan melakukan tugas lain pada script Anda. Editor kode ini bekerja seperti editor teks lain yang pernah Anda lihat. Jendela source mengenali dan menyoroti berbagai elemen kode Anda, misalnya (menggunakan warna berbeda untuk elemen berbeda), dan juga membantu Anda menemukan tanda kurung yang cocok dalam script Anda atau memberikan tanda merah pada baris perintah yang memuat kode yang salah.

Console: Di sudut kiri bawah, Anda menemukan Console. Console di RStudio identik dengan Console di RGui. Di sinilah kita bisa melihat script yang kita tuliskan di layar Source di jalankan secara interaktif dengan R.

Workspace and History: Sudut kanan atas adalah ikhtisar praktis dari ruang kerja Anda, tempat Anda dapat memeriksa variabel yang Anda buat di setiap sesi serta nilainya. Ini juga merupakan area di mana Anda dapat melihat riwayat perintah yang Anda keluarkan di R.

File, plot, package, dan help: Di kanan bawah sudut, kita memiliki akses ke beberapa alat:

- **File**: Di sini kita dapat menelusuri folder dan file di komputer.
- **Plot**: Di sini R menampilkan plot yang kita buat berdasarkan script (perintah) yang kita berikan.
- **Packages**: layar untuk melihat daftar semua paket yang kita instal. Paket adalah kumpulan kode mandiri yang menambahkan fungsionalitas ke R, mirip dengan cara add-in menambahkan fungsionalitas ke Microsoft Excel.
- **Help**: Di sini kita dapat menelusuri sistem bantuan bawaan R.

Paket (Packages) adalah kumpulan fungsi R, data, dan kode yang dikompilasi dalam format yang terdefinisi dengan baik. Ada lebih dari 5.500 paket yang dapat diunduh dari <http://cran.r-project.org/web/packages>. Paket-paket ini menyediakan berbagai kemampuan baru yang luar biasa, mulai dari analisis data geospasial hingga pemrosesan spektrum massa protein hingga

analisis tes psikologis. Direktori tempat paket disimpan di komputer disebut library. Fungsi `.libPaths()` menunjukkan di mana library berada, dan fungsi `library()` menunjukkan paket apa saja yang telah disimpan di library. Sebagai contoh kita akan melakukan analisis IRT dengan R. Sebelum melakukan analisis harus dipanggil dulu library yang menyediakan paket perintah (*packages*) yang tersedia di R. Misalnya:

```
> library(ltm) ##analisis parameter butir IRT model 1 dan 2PL  
> library(equateIRT) ##penyetaraan tes dengan metode IRT
```

R lebih merupakan bahasa pemrograman daripada aplikasi. Saat Anda mengunduh R, Anda secara otomatis mengunduh aplikasi console yang sesuai dengan sistem operasi Anda. Namun, aplikasi ini hanya memiliki fungsionalitas dasar, dan sampai batas tertentu berbeda dari satu sistem operasi ke sistem operasi berikutnya. Berbeda dengan Rstudio yang merupakan aplikasi lintas platform dengan beberapa fitur (ciri khas) yang sangat rapi untuk mendukung R.

D. Dasar-dasar R

R memiliki empat kelompok operator dan fungsi matematika yang berbeda:

- Operator aritmatika dasar: Operator ini digunakan di hampir semua bahasa pemrograman.
- Fungsi matematika: kita dapat menemukan fungsi lanjutan ini pada kalkulator teknis.
- Operasi vektor: Operasi vektor adalah fungsi yang membuat perhitungan pada vektor lengkap, seperti `sum()`. Setiap hasil bergantung pada lebih dari satu nilai vektor.
- Operasi matriks: Fungsi-fungsi ini digunakan untuk operasi dan perhitungan

1. Skalar dan Vektor

Operasi aritmatika untuk scalar di R, menggunakan symbol sebagaimana biasanya digunakan di kalkulator seperti.

Tabel 3. Simbol operasi matematika di R

Simbol	Keterangan	Simbol	Keterangan
+	Operasi penjumlahan	>	Lebih dari
-	Operasi pengurangan	<	Kurang dari
*	Operasi perkalian	>=	Lebih besar atau sama dengan
/	Operasi pembagian	<=	Lebih kecil atau sama dengan
%%	Pembagian integer	==	Setara
%/%	Sisa dari pembagian integer	!=	Tidak setara
		&	Berdasarkan elemen dan
			Berdasarkan elemen atau

Contoh:

Operasi penjumlahan

```
> 15+20
```

```
[1] 35
```

Operasi perkalian

```
> 8*40
```

```
[1] 320
```

Operasi pembagian

```
> 35/9
```

```
[1] 3.888889
```

Operasi perpangkatan (eksponensial)

```
> 4^8
```

```
[1] 65536
```

```
> 4**8
```

```
[1] 65536
```

Operasi modulo (sisa pembagian)

```
> 15%%2
```

```
[1] 1
```

Operasi pembagian yang dibulatkan ke bawah (bilang bulat)

```
> 15%/%2
```

```
[1] 7
```

Vektor di R sangat penting terutama untuk menangani banyak item. Vektor terdiri dari vektor numeric, string dan logika. Secara numerik, vektor merupakan kumpulan pengamatan atau

pengukuran mengenai satu variabel, misalnya, berat badan 100 orang mahasiswa Universitas Hamzanwadi atau jumlah bubur ayam yang Anda konsumsi setiap hari. Struktur data yang lebih rumit dapat terdiri dari beberapa vektor, dan untuk membuat fungsi vektor di R adalah dengan mengetikkan huruf c (mendefinisikan vektor) dengan entri yang diinginkan dalam tanda kurung yang dipisahkan dengan koma.

Contoh berat badan 5 orang mahasiswa Universitas Hamzanwadi:

```
> BB <- c(40,55,60,65,80)
> BB
[1] 40 55 60 65 80
##Tinggi Badan
> TB <- c(1.60,1.65,1.70,1.72,1.75)
> TB
[1] 1.60 1.65 1.70 1.72 1.75
```

Kita bisa melakukan operasi matematika dengan vektor asalkan vektor tersebut memiliki panjang yang sama, seperti contoh di atas, kita bisa mencari indeks massa tubuh (BMI) masing-masing orang, di mana BMI didefinisikan sebagai berat badan dibagi dengan kuadrat tinggi badan.

```
> BMI <- BB/TB^2
[1] 15.62500 20.20202 20.76125 21.97134 26.12245
```

Vektor string merupakan vektor yang elemennya adalah huruf atau kata yang diberi tanda petik ganda. Contoh

```
> hari <- c("Senin","Selasa","Rabu","Kamis",
+          "Jumat", "Sabtu", "Ahad")
> hari
[1] Senin Selasa Rabu Kamis Jumat Sabtu Ahad
```

2. Fungsi Matematika

Hampir semua fungsi matematika tersedia di R dengan argumen dalam tanda kurung seperti.

Tabel 4. Fungsi matematika di R

Argumen	Keterangan
Abs()	absolute value
Exp()	Exponential
Gamma()	Fungsi Gamma
Lgamma()	log of gamma function
Log()	Fungsi logaritma
log10()	Logaritma dengan basis 10
Sign()	Fungsi signum
Sqrt()	Square root (akar kuadrat)
Floor()	Integer terbesar, kurang dari atau sama dengan
Ceiling()	Integer terkecil, lebih dari atau sama dengan.
Trunk()	pemotongan ke bilangan bulat terdekat
Factorial()	Fungsi factorial
Lfactorial()	Log factorial

Contoh penerapan fungsi matematika di R

```
> x <- c(1,5,8,7,6)
> x
[1] 1 5 8 7 6
> abs(x)
[1] 1 5 8 7 6
> x^2
[1] 1 25 64 49 36
> log(x)
[1] 0.000000 1.609438 2.079442 1.945910 1.791759
> log(x, base=2) ## Log x dengan basis 2
[1] 0.000000 2.321928 3.000000 2.807355 2.584963
> exp(x)
[1] 2.718282 148.413159 2980.957987 1096.633158 403.428793
> sqrt(x)
[1] 1.000000 2.236068 2.828427 2.645751 2.449490
> factorial(x)
[1] 1 120 40320 5040 720
```

Fungsi pembulatan angka di R

Meskipun R dapat menghitung secara akurat hingga 16 digit, namun kita dapat menggunakan beberapa fungsi di R untuk membulatkan angka. Untuk membulatkan angka menjadi dua digit

setelah titik desimal, misalnya gunakan fungsi `round()` sebagai berikut:

```
> round(38.131631463713, digits=2)
[1] 38.13
```

Jika Anda ingin menentukan jumlah angka penting dipertahankan, terlepas dari ukuran nomornya, Anda menggunakan fungsi `signif()` sebagai gantinya:

```
> signif(38.131631463713, digits=5)
[1] 38.132
```

Selain dengan `round()`, kita juga bisa menggunakan tiga fungsi lainnya untuk melakukan pembulatan:

- `floor(x)` dibulatkan ke bilangan bulat terdekat yang lebih kecil dari x . Jadi `floor(312.38)` menjadi 312 dan `floor(-312.38)` menjadi -313
- `ceiling(x)` dibulatkan ke bilangan bulat terdekat yang lebih besar dari x . contoh `ceiling(312.38)` menjadi 313 dan `ceiling(-312.38)` menjadi -312.
- `trunc(x)` dibulatkan ke bilangan bulat terdekat dalam arah 0. Jadi `trunc(312.38)` menjadi 312 dan `trunc(-312.38)` menjadi -312.

Fungsi trigonometri

```
> sin(30)
[1] -0.9880316
> cos(30)
[1] 0.1542514
> tan(30)
[1] -6.405331
```

kode ini tidak memberi hasil yang benar, karena R selalu bekerja dengan sudut dalam radian, bukan dalam derajat. Dengan demikian gunakan nilai pi ($\pi = 3.141593$)

```
> sin(30*pi/180)
[1] 0.5
> cos(45*pi/180)
```

Materi lengkap tentang R dan perintah (script) yang digunakan untuk berbagai keperluan, seperti analisis statistik, analisis faktor dan analisis persamaan structural (SEM), Anda bisa membaca buku-buku R yang khusus untuk itu.

Matriks

Matriks merupakan kumpulan dari beberapa vektor dengan struktur unik dan properti dimensi tambahan, atau dengan kata lain matriks adalah vektor dua dimensi yang bersifat atomik, yang berarti bahwa semua item dalam baris dan kolomnya harus termasuk dalam kelas yang sama. Matriks dapat dibuat dengan berbagai cara. Salah satunya memerlukan pembuatan vektor dan kemudian menggunakan fungsi `dim()` untuk mengatur dimensinya.

Contoh:

```
> Mate <- c(2,4,6,8)
> dim(Mate) <- c(2, 2)
> Mate
```

```
      [,1] [,2]
[1,]    2    6
[2,]    4    8
```

```
> Matika <- c(3,5,7,9)
> dim(Matika) <- c(2, 2)
> Matika
```

```
      [,1] [,2]
[1,]    3    7
[2,]    5    9
```

Operasi Matematika pada matrik

```
> Jumlah <- Mate+Matika
> Jumlah
```

```
      [,1] [,2]
[1,]    5   13
[2,]    9   17
```

```
> Kali <- Mate*Matika
> Kali
```

```
      [,1] [,2]
[1,]    6   42
[2,]   20   72
```

```
> Bagi <- Mate/Matika
> Bagi
```

```

      [,1]      [,2]
[1,] 0.6666667 0.8571429
[2,] 0.8000000 0.8888889
> Kurang <- Mate-Matika
> Kurang
      [,1] [,2]
[1,]   -1  -1
[2,]   -1  -1

```

Menggabungkan dua matrik bisa dilakukan dengan perintah berdasarkan baris (rbind) atau colom (cbind).

Contoh:

```

> M1 <- c(2,3,6,7)
> M2 <- c(3,5,8,9)
> rbind(M1,M2)
      [,1] [,2] [,3] [,4]
M1     2   3   6   7
M2     3   5   8   9
> cbind(M1,M2)
      M1 M2
[1,]  2  3
[2,]  3  5
[3,]  6  8
[4,]  7  9

```

Mengganti nama kolom

```

> gbr <- rbind(M1,M2)
> colnames(gbr) <- c("Wawan", "Sadityo", "Nugroho", "Siti")
> gbr
      Wawan Sadityo Nugroho Siti
M1     2     3     6     7
M2     3     5     8     9

```

Data Frame

Data Frame digunakan untuk menyimpan data dalam format tabel yang memiliki struktur seperti matriks karena data frame juga dua dimensi, tetapi dapat memiliki tipe data yang berbeda di setiap kolom, seperti halnya daftar. Pembatasan logis dalam data frame adalah bahwa setiap kolom hanya dapat berisi objek dari kelas yang sama. Untuk membuat data frame menggunakan fungsi `data.frame()`.

Contoh:

```
> x <- c("Muhammad","Ali","Gunawan","Amalia","Fitri")
> y <- c(100,90,70,80,85)
> z <- c(TRUE,TRUE,FALSE,TRUE,TRUE)
> Cth <- data.frame>Nama=x, IQ=y, KET=z)
> Cth
  Nama IQ  KET
1 Muhammad 100 TRUE
2     Ali   90 TRUE
3 Gunawan  70 FALSE
4  Amalia  80 TRUE
5   Fitri  85 TRUE
```

Sequence (urutan)

Pembuatan urutan adalah tugas yang sederhana namun sangat penting, yaitu menetapkan dasar pada tugas yang lebih penting, seperti referensi dalam subset struktur atau vektorisasi. Cara paling sederhana untuk membuat urutan adalah dengan menggunakan operator “:”

Contoh: Misalkan kita ingin membuat vektor numeric dari 3 sampai 15, maka script yang digunakan adalah

```
> x <- 3:15
> x
[1] 3 4 5 6 7 8 9 10 11 12 13 14 15
```

Atau kita juga bisa menggunakan script seq seperti berikut untuk melakukan urutan dari nilai tertentu sampai nilai yang kita inginkan dengan jarak tertentu juga, seperti contoh berikut.

```
> x <- seq(from=2, to=15, by=3)
> x
[1] 2 5 8 11 14
```

Jika kita ingin membagi membagi dari-ke ruang interval yang ditentukan, digunakan “length”

```
> y <- seq(from=3, to=20, length=5)
> y
[1] 3.00 7.25 11.50 15.75 20.00
```

Terkadang kita ingin membuat urutan angka yang berulang secara kelompok, misalnya urutan angka 1 sampai 5 sebanyak 3 kali.


```
> p <- rep(1:5, 3)
> p
[1] 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
```

E. Simulasi (Membangkitkan Data) Dengan R

Penelitian bidang pengukuran dan penilaian yang melibatkan siswa di suatu sekolah atau daerah terkadang menggunakan data bangkitan atau simulasi. Simulasi dilakukan mengingat pemberian perlakuan yang dicobakan (diubah-ubah), karena tidak memungkinkan untuk dilakukan pada kehidupan nyata, seperti mengacak jumlah peserta tes, jumlah butir tes, materi pembelajaran dan sebagainya. Sebagaimana diungkap Rubinstein dan Kroese (2007) bahwa melalui simulasi komputer, seseorang dapat mempelajari perilaku sistem kehidupan nyata yang terlalu sulit untuk diperiksa secara analitik. Selain itu, menurut Retnawati (2008) studi simulasi dapat memberikan sumbangan yang sangat besar pada perkembangan teori respons butir bila dibandingkan dengan menggunakan data nyata (riil) di lapangan.

Langkah-langkah umum dalam studi simulasi berorientasi kejadian (respons benar dan salah dalam menjawab soal) menurut Kroese, et.al. (2011) sebagai berikut.

1. Buat model matematika untuk sistem tersebut: misalkan model matematika untuk mengestimasi kemampuan peserta didik (θ), karakteristik butir (model 1 parameter logistik, 2 parameter logistik atau 3 parameter logistik)
2. Identifikasi status sistem, bergantung pada aspek sistem mana yang sedang diselidiki:
3. Siapkan struktur data yang sesuai: antrian, daftar, variabel, penghitung statistik, dan lain-lain.
4. Identifikasi kemungkinan kejadian dan interaksinya, misalnya menggunakan grafik kejadian/peristiwa (kurve karakteristik butir, kurve informasi butir, kurve fungsi informasi dan lain sebagainya).
5. Siapkan daftar kejadian yang sesuai dan parameter simulasi utama.
6. Menerapkan parameter-parameter/model untuk setiap kemungkinan kejadian. Gambar diagram alir untuk setiap model jika perlu.
7. Uji dan debug, ini adalah langkah yang sangat penting dan seringkali memakan waktu. Pada saat memeriksa kesalahan

pemrograman, verifikasi bahwa sistem berperilaku baik melalui program dan/atau dengan memplot realisasi proses status. Bandingkan hasil simulasi dengan kasus khusus yang diketahui (informasi sebelumnya).

8. Lakukan analisis statistik terhadap data.
9. Laporkan hasilnya.

Agar hasil simulasi memberikan informasi yang berguna dan sesuai dengan kondisi sebenarnya di lapangan, Rubinstein dan Kroese (2007) menegaskan bahwa hasil simulasi harus merepresentasikan model yang diteliti secara valid, tepat dan akurat. Simulasi semacam ini hanya dapat dilakukan dengan metode yang lebih canggih (Fox, 2010) seperti halnya metode yang membangkitkan data dengan parameter hasil penelitian sebelumnya (posterior), metode simulasi ini dikenal sebagai Markov Chain Monte Carlo (MCMC).

Markov Chain Monte Carlo (MCMC) adalah metode umum untuk estimasi pengambilan sampel dari distribusi arbitrer. Ide utamanya adalah untuk menghasilkan rantai Markov yang distribusi pembatasnya sama dengan distribusi yang diinginkan (Kroese, Taimre & Botev, 2011) dengan cara membangun urutan yang menyatu dalam distribusi ke distribusi posterior (target). Kemudian rata-rata sampel dihitung untuk memperkirakan ekspektasi posterior (Fox, 2010). Pada dasarnya MCMC adalah integrasi Monte Carlo menggunakan rantai Markov di mana pengambilan sampel digunakan untuk melakukan integrasi dan mengestimasi parameter model dengan estimasi Bayesian.

Ada tiga prinsip yang perlu diperhatikan peneliti saat merancang dan melaksanakan studi simulasi Monte Carlo (Bulut dan Sunbul, 2017). Prinsip-prinsip tersebut adalah keaslian, kelayakan, dan reproduktifitas. Keaslian studi simulasi Monte Carlo mengacu pada sejauhmana studi simulasi tersebut mencerminkan kondisi nyata. Misalnya, asumsikan bahwa seorang peneliti ingin menyelidiki dampak panjang tes pada estimasi kemampuan yang diperoleh dari model IRT tertentu. Peneliti memilih 30, 60, 90, dan 300 item sebagai nilai hipotesis untuk faktor panjang tes. Karena tes 300 item sangat tidak mungkin terjadi dalam kehidupan nyata, peneliti mungkin harus mempertimbangkan untuk menghilangkan opsi ini dari studi simulasi. Keaslian studi simulasi Monte Carlo juga terkait dengan kebutuhan faktor simulasi. Melanjutkan contoh yang sama, peneliti mungkin mempertimbangkan ukuran sampel sebagai faktor potensial untuk studi simulasi tentang pemulihan perkiraan

kemampuan; tetapi ukuran sampel diketahui tidak berpengaruh pada estimasi kemampuan (sifat laten) ketika parameter item sudah diketahui (Bulut, 2013; Bulut, Davison, & Rodriguez, 2017). Oleh karena itu, peneliti tidak perlu memasukkan ukuran sampel sebagai faktor simulasi dalam penelitian.

Kelayakan studi Monte Carlo mengacu pada keseimbangan antara tujuan studi simulasi dan ruang lingkup studi simulasi. Kombinasi dari banyak faktor simulasi dan jumlah ulangan yang tinggi seringkali dapat menghasilkan studi simulasi yang sangat kompleks dan sulit diselesaikan serta diringkas dalam jangka waktu yang wajar. Oleh karena itu, peneliti harus menentukan faktor simulasi mana yang penting dan berapa banyak ulangan yang dapat dilakukan berdasarkan ruang lingkup studi simulasi. Reprodusibilitas studi simulasi Monte Carlo mengacu pada kemungkinan bahwa peneliti yang melakukan studi simulasi dapat mereplikasi temuan yang sama di lain waktu atau bahwa peneliti lain yang memiliki akses ke parameter simulasi dapat mereplikasi (atau setidaknya mendekati) temuan. Feinberg & Rubright (2016) mengatakan bahwa untuk memastikan reproduktifitas, peneliti harus menentukan parameter sebelum menghasilkan data dan menyimpan catatan parameter terpilih dalam studi simulasi. Peneliti dapat menggunakan parameter untuk mereplikasi temuan di kemudian hari atau memberikannya kepada peneliti lain yang tertarik untuk mereplikasi temuan.

Simulasi adalah transkripsi langsung ke istilah komputasi dari proses stokastik alami (seperti dalam contoh *solitaire*/permainan kartu) atau simulasi dengan bilangan random pada aplikasi Excel dengan perintah “@randbetween(bottom;up)”. Simulasi Monte Carlo sebaliknya adalah solusi dengan metode probabilistik dari masalah nonprobabilistik seperti simulasi dengan pemodelan dengan faktor penskalaan $D=1,702$. Kalos dan Whitlock (2008) mengatakan bahwa Monte Carlo sebenarnya merupakan simulasi dengan parameter-parameter yang telah ditetapkan berdasarkan hasil penelitian sebelumnya atau kejadian-kejadian yang dimanipulasi sedemikian rupa secara algoritmik agar sesuai dengan kondisi real (nyata) di lapangan.

Aplikasi R dapat digunakan untuk tujuan pembangkitan data, sesuai dengan kebutuhan atau data sebelumnya. Adapun script yang digunakan sebagaimana ditulis Horton & Kleinman (2015),

```

> intercept = -1
> beta = 0.5
> n = 5000
> xtest = rnorm(n, mean=0, sd=1)
> linpred = intercept + (xtest * beta)
> prob = exp(linpred)/(1 + exp(linpred))
> ytest = ifelse(runif(n) < prob, 1, 0)
> summary(ytest)
> ytest

```

Baker & Kim (2017), menuliskan script untuk membangkitkan data dengan model 2 PL (dengan sedikit modifikasi penulis) sebagai berikut.

```

> par(lab=c(7,3,3))
> theta <- seq(-3, 3, .1)
> theta

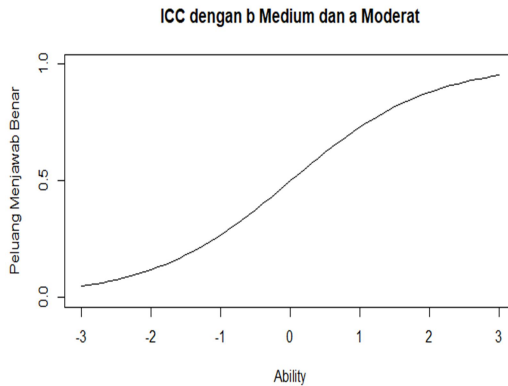
[1] -3.0 -2.9 -2.8 -2.7 -2.6 -2.5 -2.4 -2.3 -2.2 -2.1 -2.0 -1.9 -1.8 -1.7 -1.6 -1.5
[17] -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1
[33] 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7
[49] 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9 3.0

> bmedium <- 0
> amoderate <- 1
> P <- 1 / (1 + exp(-amoderate * (theta - bmedium)))
> P

[1] 0.04742587 0.05215356 0.05732418 0.06297336 0.06913842 0.07585818 0.08317270
[8] 0.09112296 0.09975049 0.10909682 0.11920292 0.13010847 0.14185106 0.15446527
[15] 0.16798161 0.18242552 0.19781611 0.21416502 0.23147522 0.24973989 0.26894142
[22] 0.28905050 0.31002552 0.33181223 0.35434369 0.37754067 0.40131234 0.42555748
[29] 0.45016600 0.47502081 0.50000000 0.52497919 0.54983400 0.57444252 0.59868766
[36] 0.62245933 0.64565631 0.66818777 0.68997448 0.71094950 0.73105858 0.75026011
[43] 0.76852478 0.78583498 0.80218389 0.81757448 0.83201839 0.84553473 0.85814894
[50] 0.86989153 0.88079708 0.89090318 0.90024951 0.90887704 0.91682730 0.92414182
[57] 0.93086158 0.93702664 0.94267582 0.94784644 0.95257413

> plot(theta, P, type="l", xlim=c(-3,3), ylim=c(0,1),
+       xlab="Ability", ylab="Peluang Menjawab Benar",
+       main= "ICC dengan b Medium dan a Moderat")

```



Gambar 12. ICC butir tes dengan data simulasi ($a = 1$; $b = 0$)

F. Paket Perintah (*Packages*) R

Packages (paket) yang tersedia untuk menganalisis butir soal dengan pendekatan teori respons butir (IRT) begitu banyak tersedia di R. Beberapa di antaranya yang sering digunakan adalah “*eRm*”, “*ltm*”, dan “*mirt*” dan masih banyak lagi yang lainnya.

Pembaca perlu mempertimbangkan lebih mendalam bahwa setiap paket (*packages*) memiliki asumsi-asumsi yang harus dipenuhi untuk menjalankan operasi fungsi (*script*) yang ada di R. Hal ini dikarenakan bahwa baik estimasi parameter butir maupun estimasi parameter orang (kemampuan laten) memiliki banyak sekali metode di antaranya: 1) estimasi kebolehjadian maksimum gabungan (*Joint Maximum Likelihood_JML*); 2) *Marginal Maximum Likelihood* (MML), 3) *Conditional Maximum Likelihood* (CML); dan 4) *Marginal Likelihood Ratio* (LR). Keterangan lebih lanjut mengenai metode estimasi dapat Anda baca di buku-buku statistik yang secara khusus membahas tentang itu.

3 Uji Asumsi IRT dengan R

“Guru yang benar-benar bijaksana tidak meminta Anda untuk memasuki rumah kebijaksanaannya, tetapi justru menuntun Anda ke ambang batas pikiran Anda.”

-Khalil Gibran-

Uji asumsi perlu dilakukan untuk melihat apakah data yang diperoleh dari hasil penelitian memenuhi syarat atau layak dianalisis menggunakan teori respons butir atau tidak, yaitu: a) unidimensi, b) independensi lokal dan c) invariansi.

A. Unidimensi

Asumsi unidimensi dapat dibuktikan dengan menggunakan analisis faktor, yakni dengan melihat nilai *eigen* pada matrik varian kovarian inter-butir (Retnawati, 2014, lihat juga de Gruijter & van der Kamp, 2008). Tahapan awal pada analisis faktor adalah menganalisa kecukupan sampel dengan menggunakan Kaiser-Meyer-Olkin (KMO) dan uji Bartlet. Berikut paket dan script yang digunakan untuk analisis faktor menggunakan R.

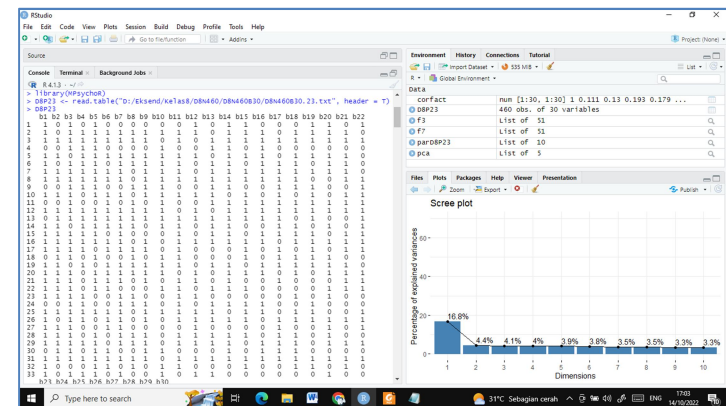
```
##Exploratory Factor Analysis (EFA)--> Buku Analisis Butir  
Pilihan Ganda  
##Langkah-langkah  
##Instal paket ("corpor") --> Matrik korelasi parsial  
##Instal paket ("GPArotation") --> Rotasi faktor  
##Instal paket ("psych") --> Analisis Faktor  
##Instal paket ("MASS") --> Analisis Faktor  
##Instal paket ("parellel") --> plot atau grafik  
##Jika sudah terinstal, silahkan diaktifkan.  
> ##Jika sudah terinstal, silahkan diaktifkan.  
> library(corpor)  
> library(GPArotation)  
> library(psych)  
> library(MASS)
```

- > library(parallel)
- > library(factoextra)
- > library(FactoMiner)
- > library(MPpsychR)

##Panggil data yang akan dianalisis, kami menggunakan contoh data bangkitan dengan nama D8P23 (Data kelas 8 Paket data ke 23), hanya sekedar contoh.

- > D8P23 <- read.table("D:/Eksend/Kelas8/D8N460/D8N460B30/
+ D8N460B30.23.txt", header = T)
- > D8P23

Pada layar Console ditampilkan seperti gambar 13 berikut (R-Studio).



Gambar 13. Tampilan data frame pada layar Console

Analisis kecukupan sampel dengan KMO (Kaiser-Meyer-Olkin) menggunakan script berikut.

> KMO(D8P23)

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = D8P23)

Overall MSA = 0.87

MSA for each item =

```

b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14 b15 b16
0.86 0.88 0.87 0.86 0.80 0.85 0.87 0.90 0.90 0.86 0.86 0.88 0.86 0.79 0.86 0.89
b17 b18 b19 b20 b21 b22 b23 b24 b25 b26 b27 b28 b29 b30
0.84 0.85 0.88 0.88 0.90 0.84 0.85 0.84 0.90 0.83 0.87 0.91 0.86 0.88

```

```
> cortest.bartlett(D8P23,n=460)
```

```
R was not square, finding R from data
```

```
$chisq
```

```
[1] 1669.26
```

```
$p.value
```

```
[1] 9.762813e-144
```

```
$df
```

```
[1] 435
```

```
> f3 <- fa(D8P23,nfactors=3, SMC=TRUE, min.err = .001,
+ max.iter=100, fm="ml", rotate = "none", n.obs = 460)
> print(f3,digits=2)
```

```
Factor Analysis using method = ml
```

```
Call: fa(r = D8P23, nfactors = 3, n.obs = 460, rotate =
"none",
```

```
SMC = TRUE, min.err = 0.001, max.iter = 100, fm = "ml")
```

```
Standardized loadings (pattern matrix) based upon correlation
matrix
```

	ML1	ML2	ML3	h2	u2	com
b1	0.41	0.07	-0.14	0.194	0.81	1.3
b2	0.38	-0.01	0.00	0.142	0.86	1.0
b3	0.33	0.17	0.14	0.159	0.84	1.9

b28	0.47	0.10	-0.15	0.255	0.74	1.3
b29	0.33	-0.07	0.13	0.134	0.87	1.4
b30	0.46	0.18	0.01	0.238	0.76	1.3

	ML1	ML2	ML3
SS loadings	4.21	0.56	0.48
Proportion Var	0.14	0.02	0.02
Cumulative Var	0.14	0.16	0.18
Proportion Explained	0.80	0.11	0.09
Cumulative Proportion	0.80	0.91	1.00

```
Mean item complexity = 1.4
```

```
Test of the hypothesis that 3 factors are sufficient.
```

```
The degrees of freedom for the null model are 435 and the
objective function was 3.72 with Chi Square of 1669.26
```


The degrees of freedom for the model are 348 and the objective function was 0.75

The root mean square of the residuals (RMSR) is 0.03
The df corrected root mean square of the residuals is 0.04

The harmonic number of observations is 460 with the empirical chi square 470.43 with prob < 1.3e-05
The total number of observations was 460 with Likelihood Chi Square = 336.54 with prob < 0.66

Tucker Lewis Index of factoring reliability = 1.012
RMSEA index = 0 and the 90 % confidence intervals are 0.015

BIC = -1797.12

Fit based upon off diagonal values = 0.94

Measures of factor score adequacy

	ML1	ML2	ML3
Correlation of (regression) scores with factors	0.92	0.69	0.62
Multiple R square of scores with factors	0.84	0.48	0.38
Minimum correlation of possible factor scores	0.68	-0.04	-0.24

```
> parD8P23 <- fa.parallel(D8P23, fa="pc", n.iter=500,  
+   ylabel = "Eigenvalues", quant = .50)
```

Parallel analysis suggests that the number of factors = NA
and the number of components = 1

Hasil analisis faktor (*exploratory factor analysis*) paket data sebagaimana telah dipaparkan di atas menunjukkan bahwa paket data dengan nama D8P23 memenuhi syarat kecukupan sampel, di mana rerata nilai Kaiser-Meyer-Olkin (KMO) lebih besar dari 0,5000 yaitu sebesar 0,87 sebagaimana berikut.

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = D8P23)

Overall MSA = 0.87

Begitu juga dengan uji Bartlett

```
> cortest.bartlett(D8P23,n=460)
```

R was not square, finding R from data

\$chisq

[1] 1669.26

\$p.value

[1] 9.762813e-144

\$df

[1] 435

Nilai p yang dihasilkan jauh lebih kecil dari 0,05 yaitu 9.762813e-144 dengan derajat kebebasan ($df/dk = 435$), artinya bahwa paket data memenuhi syarat kecukupan sampel (hubungan antar variable/konstruk atau dalam hal ini item yang dianalisis Valid). Selanjutnya, output R menunjukkan bahwa semua butir yang dianalisis juga memenuhi kriteria kecukupan sampel (MSA masing-masing butir jauh lebih besar dari nilai korelasi diagonal butir di atas 0,5. Sehingga dapat disimpulkan bahwa data memenuhi syarat kecukupan sampel (Valid).

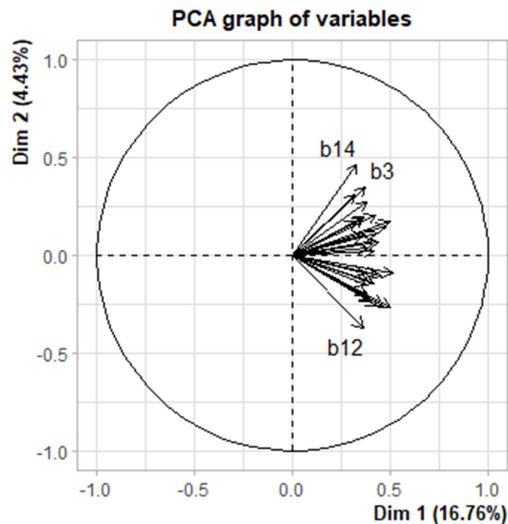
Sering muncul pertanyaan, apakah analisis faktor memang digunakan untuk membuktikan validitas instrumen (*construct validity*) dan reliabilitas tes atau hanya sekadar membuktikan asumsi IRT? Analisis faktor (EFA dan CFA) merupakan alat analisis yang akurat untuk membuktikan apakah tes atau soal yang digunakan “valid” dari segi materi atau konstruknya (struktur variable penyusunnya), dengan melihat nilai hubungan antar variable (nilai KMO dan MSA) (Sylva, 2020). Sebagaimana diketahui bahwa skor perolehan peserta tes bervariasi dan merupakan penjumlahan dari skor sebenarnya dengan kesalahan, jika nilai KMO kurang dari atau sama dengan 0,5000 artinya kesalahan pengukuran dan jawaban murni peserta tes berimbang (sama-sama 50%), tentu saja kita tidak ingin mengambil resiko yang tidak diharapkan (reliabilitas tes rendah) (Watkins, 2021).

b1	0,86	b7	0,87	b13	0,86	b19	0,88	b25	0,9
b2	0,88	b8	0,9	b14	0,79	b20	0,88	b26	0,83
b3	0,87	b9	0,9	b15	0,86	b21	0,9	b27	0,87
b4	0,86	b10	0,86	b16	0,89	b22	0,84	b28	0,91
b5	0,80	b11	0,86	b17	0,84	b23	0,85	b29	0,86
b6	0,85	b12	0,88	b18	0,85	b24	0,84	b30	0,88

Selanjutnya, untuk mengetahui jumlah komponen atau dimensi yang diukur oleh paket data dilakukan dengan melihat nilai *eigen* (*factor loading* atau *SS loadings*) berikut.

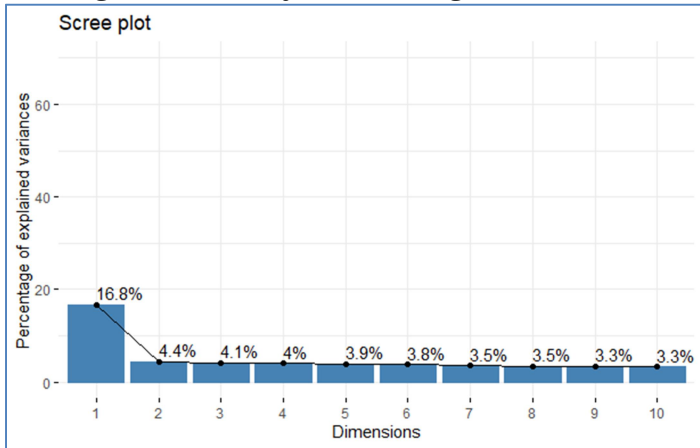
	ML1	ML2	ML3
SS loadings	4.21	0.56	0.48
Proportion Var	0.14	0.02	0.02
Cumulative Var	0.14	0.16	0.18
Proportion Explained	0.80	0.11	0.09
Cumulative Proportion	0.80	0.91	1.00

Nilai eigen pada komponen 1 atau SS loadings (dengan ML = Maximum Likelihood) sebesar 4.21, komponen 2 sebesar 0,56 dan komponen 3 sebesar 0,48, dengan demikian dapat dikatakan bahwa paket data yang dianalisis memenuhi asumsi unidimensi, karena ada satu faktor yang nilai eigen (SS loading > 1) yaitu komponen 1. Pada Proportion Var dapat dilihat bahwa komponen 1 dapat menjelaskan variabilitas sebesar 0,14 atau 14% dan jika digabung dengan komponen 2 maka kedua komponen tersebut bisa menjelaskan variansi sebesar 16%. Pada gambar atau grafik PCA graph of variables berikut menunjukkan bahwa data yang dianalisis terdiri dari 2 komponen dengan komponen 1 merupakan faktor yang dominan



Gambar 14. Grafik dimensi data menggunakan PCA

Pada gambar 14 menunjukkan bahwa komponen satu (dimensi 1) dapat menjelaskan 16,76% variansi yang diukur sedangkan komponen dua (dimensi 2) mampu menjelaskan variansi sebesar 4,43%. Artinya data yang dianalisis berdimensi satu (unidimensi) karena terdapat hanya satu faktor dominan yang diukur. Pembuktian unidimensi juga bisa menggunakan grafik *Scree Plot* (Script untuk analisis faktor dapat dilihat di lampiran 06)yang dihasilkan sebagaimana ditunjukkan oleh gambar 15 berikut.



Gambar 15. *Scree Plot*

Gambar 15 tersebut di atas menunjukkan bahwa pola yang dibentuk oleh grafik *Scree Plot* curam pada komponen 1 dan mulai landai pada komponen 2 dan seterusnya. Hal ini membuktikan bahwa terdapat satu faktor atau komponen yang dominan pada perangkat paket data (Retnawati, 2014).

Interpretasi ini didasarkan pada pendapat Mair (2018) yang menyatakan bahwa dengan menggunakan *scree plot*, ada dua hal yang harus diperhatikan. Pertama, pencarian batas yang jelas antara struktur sistematis (batuan) dan struktur acak (*scree*), dengan memperhatikan titik siku yang menunjukkan jumlah faktor dari siku ke kiri. Dengan cara ini, komponen pertama akan terlihat sangat mendominasi. Kedua, menurut Guttman (1954, dalam Thompson, 2004) dan Kaiser (1960, dalam Mair, 2018), nilai eigen harus lebih besar dari 1. Alasan di baliknya adalah bahwa faktor yang memiliki nilai eigen kurang dari 1 variabilitasnya rendah.

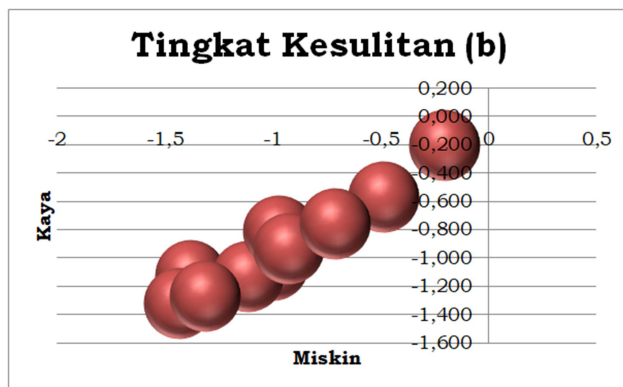
B. Independensi lokal

Independensi lokal berarti bahwa ketika kemampuan yang memengaruhi kinerja tes dipertahankan konstan, respons peserta tes terhadap pasangan item apa pun secara statistik independen. Dengan kata lain, setelah mempertimbangkan kemampuan peserta tes, tidak ada hubungan antara jawaban peserta tes pada item yang berbeda. Sederhananya, menurut Hambleton, et.al. (1991) bahwa kemampuan yang ditentukan dalam model adalah satu-satunya faktor yang memengaruhi jawaban peserta tes pada item tes.

Asumsi independensi lokal secara otomatis terpenuhi/terbukti setelah asumsi unidimensi data respons peserta terhadap suatu tes terbukti (Retnawati, 2014), sebab kedua konsep ini setara atau ekuivalen (Lord, 1980; Lord & Novick, 1968 dalam Hambleton, et.al., 1991). Karena kedua konsep ini (unidimensi dan independensi lokal) setara, maka secara logika matematis dapat dikatakan jika $A=B$ dan $B=C$ maka $A=C$.

C. Invariansisi

Asumsi/prasyarat ketiga dari teori respons butir adalah invariansi parameter butir dan parameter kemampuan. Pembuktian asumsi ini dilakukan dengan memperkirakan parameter butir pada kelompok peserta tes yang berbeda (misalnya: jenis kelamin, status sosial ekonomi, daerah asal/tempat tinggal peserta tes, dan lain sebagainya). Gambar 16 berikut ini menampilkan diagram pencar tingkat kesulitan butir soal yang dibuat berdasarkan status sosial ekonomi orang (sebut saja “Kaya” dan “Miskin”).



Gambar 16. Scatter plot invariansi

Pada gambar 16 tersebut di atas, dapat dilihat bahwa tidak terjadi variansi tingkat kesulitan (b) berdasarkan status sosial ekonomi orang tua siswa (“Kaya” dan “Miskin”). Hal ini terlihat dari titik (*point*) tingkat kesulitan butir tes yang dijawab oleh siswa yang ekonomi orang tuanya mapan (“Kaya”) dan tingkat kesulitan butir tes siswa yang ekonominya kurang (“Miskin”) berada pada/sekitar garis lurus, hal ini memberi arti bahwa tingkat status sosial ekonomi orang tua siswa tidak memengaruhi tingkat kesulitan butir tes yang diujikan.

4 Analisis Karakteristik Butir Dikotomus

“Jalanilah kehidupan di dunia ini tanpa membiarkan dunia hidup di dalam dirimu, karena ketika perahu berada di atas air, ia mampu berlayar dengan sempurna, tetapi ketika air masuk ke dalamnya, perahu itu tenggelam.”

-Ali bin Abi Thalib-

A. Definisi dan Konsep

Model jawaban dikotomi adalah model jawaban yang terdiri dari dua kategori yaitu benar-salah (benar = 1 dan salah = 0) atau data pilah (diskrit), biasa digunakan pada tes dengan model jawaban pilihan ganda (*multiple choice*). Model IRT dikotomus terdiri dari beberapa model, di antaranya model 1 parameter logistik (1PL), model 2 parameter logistik (2PL), model 3 parameter logistik (3PL) dan model 4 parameter logistik (4PL).

B. Model Rasch dan IRT 1 Parameter Logistik

Wright dan Stone (1999) dengan mengutip Rasch (1960) menjabarkan bahwa model pengukuran Rasch secara fundamental dikembangkan oleh Rasch 1960/1980, 1961, 1967 dan 1977 dengan aksioma objektivitasnya yang sangat lugas bahwa “Mirip dengan bagaimana satu hal menjadi lebih sulit daripada yang lain berarti bahwa untuk setiap individu kemungkinan menyelesaikan item kedua dengan benar juga akan berkurang, seseorang yang lebih mampu daripada yang lain harus memiliki kemungkinan yang lebih baik untuk menyelesaikan setiap item dari jenis pertanyaan itu” (Boomsma, van Duijn & Snijders, 2001)

Penggunaan model pengukuran Rasch menurut Smith (1999); Bond & Fox (2015) memiliki dua asumsi utama. *Pertama*, data

harus mendekati satu dimensi (unidimensi). Artinya, sebagian besar elemen harus menghasilkan data dalam konfigurasi dasar yang sama. *Kedua*, independensi lokal. Artinya, kemungkinan jawaban yang benar untuk satu pertanyaan tidak boleh bergantung pada jawaban spesifik untuk pertanyaan lain. Independensi lokal dapat terjadi ketika satu set item didasarkan pada stimulus yang sama, seperti satu bagian pada tes membaca. Tingkat membaca suatu bagian dapat memiliki efek tetap pada semua faktor yang terkait dengan bagian tersebut. *Kedua* asumsi ini secara diatur secara ketat dalam statistik kecocokan yang tersedia dalam model Rasch.

Data yang cocok dengan model Rasch akan menghasilkan beberapa hal yang sangat berguna dalam pengukuran. *Pertama*, skor untuk semua parameter dilaporkan pada metrik umum dan skala dengan jarak yang sama. *Kedua*, semua estimasi parameter tidak tergantung pada sifat acak dari distribusi parameter. Artinya, estimasi parameter tidak tergantung pada sampel atau tes. Ini berarti bahwa data yang hilang bukanlah masalah, tidak paralel atau tumpang tindih. *Ketiga*, ekspresi probabilitas berdasarkan persamaan logit dapat digunakan untuk sampai pada nilai respons yang diharapkan dengan menggabungkan perkiraan ukuran seseorang dan perkiraan ukuran objek arbitrer. Nilai respons yang diharapkan ini dapat dibandingkan dengan respons yang diamati dan digabungkan dalam berbagai cara yang berguna untuk memberikan tes langsung kecocokan item, orang, penilai, atau subset item, orang, atau penilai model.

Kerangka kerja ini memungkinkan kita untuk mendeteksi berbagai pola respons anomali, termasuk tetapi tidak terbatas pada tebakan dan kesalahan, dan untuk mengevaluasi dan memperbaiki dampak dari respons tak terduga ini pada perilaku orang. *Keempat*, skor individu dan subjek memiliki perkiraan kesalahan standar asimtotik untuk setiap skor mentah individu berdasarkan fungsi informasi. *Kelima*, karena kesalahan standar asimtotik pengukuran ukuran orang (peserta tes) diketahui, kesalahan pengukuran dapat dipisahkan dari masalah pencocokan model data pengukuran saat menghitung koefisien reliabilitas.

Salah satu karakteristik model Rasch sebagaimana dipaparkan di atas adalah tidak tergantung pada sampel yang digunakan (Rasch, Kubinger, & Yanagida, 2011). Dimensi Rasch secara bersamaan menyusun pertanyaan dari yang paling sulit ke yang paling mudah, dan responden dari yang tertinggi hingga yang terendah. Dengan demikian, inkonsistensi atau pola yang tidak

biasa (outlier) terdeteksi dalam tanggapan responden. Dalam teori klasik, proses pengukuran berfokus pada skor nyata (x), sedangkan model Rasch menggunakan data yang mewakili peluang skor (P) dibandingkan dengan jumlah jawaban dan pertanyaan yang benar. Skor odds kemudian diubah menjadi nilai odds ratio menggunakan rumus berikut:

$$\text{Odds Rasio} = P/(1-P)$$

Dari persamaan di atas, nilai logit dapat diperoleh dengan memasukkan fungsi logaritma pada kedua ruas, sehingga persamaan di atas menjadi (Sumintono & Widhiarso, 2013).

$$\text{Logit} = \text{Log } P/(1-P)$$

Logit, W-score, atau ukuran nilai kemudian digunakan untuk menggambarkan nilai ini. Nilai logit fleksibel dan berguna untuk berbagai investigasi..

1. Kurva Karakteristik Butir (ICC) dan Kurva Informasi Butir (IIC)

Pemodelan Rasch menggabungkan algoritma untuk mewakili hasil harapan probabilistik item i dan responden v , yang secara sistematis dinyatakan sebagai berikut (untuk data dikotomis).

$$P(X_{vi} = x_{vi} | \theta_v, \beta_i) = \frac{\exp[x_{vi}(\theta_v - \beta_i)]}{1 + \exp(\theta_v - \beta_i)}$$

Keterangan:

X_{vi} = $n \times k$ matrik data, dengan elemen $x_{vi} = 0$ atau 1

β_i = k -vektor pada parameter butir

θ_v = n -vektor pada parameter orang

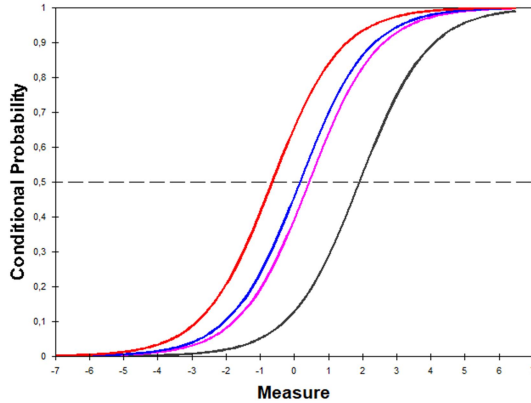
Persamaan di atas menjelaskan bahwa kemampuan responden dikurangi tingkat kesulitan soal tes dapat digunakan untuk menghitung kemungkinan satu keberhasilan. Model Rasch juga biasa dikenal dengan model IRT 1 parameter logistik. Fungsi karakteristik butir untuk model IRT 1 parameter logistik (Crocker & Algina, 2008):

$$P_i(\theta) = \frac{e^{Da(\theta-b_i)}}{1 + e^{Da(\theta-b_i)}}, i = 1,2,3, \dots, n$$

Keterangan:

- $P_i(\theta)$ = peluang seseorang dengan kemampuan(θ) menjawab butir soal ke-i dengan benar.
- b_i = parameter tingkat kesulitan untuk butir soal ke -i
- n = banyaknya butir soal dalam tes.
- e = bilangan pokok logaritma natural, yang nilainya adalah 2,718, jika dibulatkan tiga angka dibelakang koma.
- a = parameter daya pembeda (nilainya dibuat konstan atau sama dengan 1 pada semua butir pada suatu paket tes).
- D = faktor skala, yang nilainya diambil sebesar 1,7. yaitu simpangan baku distribusi logistik yang paling mendekati distribusi normal

Parameter b_i untuk item i adalah titik pada skala kemampuan di mana probabilitas jawaban yang benar adalah 0,5 (Hambleton, et.al., 1991). Parameter ini merupakan parameter lokasi atau lebih dikenal dengan parameter tingkat kesulitan suatu tes yang menunjukkan posisi ICC dalam kaitannya dengan skala kemampuan (θ). Semakin besar kemampuan (θ) yang dibutuhkan oleh seorang peserta tes untuk memiliki kemungkinan 50% menjawab pertanyaan dengan benar, semakin sulit butir soal tersebut. Butir-butir tes yang termasuk ke dalam kategori mudah ditemukan di ujung kiri atau bawah skala kemampuan, sedangkan butir-butir yang menantang ditemukan di ujung kanan atau atas, sebagaimana terlihat pada gambar 17 ICC berikut.



Gambar 17. Hubungan antara tingkat kesulitan (b_i) dan tingkat kemampuan (θ) pada model Rasch

Gambar 17 tersebut di atas menjelaskan bahwa ketika nilai kemampuan suatu kelompok ditransformasikan sedemikian hingga rata-ratanya adalah 0 dan simpangan bakunya adalah 1, nilai parameter b_i bervariasi (biasanya) dari sekitar -3,0 sampai dengan +3,0 atau $(-3 \leq b \leq +3)$. Nilai parameter b_i mendekati -3,0 ketika item tersebut sangat mudah, dan nilai parameter b_i mendekati +3,0 ketika item tersebut sangat sulit bagi kelompok peserta tes. Selanjutnya, kisaran kemampuan laten peserta tes (θ) adalah $-\infty \leq \theta \leq +\infty$, tetapi menurut Baker & Kim (2017) rentang terbatas yang digunakan biasanya adalah $-3 \leq \theta \leq +3$ atau $-4 \leq \theta \leq +4$ (de Gruijter & van der Kamp, 2008).

Model Rasch bekerja dengan konsep yang berbeda dari TCC atau IRT 1 parameter logistik. Secara statistik model Rasch tidak melibatkan adanya variable $D = 1.7$ (pembuat normal fungsi) dan parameter daya pembeda (a) yang dibuat konstan sama dengan 1. Wright (1968, dalam Crocker & Algina, 2008) memberikan penjelasan bahwa ada model ogive normal satu parameter yang analog dengan model logistik satu parameter, artinya bahwa model Rasch memiliki konsep yang berbeda dengan model 1 parameter logistik (1 PLM) pada teori respons butir (IRT).

2. Estimasi Parameter Butir dan Orang (Kemampuan)

Estimasi parameter model 1 PL dilakukan dengan fungsi likelihood

$$L(\theta, b; u) = \prod_i \prod_j P_i(\theta_j; b_i)^{u_{ij}} [1 - P_i(\theta_j; b_i)]^{1-u_{ij}}$$

Langkah selanjutnya adalah memilih metode yang dapat memaksimumkan fungsi likelihood pada persamaan di atas dengan metode kebolehdjian gabungan maksimum (*Joint Maximum Likelihood*)

$$L(\theta, \delta; u) = \prod_i \prod_j \frac{\left(\frac{\theta_j}{\delta_i}\right)^{u_{ij}}}{(1 + \theta_j/\delta_i)} = \frac{\prod_i \delta_i^{u_i} \prod_j \theta_j^{u_j}}{\prod_i \prod_j (1 + \theta_j/\delta_i)}$$

$\theta \equiv (\theta_1, \dots, \theta_N)$; $\delta \equiv (\delta_1, \dots, \delta_N)$; $u \equiv u_{ij}$. u_i dan u_j adalah penjumlahan marginal pada data matriks. Karena jumlah marjinal dapat ditunjukkan sebagai statistik yang cukup untuk parameter kesulitan dan butir, itu berlaku untuk fungsi kemungkinan yang terkait dengan probabilitas bersyarat dari (U_{ij}) dengan $U_i = u_i$.

$$L(\theta; u|u_i) = \frac{\prod_j \theta^{u_j}}{\prod_i \gamma_{ui}}$$

γ_{ui} adalah fungsi kombinatorial dari parameter kemampuan yang dikenal sebagai fungsi simetris dasar. Sebagai alternatif untuk estimasi CML, estimasi parameter item dapat ditemukan melalui maksimalisasi fungsi kemungkinan marginal, yang diperoleh dari persamaan (5) dengan mengintegrasikan melalui kerapatan (densitas) umum pada θ , yang biasanya diambil sebagai kerapatan normal dengan parameter μ dan σ :

$$L(\delta; u, \mu, \sigma) = \int_{-\infty}^{\infty} L(\delta, \theta; u) f(\theta; \mu, \sigma) d\theta$$

Cara yang efisien untuk menghitung perkiraan kemungkinan maksimum marjinal (MML) ini adalah dengan menggunakan algoritma EM (Thissen, 1982, dalam van der Linden & Hambleton, 1997). Marginalisasi seperti pada persamaan (6) adalah cara lain untuk menghilangkan dampak parameter kemampuan saat memperkirakan parameter butir.

3. Uji Kecocokan Model (Goodness of Fit)

Terdapat beberapa uji statistik untuk kecocokan model Rasch di antaranya sebagaimana ditulis van der Linden & Hambleton (1997) adalah uji rasio likelihood Andersen (1973). Tes ini dirancang untuk menguji apakah estimasi parameter item dalam kelompok skor yang berbeda sama dengan kesalahan pengambilan sampel dari estimasi. Molenaar (1983) menawarkan statistik yang sensitif terhadap perbedaan daya pembeda antar item. Van den Wollenberg (1982) memperkenalkan statistik uji Q1 dan Q2, yang sensitif terhadap pelanggaran asumsi unidimensi.

Baker & Kim (2017) menyebutkan bahwa pertimbangan penting dalam teori respons butir adalah apakah model kurva karakteristik butir tertentu cocok dengan data respons butir untuk suatu butir tertentu. Kesesuaian proporsi yang diamati dari respons yang benar dan yang dihasilkan oleh kurva karakteristik butir sesuai/nyocok dengan butir yang diukur dengan indeks kecocokan chi-kuadrat, yaitu:

$$\chi^2 = \sum_{g=1}^G f_g \frac{[p(\theta_g) - P(\theta_g)]^2}{P(\theta_g)Q(\theta_g)}$$

Keterangan:

G : Jumlah kemampuan kelompok

θ_g : Level kemampuan pada group g

f_g : Jumlah peserta ujian yang memiliki kemampuan g

$p(\theta_g)$: Proporsi yang diamati dari jawaban yang benar untuk kelompok g

$P(\theta_g)$: Probabilitas jawaban yang benar untuk kelompok g yang dihitung dari model kurva karakteristik butir (ICC) menggunakan estimasi parameter butir

$$Q(\theta_g) = 1 - P(\theta_g)$$

Aspek kontroversial yang masih menjadi perhatian para guru atau peneliti dalam mengembangkan soal/instrument penelitian adalah apakah kalau hasil analisis butir menunjukkan ada butir tertentu yang tidak cocok dengan model harus dibuang atau apakah model yang digunakan perlu direlaksasi (perubahan model dengan tidak mengubah hal mendasar) seperti mengubah model Rasch atau IRT 1 parameter logistik menjadi IRT 2 atau 3 parameter logistik. Tentu saja untuk pilihan kedua (mengubah model) akan berdampak

kepada kekakuratan hasil analisis, sebab semakin bertambahnya parameter (misal dari 1 PL ke 2PL atau 3PL) akan membutuhkan sampel yang lebih besar. Sehingga pada prakteknya, sebagian besar peneliti atau guru mengambil pilihan pertama yaitu menghapus butir yang tidak cocok dari kumpulan butir yang dianalisis.

Faktor tambahan seperti (1) karkater ketidaksesuaian, (2) ketersediaan item pengganti, (3) jumlah waktu yang tersedia untuk menulis ulang item, (4) ketersediaan sampel butir yang cukup besar untuk memperkirakan parameter butir dengan tepat untuk model yang lebih umum, dan-mungkin yang paling penting adalah (5) tujuan prosedur pengujian. Untuk itu disarankan kepada peneliti atau guru yang hendak mengembangkan instrumen penelitian ataupun penilaian di sekolah, hendaknya mempertimbangkan jumlah butir yang diujicobakan, jika terjadi adanya butir yang tidak cocok dengan model, guru atau peneliti bisa dengan segera mengganti butir yang tidak cocok tersebut dengan butir lain yang sudah disiapkan (konten pertanyaannya harus sama).

C. Model 2 Parameter Logistik (2PL)

Model respons dua parameter berdasarkan distribusi normal Ogive dibuat oleh Lord pada tahun 1952. Persamaan berikut digunakan oleh Birnbaum untuk mengubahnya menjadi model logistik dua parameter pada tahun 1968 (van der Linden & Hambleton, 1997).

$$P_i(\theta) = \frac{1}{1 + \exp\{-a_i(\theta - b_i)\}}$$

atau

$$P_i(\theta) = \frac{e^{D_i a_i (\theta - b_i)}}{1 + e^{D_i a_i (\theta - b_i)}}, i = 1, 2, 3, \dots, n$$

Keterangan:

$P_i(\theta)$ = peluang seseorang dengan kemampuan(θ) menjawab butir soal ke- i dengan benar.

b_i = parameter tingkat kesulitan untuk butir soal ke- i

n = banyaknya butir soal dalam tes.

e = bilangan pokok logaritma natural, yang nilainya adalah 2,718, jika dibulatkan tiga angka dibelakang koma.

D = faktor skala, yang nilainya diambil sebesar 1,7, yaitu simpangan baku distribusi logistik yang paling mendekati distribusi normal

1. Estimasi Parameter Butir dan Orang (kemampuan laten)

Persamaan tersebut di atas menunjukkan bahwa model dua parameter jelas merupakan generalisasi dari model satu parameter yang memungkinkan untuk membedakan item yang berbeda, bila pada model 1 parameter (model Rasch) hanya melibatkan 1 parameter yaitu tingkat kesulitan, maka model 2 parameter memuat adanya parameter daya pembeda item soal.

Parameter daya pembeda item didefinisikan secara teoretis pada skala $(-\infty, +\infty)$. Item tes yang memiliki daya pembeda negative. Baker & Kim (2017) mengungkapkan bahwa item dengan diskriminasi negatif terjadi dalam dua cara. *Pertama*, jawaban yang salah terhadap item dua pilihan (benar, salah) akan selalu memiliki parameter diskriminasi item negatif jika jawaban yang benar bernilai positif. *Kedua*, terkadang respons yang benar terhadap suatu item akan menghasilkan parameter diskriminasi item negatif. Item dengan daya pembeda bernilai negatif sebaiknya dibuang atau direvisi, karena ada sesuatu yang salah dengan butir tersebut (seperti salah ketik), jika kemungkinan menjawab dengan benar berkurang seiring dengan meningkatnya kemampuan peserta tes. Juga, tidak biasa untuk mendapatkan nilai yang lebih besar dari +2. Oleh karena itu, kisaran yang biasa digunakan untuk parameter diskriminasi butir adalah $0 \leq a \leq 2$ (Hambleton, et.al., 1991). Baker (2001) menjabarkan rentang nilai daya beda butir tes sebagai berikut:

Tabel 5. Rentang nilai daya beda butir tes

Label Verbal	Rentang Nilai
Tidak ada	0
Sangat rendah	0,1 – 0,34
Rendah	0,35 – 0,64
Sedang	0,65 – 1,34
Tinggi	1,35 – 1,69
Sangat tinggi	>1,70
Sempurna	+ tidak terbatas

Hal ini menyiratkan bahwa nilai yang rendah akan menghasilkan kurva karakteristik butir yang naik secara bertahap sebagai fungsi kemampuan, tetapi nilai yang tinggi (daya pembeda) akan menghasilkan kurva karakteristik butir yang cukup curam. Tingkat informasi yang besar yang disampaikan suatu item tentang kemampuan responden atau peserta tes tercermin dalam daya pembedanya yang kuat. Akibatnya, nilai a menunjukkan seberapa banyak informasi tentang tingkat kemampuan peserta tes yang dapat diperoleh dari suatu item.

Estimasi parameter butir mode 2PL dilakukan dengan estimasi parameter butir untuk 1 dan 3PL, yaitu dengan menggunakan fungsi likelihood,

$$L(\theta, a, b; u) = \prod_i \prod_j P_i(\theta_j; a_i, b_i)^{u_{ij}} [1 - P_i(\theta_j; a_i, b_i)]^{1-u_{ij}}$$

Fungsi likelihood pada persamaan di atas, dapat dimaksimalkan dengan *Joint Maximum Likelihood* (JML), *Marginal Maximum Likelihood* (MML) atau CML (*Conditional Maximum Likelihood*).

a. Joint Maximum Likelihood (JML)

Metode JML mengestimasi secara bersamaan parameter orang (kemampuan peserta tes) dengan butir tes. Asumsi penggunaan metode kebolehjadian gabungan maksimum (JML) adalah pola skor pada peserta tes yang berbeda adalah independen. Fungsi kebolehjadian gabungan dirumuskan sebagai (Hessen, 2009).

$$L^J(\theta_1, \dots, \theta_n) = \prod_{v=1}^n \prod_{i=1}^k P_i(\theta_v)^{x_{vi}} Q_i(\theta_v)^{1-x_{vi}}$$

Di mana n = jumlah sampel (peserta tes) dan k adalah jumlah butir.

Kelebihan:

- Menarik secara konseptual dan secara komputasi kurang intensif
- Bebas distribusi

Kelemahan:

- Tidak ada estimasi parameter orang untuk skor nol dan skor sempurna
- Estimasi parameter tidak konsisten
- Sering gagal (tidak bisa digunakan) pada model 3PL.

b. Marginal Maximum Likelihood (MML)

Estimasi kemungkinan maksimum marjinal (MML), parameter orang (kemampuan peserta tes) dihilangkan dari proses estimasi dengan integrasi distribusi parameter orang. Maka pola respons peserta tes yang mungkin dari k butir dikotomus adalah 2^k . Peluang marginal pola jawaban $x_p = (x_{1p}, \dots, x_{kp})$ adalah.

$$P(x_p) = \int \left\{ \prod_{i=1}^k P_i(\theta)^{x_{ip}} Q_i(\theta)^{1-x_{ip}} \right\} g(\theta) d\theta$$

Sehingga fungsi kebolehdian marginal dapat dituliskan sebagai berikut.

$$L^M = \prod_{p=1}^{2^k} P(x_p)^{n_p}$$

Jika persamaan tersebut digunakan untuk mengestimasi parameter butir IRT model 2 PL, rumus matematika untuk fungsi kebolehdian maksimum marginal dapat dituliskan sebagai berikut.

$$L^M(a, b) = \prod_{p=1}^{2^k} \left[\int \left\{ \prod_{i=1}^k \frac{e^{x_{ip}a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}} \right\} g(\theta) d\theta \right]^{n_p}$$

Rata-rata dari distribusi ini ditetapkan sama dengan 0,0. Standar deviasi sama dengan 1,0. Hal ini berlaku juga untuk model 1 PL dan 3PL, model yang ditentukan pada skala interval. Jadi, untuk analisis Rasch, kita mulai dengan model 1PL (yaitu, model 2PL dengan parameter diskriminasi konstan). Estimasi parameter kesulitan MML akhir ditransformasikan ke skala dengan parameter diskriminasi sama dengan 1.

Kelebihan:

- Estimasi parameter butir konsisten
- Estimasi parameter orang (kemampuan peserta tes) untuk skor nol dan skor sempurna tercakup.

Kelemahan : tidak bebas distribusi (Hessen, 2009)

c. Conditional Maximum Likelihood (CML)

Estimasi kemungkinan maksimum bersyarat (CML), parameter orang dihilangkan dari proses estimasi dengan

pengkondisian pada skor total. Metode ini tersedia untuk model logistik dengan hanya parameter kesulitan. Jadi, CML dimungkinkan dengan model Rasch. Andersen (1983), Verhelst dan Glas, (1995) sebagaimana dikutip van der Linden & Hambleton (1997) menunjukkan bahwa CML juga dimungkinkan ketika dalam model dua parameter kemiringan diasumsikan diketahui.

Dalam kasus umum untuk sejumlah n item, banyaknya jawaban benar s_i untuk butir i ($i = 1, \dots, n$) didasarkan pada kelompok N orang dengan $0 < t < n$ jawaban benar. Kebolehjadian bersyarat sama dengan.

$$L_{CML} = C \frac{\prod_{i=1}^n \varepsilon_i^{s_i}}{\prod_{p=1}^N \gamma_{t_p}(\varepsilon)}$$

C adalah faktor yang tidak bergantung pada parameter item dan orang. Persamaan estimasi untuk estimasi ε_i adalah

$$s_i = \sum_{t=1}^{n-1} N_t P(x_i = 1|t) = \sum_{t=1}^{n-1} N_t \frac{\varepsilon_i \gamma_{t-1}^{(i)}(\varepsilon)}{\gamma_t(\varepsilon)}$$

Di mana N_t menunjukkan jumlah orang dengan skor t ($t = 1, \dots, n-1$). Jadi, dalam CML skor item s_i ditetapkan sama dengan skor item yang diharapkan berdasarkan probabilitas bersyarat yang benar untuk berbagai skor total. Persamaan estimasi untuk parameter item i ($i = 1, \dots, n$) diselesaikan secara iteratif di bawah satu pembatasan yang diperlukan untuk memperbaiki skala laten.

Kelebihan :

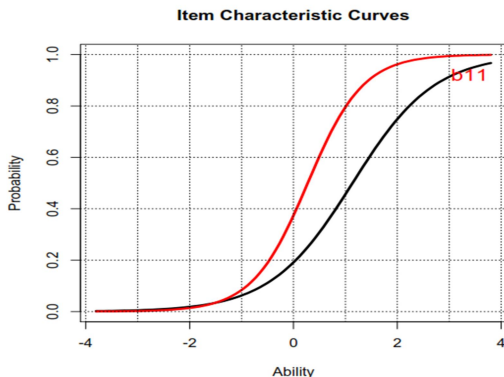
- Bebas distribusi
- Hasil estimasi parameter butir konsisten

Kelemahan: tidak ada estimasi parameter orang untuk skor nol dan skor sempurna (benar semua) (Hessen, 2009)

2. Kurva Karakteristik Butir (ICC) dan Informasi Butir (IIC)

Gambar berikut menunjukkan fungsi respons butir dari dua butir (butir 7 dan butir 11) yang memiliki nilai daya beda yang berbeda. Nilai a untuk butir 7 = 1,267 ($a_7 = 1,267$) dan nilai a untuk butir 11 = 1,876 ($a_{11} = 1,876$). Bagi kedua butir ini, nilai tingkat

kesulitan $b_7 = 1,137$ sedangkan $b_{11} = 0,275$ Perhatikan bahwa nilai daya beda a menentukan kemiringan fungsi respons butir; saat a bertambah, fungsi respons butir menjadi lebih curam dan butir tersebut menjadi butir dengan daya beda yang semakin tinggi. Dalam situasi ini, butir 7 memiliki daya beda yang lebih rendah dari butir 11, sehingga butir 11 memberikan lebih banyak informasi mengenai tingkat kemampuan peserta tes dibanding butir 7.



Gambar 18. Kurva Karakteristik Butir Model 2PL

Pada IRT, indeks keandalan (koefisien reliabilitas) tes dinyatakan dengan fungsi informasi tes, yaitu merupakan penjumlahan fungsi informasi semua butir tes (Mardapi, 2012). Fungsi informasi memberitahu kita tentang seberapa baik tingkat kemampuan peserta tes yang sedang diestimasi. Perlu diketahui bahwa, fungsi informasi tidak bergantung pada distribusi skala kemampuan peserta tes.

Jumlah informasi yang dihasilkan oleh tes pada tingkat kemampuan tertentu dapat diperoleh, karena tes digunakan untuk mengestimasi kemampuan peserta tes. Sebuah tes adalah kumpulan satu set item; oleh karena itu, informasi tes pada tingkat kemampuan yang diberikan hanyalah jumlah dari informasi item pada tingkat itu (Baker & Kim, 2017). Sehingga Total informasi untuk setiap item tes mewakili jumlah informasi pada tes (Hambleton & Swaminathan, 1985):

$$I(\theta) = \sum_{j=1}^J I_j(\theta)$$

$$I_j(\theta) = \sum_{j=1}^J \frac{P'_i(\theta)^2}{P_i(\theta)Q_i(\theta)}$$

Keterangan:

- $I(\theta)$: adalah jumlah informasi tes pada tingkat kemampuan
 $I_j(\theta)$: adalah banyaknya informasi butir j pada tingkat kemampuan
 J : Jumlah butir pada tes
 $P'_i(\theta)$: *derivative* dari $P_i(\theta)$

Untuk informasi butir mode 2PL, dapat dituliskan sebagai berikut.

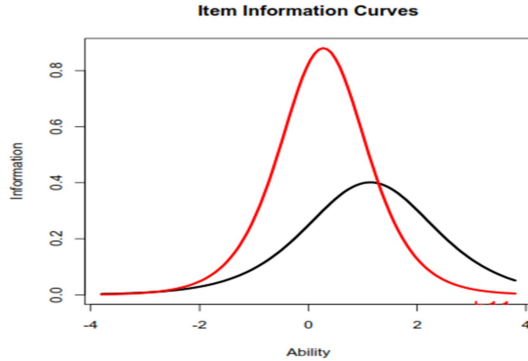
$$I_j(\theta) = a_j^2 P_j(\theta) Q_j(\theta)$$

Keterangan:

- a_j = daya pembeda butir ke j.
 $P_j(\theta)$ = Peluang seseorang dengan kemampuan θ menjawab butir tes j dengan benar $[1/(1+\exp(-L_j))]$.
 $L_j = a_j(\theta - b_j)$
 $Q_j(\theta) = 1 - P_j(\theta)$

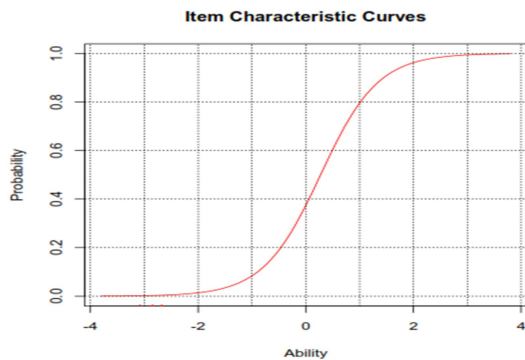
Lebih lanjut, Mardapi (2012) menyatakan bahwa jika item tes memenuhi model, pengembang tes akan memilih item dengan fungsi informasi terbesar, dengan demikian fungsi informasi tes akan bermanfaat. Standard Error of Measurement (Standard Error, SE) dan fungsi informasi memiliki hubungan kuadratik, artinya semakin besar fungsi informasi maka semakin kecil SE, dan semakin kecil fungsi informasi maka semakin besar SE.. Apabila nilai fungsi informasi dinyatakan dengan $I_i(\theta)$ maka nilai estimasi SE θ adalah (Hambleton & Swaminathan, 1985):

$$SE(\theta) = \frac{1}{\sqrt{I_i(\theta)}}$$



Gambar 19. Kurva fungsi informasi daya beda butir 7 dan butir 11

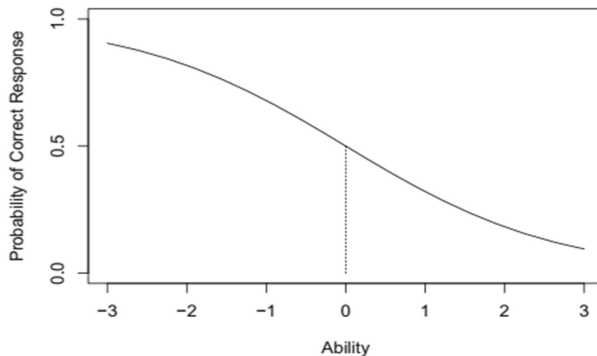
Model 2-PL dan 1-PL merupakan bagian dari model 3-PL. Model 2-PL merupakan kasus khusus dari 3-PL, yakni ketika $c = 0$. Model 1-PL merupakan kasus khusus dari 2-PL, yakni ketika $a = 1$ atau a merupakan tetapan untuk keseluruhan butir tes. Kurva karakteristik butir yang dianalisis dengan model 2-PL yang memiliki $a_{11} = 1,876$; $b_{11} = 0,275$ ditunjukkan oleh Gambar 20 berikut.



Gambar 20. ICC model 2PL ($a=1,876$ dan $b = 0,275$)

Pada gambar 20 di atas, dan hampir sebagian butir tes memiliki daya beda positif, jarang sekali kita menemukan daya beda butir tes yang negative. Menurut Baker & Kim (2017), item dengan diskriminasi negatif terjadi dalam dua cara. Pertama, jawaban yang salah terhadap item pilihan dua akan selalu memiliki parameter

diskriminasi item negatif jika jawaban yang benar bernilai positif. Kedua, terkadang respons yang benar terhadap suatu item akan menghasilkan parameter diskriminasi item negatif. Ini memberitahu Anda bahwa ada sesuatu yang salah dengan item tersebut. Seperti digambarkan berikut.



Gambar 21. Butir tes dengan $b = 0$ dan $a = -0,75$

Kesalahan tersebut, bisa jadi karena penulisannya kurang baik atau ada beberapa kesalahan informasi yang lazim di kalangan siswa berkemampuan tinggi. Bagaimanapun, ini adalah peringatan bahwa item tersebut membutuhkan perhatian. Untuk sebagian besar topik teori respons item yang menarik, nilai parameter diskriminasi item akan positif.

3. Uji Kecocokan Model (*Goodness of Fit*)

Pendekatan penilaian kecocokan dengan model dua dan tiga parameter sering melibatkan pengumpulan berbagai bukti tentang kecocokan model, termasuk uji statistik dan kemudian membuat penilaian yang tepat tentang kecocokan model dan kegunaan model dengan kumpulan data tertentu (Hambleton, 1989). Pendekatan ini sering mencakup penerapan beberapa model untuk menguji data dan membandingkan hasil dan konsekuensi praktis dari setiap perbedaan.

Biasanya dilakukan tiga jenis analisis, yaitu: (1) pemeriksaan asumsi model, (2) pemeriksaan invariansi parameter model, dan (3) pemeriksaan prediksi dari model. Ide dasarnya adalah membiasakan diri dengan model (dan model lain yang mungkin), uji kecocokan

data dengan model, dan konsekuensi dari ketidakcocokan model, sebelum membuat keputusan tentang kegunaan model tertentu.

Pemeriksaan asumsi model dilakukan dengan analisis faktor (*exploratory factor analysis_EFA dan confirmatory factor analysis-CFA*) yaitu pembuktian apakah asumsi unidimensi valid atau tidak. Biasanya pengecekan dilakukan dengan melihat: (1) nilai eigen dari matriks korelasi antar item, (2) residu setelah memasang model analitik faktor linier atau nonlinier satu faktor, atau (3) statistik butir IRT yang diperoleh dalam tes total dan dalam subset item dari tes yang berpotensi mengukur kemampuan kedua. Pembahasan mengenai analisis faktor dapat dilihat pada bagian berikutnya .

Pemeriksaan invariansi parameter model dapat dilakukan dengan cara membuktikan bahwa model yang diestimasi dapat diterima, di mana estimasi kemampuan peserta tes harus sama (terlepas dari kesalahan pengukuran) dari sampel butir yang berbeda dalam tes (misalnya, mudah versus sulit). Estimasi parameter item harus hampir sama (kecuali untuk kesalahan pengambilan sampel) dari sampel peserta ujian yang berbeda dari populasi peserta ujian yang menjadi tujuan pengujian (misalnya, pria dan wanita, desa atau kota dan sebagainya). Pemeriksaan invariansi kemampuan dan parameter butir memberikan informasi berharga mengenai kecocokan model (Hambleton, Swaminathan dan Rogers, 1991).

Prediksi dari model juga dapat diperiksa/diuji dengan menggunakan analisis residu atau residu standar untuk butir dan juga orang (kemampuan) biasanya dilakukan melalui uji plot residu (Hambleton, 1989). Uji statistik penyimpangan dalam data dari prediksi model juga dapat dilakukan. Perbandingan residual dari penggunaan beberapa model sangat membantu dalam menentukan kecocokan model. Studi lain yang mungkin dilakukan termasuk perbandingan distribusi skor prediksi dengan distribusi aktual (dari beberapa model) dan konsekuensi dari model fit (seperti peringkat kemampuan peserta tes di bawah model yang berbeda).

D. Model 3 Parameter Logistik (3PL)

Pada model logistik 3 parameter (3 PL), parameter tebakan (guessing) untuk respons butir oleh peserta tes menjadi salah satu perhatian. Tidak sedikit peserta tes mendapatkan skor yang lebih tinggi (mampu menjawab soal yang sulit) dari kemampuan yang dimilikinya. Pembaca sepertinya tidak terlalu asing dengan istilah menebak (guessing) dalam menjawab soal. Pada model 3 PL,

tindakan menebak (guessing) dalam menjawab soal ini dilambangkan dengan c (parameter c).

Model 3 parameter logistik (3 PL) dengan demikian melibatkan parameter tingkat kesulitan (b_i), daya pembeda (a_i) dan tebakan (c_i). Parameter c (tebakan) menunjukkan bahwa peserta ujian dengan kemampuan terendah dan tertinggi mungkin memiliki probabilitas yang sama untuk menjawab butir soal dengan benar yaitu dengan cara menebak. Parameter c (tebakan) memiliki rentang nilai teoretis $0 \leq c \leq +1,0$. Namun dalam prakteknya, menurut Baker & Kim (2017) bahwa nilai c di atas 0,35 tidak dapat dipertimbangkan sehingga kisaran nilai parameter c (tebakan) adalah $0 \leq c \leq +0,35$.

Adapun persamaan matematika model 3 parameter logistik dapat ditulis (van der Linden & Hambleton, 1997).

$$P_i(\theta) = c_i + (1 - c_i) \frac{1}{1 + \exp\{-a_i(\theta - b_i)\}}$$

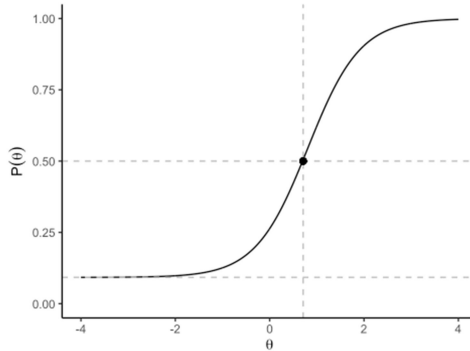
Atau

$$P_i(\theta) = c_i + \left\{ (1 - c_i) \frac{e^{D_i a_i (\theta - b_i)}}{1 + e^{D_i a_i (\theta - b_i)}} \right\}, i = 1, 2, 3 \dots n$$

Keterangan:

- $P_i(\theta)$ = peluang seseorang dengan kemampuan(θ) menjawab butir soal ke- i dengan benar.
- b_i = parameter tingkat kesulitan untuk butir soal ke - i
- e = bilangan pokok logaritma natural, yang nilainya adalah 2,718, jika dibulatkan ketiga angka dibelakang koma.
- D = faktor skala, yang nilainya diambil sebesar 1,7. yaitu simpangan baku distribusi logistik yang paling mendekati distribusi normal.
- c_i = parameter tebakan (*pseudo guessing*) pada butir soal ke- i

Gambar 22 di bawah ini menunjukkan kurva karakteristik item (ICC) untuk IRT mode 3PL.



Gambar 22. Kurva Karakteristik Butir Model 3PL

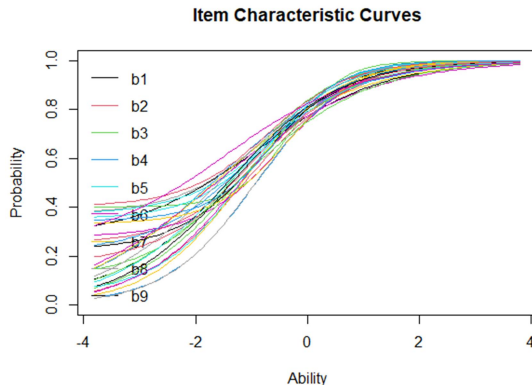
Sumbu x menunjukkan kemampuan laten (θ) mulai dari -4 sampai 4, dengan 0 sebagai kemampuan rata-rata pada populasi yang diteliti. Sumbu y menunjukkan probabilitas menjawab item dengan benar. ICC dengan demikian mewakili probabilitas menjawab item dengan benar pada kemampuan laten tertentu.

1. Estimasi parameter butir dan orang

Estimasi parameter butir untuk model 3PL menggunakan fungsi likelihood berikut.

$$L(\theta, a, b, c; u) = \prod_i \prod_j P_i(\theta_j; a_i, b_i, c_i)^{u_{ij}} [1 - P_i(\theta_j; a_i, b_i, c_i)]^{1-u_{ij}}$$

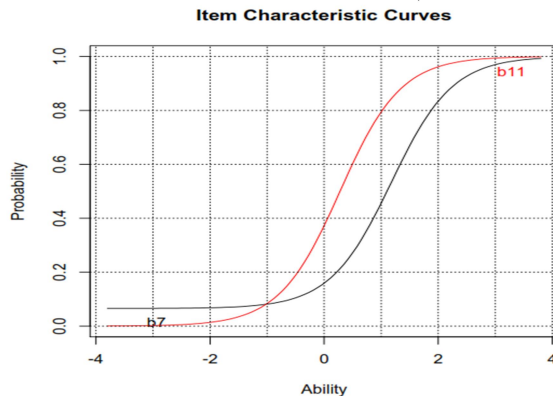
2. Kurva Karakteristik Butir (ICC) dan Kurva Informasi Butir (IIC)



Gambar 23. Item Characteristic Curves (ICC) IRT model 3 PL

Pada gambar 23 tersebut di atas, terlihat parameter tebakan memberikan (kemungkinan/peluang) bukan nol lebih rendah dari asimtot pada kurva karakteristik butir dan mewakili probabilitas peserta ujian dengan kemampuan rendah menjawab item dengan benar. Parameter tebakan dimasukkan ke dalam model untuk memperhitungkan kinerja di ujung bawah kontinum kemampuan, di mana menebak merupakan faktor dalam kinerja tes pada item tes pilihan ganda. Menurut Hambleton, et.al. (1991), parameter tebakan mengasumsikan nilai yang lebih kecil dari nilai yang akan dihasilkan jika peserta ujian menebak secara acak item tersebut.

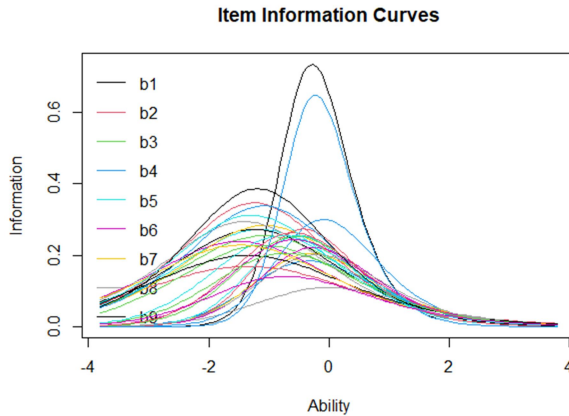
Indeks tebakan semu (*pseudoguessing*) pada model 3-PL, memungkinkan peserta yang memiliki kemampuan rendah mempunyai peluang untuk menjawab butir soal dengan benar. Nilai c merefleksikan nilai terendah dari fungsi respons butir saat kemampuan menjadi sangat rendah (dikenal sebagai asimtot fungsi respons butir). Jadi, jika $c = 0,2$, maka peluang menjawab benar bagi individu dengan kemampuan yang sangat rendah ialah 0,2. Karena nilai c merefleksikan hasil perilaku menebak jawaban, hal ini disebut dengan parameter *pseudo-guessing*. Indeks tebakan semu pada tes pilihan ganda terletak di sekitar seperbanyaknya pilihan jawaban ($c = 1/k$). Misalnya pada tes dengan pilihan 4 jawaban, maka nilai c_i terletak di sekitar $\frac{1}{4}$ atau 0,25 (Retnawati, 2014).



Gambar 24. Perilaku menebak di butir 7

Gambar 24 berikut menunjukkan fungsi respons butir dari dua butir berbeda (butir 7 dan butir 11) yang memiliki nilai c yang berbeda; nilai c untuk butir 7 ialah 0,065 ($c_7 = 0,065$) dan nilai c

untuk butir 11 ialah ($c_{11} = 0$). Namun, nilai $c_7 = 0,065$ untuk butir 7 menyebabkan batas bawah fungsi respons butir 7 menjadi lebih tinggi daripada butir 11, hal ini mengindikasikan adanya perilaku menebak di butir 7.



Gambar 25. Kurva fungsi informasi

Fungsi informasi butir untuk model IRT 3PL, dapat dituliskan sebagai berikut.

$$I_j(\theta) = a_j^2 \left[\frac{Q_j(\theta)}{P_j(\theta)} \right] \left[\frac{(P_j(\theta) - c_j)^2}{(1 - c_j)^2} \right]$$

Keterangan:

$$P_j(\theta) = c_j + (1 - c_j) \frac{1}{1 + \exp(-L_j)}$$

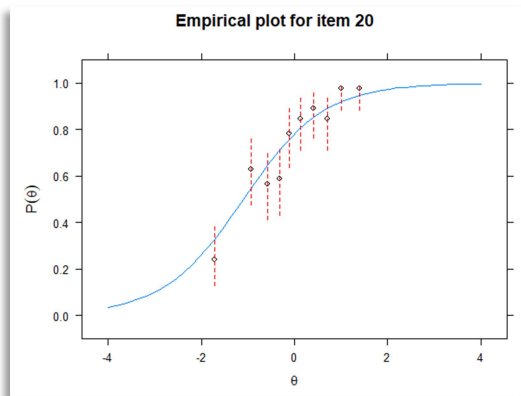
$$L_j = a_j(\theta - b_j)$$

$$Q_j(\theta) = 1 - P_j(\theta)$$

3. Analisis Kecocokan Model data dengan Model

Analisis kecocokan mode 3PL sama dengan model 2PL dan 4PL. Untuk uji kecocokan butir dengan model digunakan beberapa metode di antaranya adalah Chi-kuadrat M2. Dengan ketentuan, jika nilai p pada chi-kuadrat M2 lebih besar dari nilai $\alpha = 0,05$, maka semua butir diyatakan cocok dengan model. Pada output chi-kuadrat M2 juga tersedia nilai RMSEA. Jika nilai RMSEA $< 0,05$, dapat dikatakan butir tersebut cocok dengan mode.

Selain itu, uji kecocokan butir dengan model juga dapat dilakukan dengan membandingkan grafik ICC berbasis model dengan ICC empiris, sebagai contoh pada salah satu butir nomor 26 berikut.



Gambar 26. Kurva empiris item nomor 20.

Pada grafik ICC berbasis model (grafik warna biru muda) dan ICC empiris (titik-titik dan garis putus-putus berwarna merah) terlihat bahwa ICC empiris mendekati atau berada pada kisaran garis ICC berbasis model. Dengan demikian butir nomor 20 tersebut dapat dikatakan cocok dengan model.

Untuk uji kecocokan orang (kemampuan orang) dengan model dapat menggunakan metode MLE (*Maximum Likelihood Estimation*), EAP dan MAP.

E. Model 4 Parameter Logistik (4PL)

IRT model 4 parameter logistik adalah penyempurnaan dari model-model sebelumnya (1PL, 2PL, dan 3PL), yaitu dengan memasukkan parameter “kecerobohan” atau “carelessness” sebagai parameter ke empat dan biasanya dilambangkan dengan d . IRT model 4 PL dirumuskan sebagai berikut.

$$P(X_{ij} = 1|\theta_j) = g_i + (u_i - g_i) \frac{\exp[a_i(\theta_j - b_i)]}{1 + \exp[a_i(\theta_j - b_i)]}$$

Keterangan:

$P(X_{ij} = 1|\theta_j)$: Peluang peserta tes dengan kemampuan j menjawab benar butir ke i

- g_i : parameter tebakan butir ke $-i$
- u_i : parameter “kecerobohan” atau “*carelessness*” butir ke- i
- a_i : parameter daya pembeda butir ke- i
- b_i : parameter tingkat kesulitan butir ke- i
- D : faktor skala, yang nilainya diambil sebesar 1,7, yaitu simpangan baku distribusi logistik yang paling mendekati distribusi normal.
- e : bilangan pokok logaritma natural, yang nilainya adalah 2,718, jika dibulatkan ketiga angka dibelakang koma

1. **Estimasi Parameter Butir dan Orang**

Contoh kasus:

Seorang peneliti ingin menganalisis tes yang digunakan sebagai instrument dalam penelitiannya. Sebut saja tes kemampuan literasi matematika siswa MTs. Muallimat NWDI Pancor. Sebelum digunakan sebagai alat pengumpul data, tes tersebut diujicobakan kepada 460 orang siswa MTs NWDI Pancor, jumlah butir sebanyak 30 butir. Peneliti tersebut hendak menggunakan IRT model 4 parameter logistik.

Pertanyaannya:

- 1) Bagaimana karakteristik butir tes yang digunakan?
- 2) Bagaimana karakteristik kemampuan peserta tes yang dijadikan sampel dalam penelitian ini?
- 3) Apakah butir tes yang diujicobakan cocok atau sesuai dengan model 4PL?
- 4) Apakah pola jawaban peserta tes cocok atau sesuai dengan model 4PL?

Solusi:

Estimasi parameter butir dan parameter kemampuan dihitung menggunakan rumus model 4 parameter di atas, metode estimasi yang digunakan metode EAP. Analisis data menggunakan aplikasi R (misalnya) dengan paket “mirt”.

1) Estimasi parameter butir.

Hasil analisis data uji coba tes kemampuan literasi matematika diperoleh sebagai berikut (untuk langkah dan perhitungan lengkap, bisa dilihat pada bagian analisis butir IRT model 4PL)

Tabel 6. Hasil analisis estimasi parameter butir model 4PL

Butir	a.	b.	g.	u.
b1	2,447	-0,394	0,316	1,000
b2	1,595	-1,388	0,010	0,874
b3	1,492	-1,027	0,077	0,925
...
b28	2,139	-0,502	0,306	1,000
b29	1,169	-1,242	0,009	1,000
b30	0,988	-1,162	0,239	0,999

Berdasarkan tabel 6 hasil analisis estimasi parameter butir tersebut di atas, diketahui bahwa butir 1 memiliki daya pembeda ($a_1 = 2,447$), tingkat kesulitan ($b_1 = -0,394$), parameter tebakan ($g_1 = 0,316$) dan parameter “kecerobohan” atau “carelessness” ($u_1 = 1,000$) dan seterusnya.

2) Estimasi parameter orang (kemampuan laten)

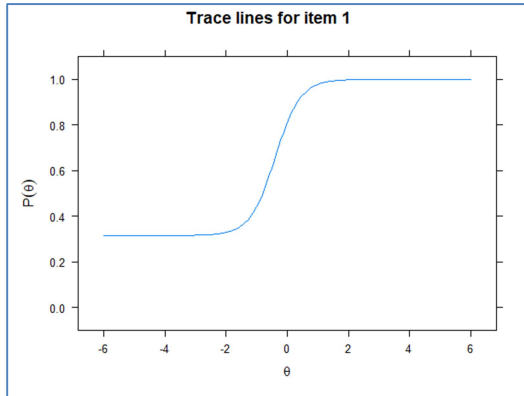
Ringkasan hasil estimasi parameter orang (kemampuan laten) menggunakan script R dapat dilihat pada tabel 6 berikut.

Tabel 7. Hasil analisis estimasi kemampuan orang (peserta tes)

Kode Siswa	F1	SE_F1	Kode Siswa	F1	SE_F1
A8S1	0,643	0,491	A8S231	-0,792	0,286
A8S2	-0,797	0,318	A8S232	0,186	0,415
...
A8S228	1,250	0,644	A8S458	0,576	0,423
A8S229	-0,398	0,340	A8S459	-0,397	0,371
A8S230	0,051	0,368	A8S460	-0,355	0,299

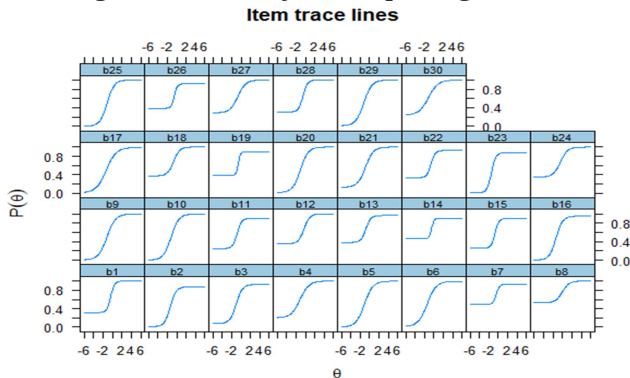
Dari hasil estimasi parameter kemampuan yang ditunjukkan pada tabel 7 di atas, terlihat bahwa peserta dengan nomor urut 1 atau kode A8S1 memiliki kemampuan laten (θ) = 0,643 dengan kesalahan baku (SE) = 0,491; peserta tes dengan kode A8S460 memiliki kemampuan (θ) = -0,355 dengan kesalahan baku (SE) = 0,299.

Untuk mempermudah pembacaan laporan penelitiannya, peneliti juga menampilkan grafik (ICC, IIC dan TRF) berikut.



Gambar 27. Kurva karakteristik butir nomor 1

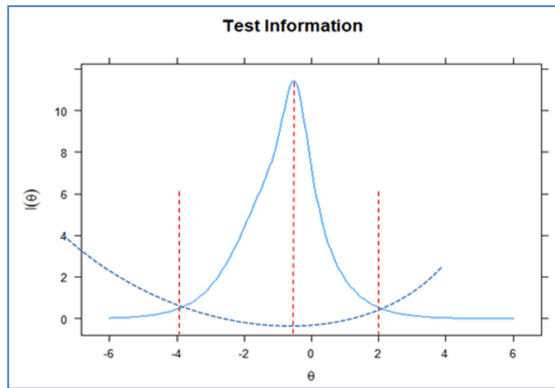
Gambar 27 di atas, menjelaskan grafik ICC untuk model 4PL, daya pembeda ($a_1 = 2,447$), tingkat kesulitan ($b_1 = -0,394$), parameter tebakan ($g_1 = 0,316$) dan parameter “kecerobohan” atau “carelessness” ($u_1 = 1,000$). Akan tetapi, menampilkan grafik ICC satu per satu pada badan naskah laporan penelitian, tentunya kurang efektif dan efisien. Dibutuhkan tabel matrik ICC yang memuat ICC semua butir soal, sebagaimana ditunjukkan pada gambar 28 berikut.



Gambar 28. Trace line butir soal atau ICC secara keseluruhan

IRT tidak hanya berbicara mengenai kurva karakteristik butir soal, di dalamnya juga kita bisa mendapatkan grafik informasi tes, yaitu sejauh mana tes menggambarkan atau

memberikan informasi yang dibutuhkan oleh pembuat tes (alat ukur) guna menjawab permasalahan penelitiannya.



Gambar 29. Kurva informasi tes.

Gambar 29 di atas menunjukkan bahwa tes atau alat ukur yang digunakan memberikan informasi yang cukup mengenai kemampuan peserta tes, terutama sekali informasi mengenai kemampuan peserta yang berada pada rentang $\theta = -4$ dan $\theta = +2$

3) Analisis kecocokan butir dengan model

Luaran hasil analisis kecocokan butir dengan model dapat dilihat pada tabel 8 berikut.

Tabel 8. Output hasil analisis kecocokan butir dengan model.

Butir	S_X2	df.S_X2	RMSEA.S_X2	p.S_X2
b1	18,130	16,000	0,017	0,316
b2	18,734	17,000	0,015	0,344
b3	28,135	17,000	0,038	0,043
...
b28	15,130	17,000	0,000	0,586
b29	9,361	16,000	0,000	0,898
b30	10,217	18,000	0,000	0,925

Pada output fungsi “itemfit()” model 4PL di atas, terlihat bahwa hampir semua butir nilai $p.S_X2 \geq$ dari 0,05 dan dengan koreksi Bonferroni di bawah tingkat kesalahan Tipe I sebesar 0,05, tidak ada item yang ditandai secara statistic (tanda bintang). Hal ini menunjukkan bahwa semua butir cocok atau sesuai dengan mode 4PL

4) Analisis kecocokan orang dengan model

Analisis kecocokan orang (kemampuan peserta tes) dengan model, dilihat dari nilai Zh yang dihasilkan, tabel 8 berikut adalah contoh bagaimana melihat kecocokan kemampuan peserta tes dengan model dari pola respons yang diberikannya.

Tabel 9. Output hasil kecocokan orang (peserta tes) dengan model

Resp.	Outfit	z.outfit	Infit	z.infit	Zh
1	0,984133	0,129173	0,957037	0,058021	0,068829
2	1,027715	0,333931	1,020372	0,261672	-0,25791
3	1,029905	0,267183	1,034398	0,307687	-0,25461
198	0,363922	-0,75758	0,542128	-0,64676	0,766184
199	0,996544	-0,06086	0,996523	-0,06314	0,077232
200	1,22015	0,569059	1,297355	0,722277	-0,57512

Statistik kecocokan orang dengan model (Iz) dihitung untuk setiap orang di bawah "Zh" menggunakan fungsi "personfit()". Jika nilai "Zh" lebih kecil dari -3 atau -2, maka hal ini mengindikasikan adanya potensi pola respons yang menyimpang atau tidak cocok dengan model. Pada output “personfit()” di atas, terlihat bahwa semua orang atau peserta memiliki pola jawaban yang sesuai atau cocok dengan model 4PL.

5 Analisis Butir Tes (Skor Dikotomis)

“Hidup itu seperti mengendarai sepeda. Untuk menjaga keseimbangan Anda, Anda harus terus bergerak.”

-Albert Einstein-

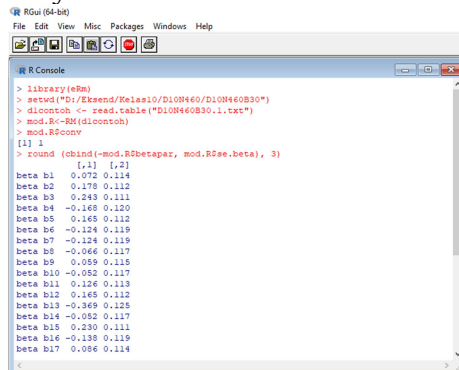
A. Analisis Butir Tes Model Rasch

1. Analisis Karakteristik Tes Model Rasch dengan Paket “eRm”

Paket “eRm” menggunakan estimasi CML (*Conditional Maximum Likelihood*). Estimasi CML tidak memerlukan spesifikasi distribusi sifat laten (θ), keterbatasan model identifikasi (ID) biasanya dihasilkan pada parameter butir. Kendala model ID yang paling populer adalah mengatur rata-rata semua kesulitan item sama dengan 0 yaitu $\sum b_i = 0$, sehingga tidak diperlukan spesifikasi tambahan dalam kode. Analisis karakteristik butir terdiri dari: 1) estimasi parameter butir, 2) estimasi parameter kemampuan orang (peserta tes), 3) uji kecocokan data dengan model.

a. Estimasi Parameter Butir

Analisis karakteristik butir bisa dilakukan dengan aplikasi R atau Rstudio. Rstudio memberikan ruang kepada pemakai untuk melakukan editing script (perintah) sedangkan di R tidak, berikut gambar tampilan layar di R dan Rstudio.

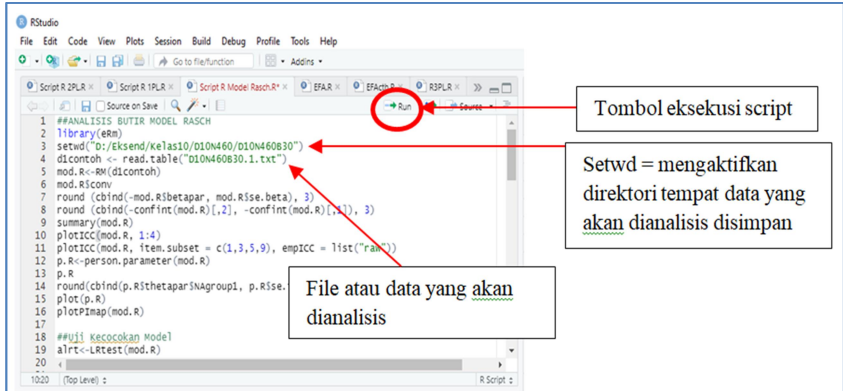


```
RStudio (64-bit)
File Edit View Misc Packages Windows Help

R R Console
> library(eRm)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> dicontoh <- read.table("D10N460B30.1.txt")
> mod.R<-RM(dicontoh)
> mod.R$conv
[1] 1
> round(cbind(mod.R$betapar, mod.R$se.beta), 3)
      [,1] [,2]
beta b1  0.072 0.114
beta b2  0.178 0.112
beta b3  0.245 0.111
beta b4 -0.168 0.120
beta b5  0.168 0.112
beta b6 -0.124 0.119
beta b7 -0.124 0.119
beta b8 -0.066 0.117
beta b9  0.059 0.115
beta b10 -0.052 0.117
beta b11  0.126 0.113
beta b12  0.148 0.112
beta b13 -0.369 0.125
beta b14 -0.052 0.117
beta b15  0.230 0.111
beta b16 -0.138 0.115
beta b17  0.086 0.114
```

Gambar 30. Analisis Butir pada Layar console R

Layar (screen) di R hanya menampilkan layar Console, untuk itu perlu kehati-hatian dan ketelitian pada saat menuliskan perintah. Berbeda titik (.) koma (,) bisa menyebabkan eksekusi perintah tidak jalan. Berbeda dengan Rstudio yang menyediakan layar untuk mengetikkan perintah (script) yang bisa diedit sebelum/setelah dieksekusi.



Gambar 31. layar script editor pada Rstudio

Agar mempermudah pembacaan script dan hasil (*output*) di buku ini, script ditulis dengan warna biru sedangkan output/hasil berwarna hitam dan tanda titik tiga antara nilai atau variabel merupakan jeda (dipotong untuk meringkas hasil). Selanjutnya script utuh dan menyeluruh diberikan pada lampiran dengan judul yang sesuai.

Contoh: Guru Fisika di MTs Mu'allimat NWDI Pancor melakukan uji coba soal yang akan digunakan untuk Ujian Akhir Sekolah (UAS), soal diujicobakan di MTs Muallimin NWDI Pancor dengan jumlah sampel (peserta tes) sebanyak 460 orang siswa dengan jumlah butir soal (panjang tes) sebanyak 30 soal. Soal dibuat dalam bentuk pilihan ganda (*multiple choices*) model Rasch.

```

> install.packages("eRm")
> library(eRm)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> d1contoh <- read.table("D10N460B30.1.txt")
> mod.R<-RM(d1contoh)
> mod.R$conv
> round(cbind(-mod.R$betapar, mod.R$se.beta), 3)
> round(cbind(-confint(mod.R)[,2], -confint(mod.R)[,1]), 3)
> summary(mod.R)

```

Berikut penjelasan script dan output yang dihasilkan tersebut.

> library (eRm) adalah paket yang digunakan untuk mengestimasi parameter butir dan orang menggunakan mode Rasch, metode estimasi paket “eRm” adalah *Conditional Maximum Likelihood* (CML) atau kebolehdjian maksimum bersyarat.

```

> round (cbind(-mod.R$betapar, mod.R$se.beta), 3)
      [,1] [,2]
beta b1  0.072 0.114
beta b2  0.178 0.112
beta b3  0.243 0.111
...     ...   ...
beta b28 -0.024 0.116
beta b29 -0.052 0.117
beta b30 -0.228 0.121

```

Output di atas menunjukkan hasil estimasi kesulitan butir (\hat{b}_i) dan kesalahan baku (s.e.) untuk masing-masing butir (butir 1 sampai dengan 30). Script di atas maksudnya adalah kita meminta R dengan paket “eRm” mengestimasi tingkat kesulitan butir dan kesalahan baku yang disimpan dalam atribut mod.R\$betapar dan mod.R\$se.beta, dibulatkan ke tiga tempat desimal. Kolom angka pertama berjudul “[,1]” adalah estimasi kesulitan soal (\hat{b}_i) dan kolom angka kedua berjudul “[,2]” adalah kesalahan standar (s.e). Hasil analisis menunjukkan bahwa estimasi kesulitan butir nomor 1 (b_1) = 0,072, dengan kesalahan baku (se_1) = 0,114; (b_{28}) = -0,024 dengan se_{28} = 0,116 dan seterusnya.

```

> round (cbind(-confint(mod.R)[,2],
+             -confint(mod.R)[,1]), 3)
      [,1] [,2]
beta b1 -0.151 0.296
beta b2 -0.042 0.398
beta b3  0.025 0.460
...     ...   ...   ...
beta b28 -0.252 0.204
beta b29 -0.281 0.177
beta b30 -0.465 0.010

```

Fungsi “**confint()**” menyediakan interval kepercayaan 95% (CI) untuk argumen yang diberikan, dalam hal ini (\hat{b}_i) dari “mod.R” Kolom angka pertama adalah batas bawah (2,5% kuantil) dan kolom angka kedua adalah batas atas (97,5% kuantil). “**round()**” digunakan untuk memberikan hasil yang dibulatkan ke tiga desimal. Hasil item 1 sesuai dengan estimasi parameter item, plus dan minus nilai z-kritis (1,96 untuk 95% CI) dikalikan kesalahan baku (s.e): $0,072 \pm (1,96 * 0,114) = [0,296, -0,151]$.

Alternatif lain untuk mengestimasi parameter butir dan kesalahan bakunya dengan interval kepercayaan (CI) 95% yaitu dengan fungsi **summary()**. Bagian pertama dari output, "parameter kesulitan butir (kategori) (eta)" adalah parameterisasi kesulitan butir di bawah batasan $\sum b_i = 0$. Namun, hasil untuk butir pertama tidak ditampilkan, karena estimasi butir pertama dapat diperoleh dengan jumlah negatif dari semua estimasi kesulitan butir lainnya. Bagian kedua dari output adalah "parameter kemudahan butir (beta)" adalah parameterisasi kemudahan butir di bawah batasan $\sum b_i = 0$ yang sama. Perhatikan parameter kemudahan sama dengan parameter kesulitan yang sesuai yang diperoleh dari “ - mod.R\$betapar.”

```

> summary(mod.R)

Results of RM estimation:

Call: RM(X = d1contoh)

Conditional log-likelihood: -5658.481
Number of iterations: 11
Number of parameters: 29

Item (Category) Difficulty Parameters (eta): with 0.95 CI:
  Estimate Std. Error lower CI upper CI
b2      0.178      0.112   -0.042   0.398
b3      0.243      0.111    0.025   0.460
b4     -0.168      0.120   -0.402   0.067
...
b28    -0.024      0.116   -0.252   0.204
b29    -0.052      0.117   -0.281   0.177
b30    -0.228      0.121   -0.465   0.010

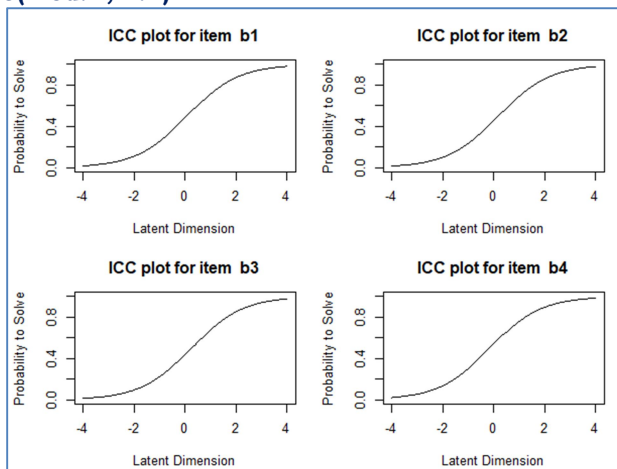
Item easiness Parameters (beta) with 0.95 CI:
  Estimate Std. Error lower CI upper CI
beta b1  -0.072      0.114   -0.296   0.151
beta b2  -0.178      0.112   -0.398   0.042
beta b3  -0.243      0.111   -0.460   -0.025
...
beta b28  0.024      0.116   -0.204   0.252
beta b29  0.052      0.117   -0.177   0.281
beta b30  0.228      0.121   -0.010   0.465

```

b. Kurva Karakteristik Butir

Persamaan IRF (*Item Response Function*) terdiri dari kurva karakteristik item (ICC) yang sesuai untuk setiap item. ICC adalah kurva estimasi (berbasis model). Untuk menguji kecocokan model secara grafis dilakukan dengan membandingkan estimasi kurva ICC dengan ICC empiris. Meskipun ICC empiris adalah istilah yang harus dibedakan dari ICC berbasis model, hal itu tidak sepenuhnya berbasis data karena ICC empiris juga didasarkan pada estimasi skor sifat laten atau merupakan bagian dari estimasi model. Jadi, peneliti, guru dan praktisi harus mempertimbangkan penggunaan plot perbandingan ICC empiris vs. ICC berbasis model ini sebagai pengujian secara grafis estimasi *Goodness of Fit* pada tingkat item. Berikut output ICC hasil analisis berbasis model dan estimasi ICC secara empiris.

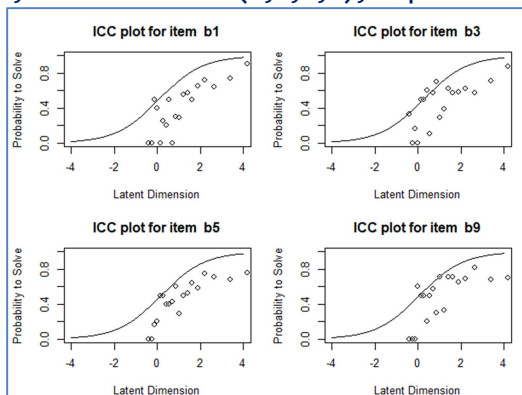
> plotICC(mod.R, 1:4)



Gambar 32. Grafik ICC berbasis model

Grafik ICC yang ditunjukkan pada gambar 32 di atas, merupakan grafik hasil analisis berbasis model untuk butir soal nomor 1 sampai dengan nomor 4. Hal ini belum cukup memberikan informasi, apakah model tersebut sesuai dengan data empiris ataukah tidak, oleh karena itu perlu dimunculkan ICC berdasarkan model empiris sebagaimana berikut.

> plotICC(mod.R, item.subset = c(1,3,5,9), empICC = list("raw"))



Gambar 33. Grafik ICC berbasis estimasi empiris

ICC berbasis model dan ICC empiris menunjukkan bahwa kemampuan laten peserta tes berada pada kisaran garis peluang menjawab benar, dengan demikian dapat dikatakan bahwa data sesuai atau cocok dengan model Rasch.

c. Estimasi Parameter Kemampuan (θ)

Untuk mengestimasi kemampuan peserta tes (θ) menggunakan paket "eRm" adalah dengan estimasi kebolehjadian maksimum (*Maximum Likelihood Estimation*). MLE tidak terbatas pada estimasi pola respons dari semua 0 dan 1 yaitu skor mentah yang sempurna (skor angka-benar). Untuk mengestimasi kemampuan laten peserta didik, "eRm" menggunakan teknik interpolasi yang disebut interpolasi spline.

```
> p.R<-person.parameter(mod.R)
> p.R
```

Person Parameters:

Raw Score	Estimate	Std.Error
6	-1.3927168159	0.4572397
7	-1.1952677212	0.4325156
8	-1.0165830509	0.4137515
...
28	2.6481740884	0.7324606
29	3.3771264936	1.0175934
30	4.1755171960	NA

Jika peneliti bermaksud mengestimasi kemampuan setiap peserta tes, perintah (script) berikut bisa digunakan dengan estimasi ML (*Maximum Likelihood*) setiap orang/peserta tes di kolom pertama dan kesalahan standarnya di kolom kedua. "**p.R\$thetapar**" memanggil estimasi skor sifat laten sebagai daftar, sedangkan "**p.R\$thetapar\$NAgrou1**" untuk estimasi skor sifat laten sebagai vektor nilai numerik. Karena fungsi "cbind()" hanya dapat menggabungkan vektor atau matriks, dan bukan daftar, maka digunakan "**p.R\$thetapar\$NAgrou1**". Setelah itu, untuk memunculkan hasil estimasi ML setiap peserta tes peneliti cukup mengetik "**coef(p.R)**".


```
> round(cbind(p.R$thetapar$NAgroup1, p.R$se.theta$NAgroup1), 3)
      [,1] [,2]
1      2.206 0.609
2      0.408 0.374
3      0.550 0.380
...     ...   ...
458    2.206 0.609
459    0.852 0.399
460    0.017 0.414
```

Luaran di atas menunjukkan hasil estimasi parameter kemampuan peserta tes (kolom 1) dan kesalahan baku (kolom 2). Peserta tes dengan nomor urut 1 memiliki kemampuan (θ) = 2,206 dan kesalahan baku (se) = 0,609; peserta tes nomor urut 459 memiliki kemampuan (θ) = 0,852 dan kesalahan baku (se) = 0,399. Selanjutnya, untuk menampilkan grafik atau kurva yang menunjukkan hubungan antara kemampuan peserta tes dengan butir soal, digunakan script berikut.

```
> plot(p.R)
```

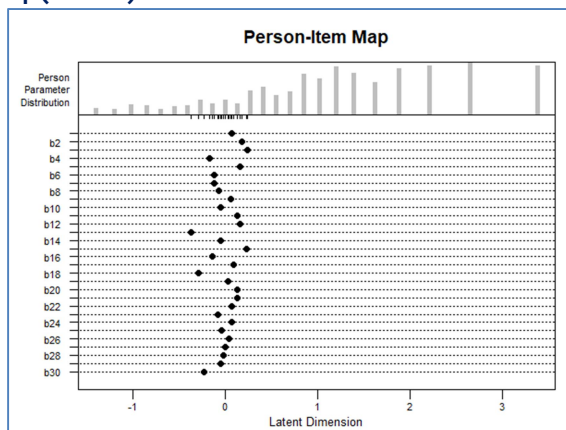


Gambar 34. Grafik parameter kemampuan orang (peserta tes)

Grafik parameter orang (kemampuan peserta tes) di atas menunjukkan bahwa semakin tinggi kemampuan peserta tes maka peluang menjawab butir soal dengan benar akan semakin besar. Peserta tes dengan kemampuan (θ) = -1 akan mampu menjawab kurang dari 10 butir soal dari 30 butir yang diujikan. Sebaliknya, peserta tes dengan kemampuan (θ) = +4 akan mampu

menjawab semua butir soal dengan benar. Untuk melihat sebaran kemampuan peserta tes, digunakan script berikut.

> plotPimap(mod.R)



Gambar 35. Sebaran kemampuan peserta tes pada semua butir soal

Pada grafik tersebut di atas, terlihat bahwa sebaran kemampuan peserta tes pada semua butir soal yang diujikan berada pada kisaran rata-rata ($\theta = 0$) yaitu -0,446 dan 0, 253.

d. Uji Kecocokan Model Data

Uji kecocokan model data bisa dilakukan dengan menggunakan: 1) uji Andersen's Likelihood Ratio (LRtest) (Andersen, 1973), 2) uji Martin's-Lof LR (Fischer & Molenaar, 1995), dan 3) Uji Wald (Fischer & Molenaar, 1995). Penggunaan berbagai metode dalam menguji kecocokan model data (*Goodness of Fit*) akan memberikan pemahaman yang utuh dan lengkap mengenai data yang dianalisis. Script yang digunakan untuk menguji kecocokan data dengan model.

```
> alrt<-LRtest(mod.R)
> alrt
> MLoef(mod.R, splitcr = "median")
> Waldtest(mod.R)
> plotGOF(alrt, conf=list())
```

Penjelasan script pada kotak perintah di atas adalah sebagai berikut.

1) Uji Andersen's

```
> alrt<-LRtest(mod.R)
> alrt

Andersen LR-test:
LR-value: 31.874
Chi-square df: 29
p-value: 0.325
```

Uji Andersen's di atas menunjukkan bahwa *p-value* yang dihasilkan lebih besar dari $\alpha = 0,05$, yaitu 0,325 (tidak signifikan), sehingga dapat dikatakan data cocok atau sesuai dengan model Rasch.

2) Martin's-Lof LR test

```
> MLoef(mod.R, splitcr = "median")

Martin-Loef-Test (split criterion: median)
LR-value: 136.784
Chi-square df: 224
p-value: 1
```

Berdasarkan output uji MLoef di atas diperoleh nilai *p-value* = 1 yang jauh lebih besar dari nilai $\alpha = 0,05$,. Sehingga dengan demikian uji homogen yang dijelaskan oleh model Rasch tidak ditolak atau data cocok/sesuai dengan model Rasch.

3) Uji Wald

Pendekatan pemisahan subjek untuk pengujian GOF (*Goodness of Fit*) untuk seluruh rangkaian item dalam hal parameter invariansis dapat diterapkan pada tingkat item menggunakan uji Wald. Dua subkelompok peserta tes dibentuk oleh median nilai mentah (default).

```
> Waldtest(mod.R)
```

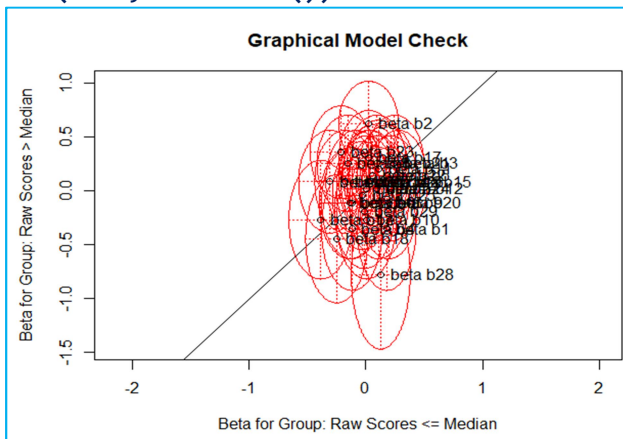
Wald test on item level (z-values):

	z-statistic	p-value
beta b1	-1.665	0.096
beta b2	2.524	0.012
beta b3	0.042	0.967
...
beta b28	-2.416	0.016
beta b29	-0.579	0.562
beta b30	1.413	0.158

Hasil tes Wald di atas menunjukkan nilai "z-statistik" dan "nilai-p" setiap item yang diuji. Tanpa penyesuaian tingkat kesalahan Tipe I, di bawah $\alpha = 0,05$ pada setiap item, pada output di atas, terlihat bahwa item 2 ($p\text{-value} = 0,012$) dan item 28 ($p\text{-value} = 0,016$) signifikan secara statistic (artinya menerima H_0 : butir tidak cocok dengan model). Koreksi Bonferroni ($\alpha/\text{jumlah item}$) menunjukkan kedua item tersebut tidak ada yang ditandai secara statistic atau hipotesis invariansi parameter item untuk item tersebut diterima.

Selain dengan menggunakan uji Andersen, Mloef, dan Wald. Kita juga bisa melakukan uji kecocokan model data dengan menggunakan grafik. Script yang digunakan adalah sebagai berikut.

> plotGOF(alrt, conf=list())



Gambar 36. Grafik analisis kecocokan model

plotGOF() adalah fungsi yang menyediakan pengujian grafis dari kesetaraan dua kelompok '(yaitu, subkelompok dengan skor mentah di bawah median dan subkelompok dengan skor mentah di atas median) estimasi kesulitan butir soal. Estimasi kesulitan butir dari kedua kelompok diplot dengan elips (pada gambar diberi warna merah) tingkat kepercayaan (*confidence*), dengan asumsi bahwa jika model Rasch sederhana merupakan deskripsi yang baik dari seluruh data, estimasi butir harus serupa atau hampir identik, dan elips harus melingkar dan memotong garis identitas. Konsisten dengan hasil uji Wald, tanpa koreksi Bonferroni, elips kepercayaan untuk item 2 dan 28 tidak menyentuh garis identitas, yang menunjukkan bahwa kedua item ini ditandai secara statistik.

e. Uji Kecocokan Butir dan Orang dengan Model

Setelah menganalisis dan mengetahui kecocokan model dengan data, analisis kecocokan dilanjutkan dengan menguji kecocokan butir dan kemampuan orang terhadap model yang digunakan.

```
> itemfit(p.R)
# nonparameteric test
> X<-as.matrix(d1contoh)
> NPtest(X, n=460, method = "T11")
> plotINFO(mod.R)
> plotINFO(mod.R, type="item")
> personfit(p.R)
```

1) Kecocokan Butir dengan Model

```
> itemfit(p.R)

Itemfit Statistics:
      Chisq  df  p-value  Outfit MSQ  Infit MSQ  Outfit t  Infit t  Discrim
b1  342.230 443   1.000    0.771    0.908   -2.352  -1.646  0.489
b2  537.599 443   0.001    1.211    1.053    2.048   0.972  0.331
b3  441.651 443   0.509    0.995    1.000   -0.025   0.015  0.391
... ..
b28 349.772 443   1.000    0.788    0.928   -2.035  -1.209  0.459
b29 422.069 443   0.756    0.951    0.976   -0.409  -0.381  0.402
b30 486.692 443   0.074    1.096    1.047    0.788   0.722  0.306
```

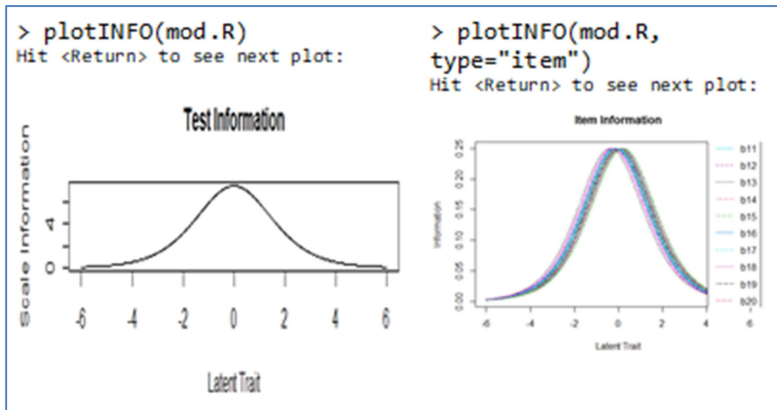
Indeks item fit yang paling terkenal dalam pemodelan Rasch adalah infit and outfit mean-square (MSQ) dan infit and outfit t-

statistics (Smith & Smith, 2004; Wright & Masters, 1982, 1990). Statistik infit adalah ukuran yang sangat kuat. Semakin mendekati nilai MSQ infit atau outfit dengan 1,0 maka semakin baik goodness of fitnya. T-statistik adalah versi standar dari MSQ. Semakin besar nilai absolut t , semakin besar kemungkinan ketidakcocokan item tersebut. Nilai absolut yang besar berangkat dari 0 seperti $|3|$ mungkin dianggap menandai item yang berpotensi tidak sesuai. Jika ada butir yang tidak sesuai atau cocok dengan model, sebaiknya butir tersebut dipertimbangkan untuk diperbaiki atau dibuang (drop).

Hasil di atas menunjukkan bahwa semua butir memiliki nilai infit and outfit mean-square (MSQ) yang mendekati 1,0. Sehingga dapat dikatakan bahwa semua butir cocok dengan model Rasch. Analisis kecocokan butir dengan model, bisa juga dilakukan dengan menggunakan uji nonparameterik sebagaimana script berikut.

```
# nonparameteric test
> X<-as.matrix(d1contoh)
> NPtest(X, n=460, method = "T11")
Nonparametric RM model test: T11 (global test - local dependence)
(sum of deviations between observed and expected inter-item
correlations)
Number of sampled matrices: 460
one-sided p-value: 0.8326087
```

p-value “satu ekor” pada output tersebut di atas lebih besar dari $\alpha = 0,05$, yang mengindikasikan bahwa model data secara global sesuai melalui pengujian korelasi inter-item sehingga dapat dikatakan data sesuai dengan model Rasch. Untuk menampilkan kurva informasi butir dan kurva informasi tes, dapat menggunakan script berikut.



Gambar 37. Informasi Tes dan Informasi Butir

Kurva informasi butir (IIC) dan kurva fungsi informasi tes (IRF) dapat juga digunakan untuk menganalisis apakah data cocok dengan model Rasch atau tidak (menggunakan paket "eRm"), estimasi parameter butir dan orang (menggunakan paket "eRm" dilakukan dengan metode estimasi CML (*Conditional Maximum Likelihood*)). Berbeda dengan menggunakan paket "mirt", di mana estimasi parameter butir maupun orang menggunakan metode estimasi MML (*Marginal Maximum Likelihood*)

2) Kecocokan orang dengan model

Ukuran infit dan outfit dapat diterapkan untuk mengevaluasi kecocokan orang (kemampuan peserta tes) dengan model. "eRm" menyediakan ukuran MSQ dan statistik t yang sesuai dan cocok digunakan untuk menguji ketidakcocokan string respons item untuk setiap orang relatif terhadap model Rasch.

```
> personfit(p.R)

Personfit Statistics:
  Chisq df p-value Outfit MSQ Infit MSQ Outfit t Infit t
1 29.905 29 0.419 0.997 0.999 0.16 0.16
2 30.388 29 0.395 1.013 1.008 0.19 0.13
3 30.471 29 0.391 1.016 1.013 0.18 0.16
...
145 28.829 29 0.474 0.961 0.993 0.08 0.15
146 30.283 29 0.400 1.009 1.009 0.27 0.25
147 25.498 29 0.652 0.850 0.989 0.16 0.30
[ reached getOption("max.print") -- omitted 302 rows ]
```

Hasil analisis tersebut di atas menunjukkan bahwa semua nilai **Outfit MSQ** dan **Infit MSQ** mendekati 1, dengan demikian string respons item untuk setiap orang relatif terhadap model Rasch sesuai/cocok.

2. Analisis Karakteristik Butir Tes Model Rasch dengan paket "mirt".

a. Estimasi Karakteristik Butir

```
> library(mirt)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> d1contoh <- read.table("D10N460B30.1.txt")
> mod.rasch<-mirt(d1contoh, model=1, itemtype="Rasch", SE=T)
> mod.rasch
> extract.mirt(mod.rasch, what="converged")
> extract.mirt(mod.rasch, what="secondordertest")
> coef(mod.rasch, IRTpars=T, simplify=T)
> fscores(mod.rasch, method="EAP", full.scores=T, full.scores.SE = T)
```

```
> mod.rasch<-mirt(d1contoh, model=1, itemtype="Rasch", SE=T)
Iteration: 11, Log-Lik: -7022.406, Max-Change: 0.00010
```

Calculating information matrix...

```
> mod.rasch
```

Call:

```
mirt(data = d1contoh, model = 1, itemtype = "Rasch", SE = T)
```

Full-information item factor analysis with 1 factor(s).

Converged within 1e-04 tolerance after 11 EM iterations.

mirt version: 1.34

M-step optimizer: nlminb

EM acceleration: Ramsay

Number of rectangular quadrature: 61

Latent density type: Gaussian

Information matrix estimated with method: Oakes


```
Second-order test: model is a possible local maximum
Condition number of information matrix = 10.27242
```

```
Log-likelihood = -7022.406
Estimated parameters: 31
AIC = 14106.81; AICc = 14111.45
BIC = 14234.88; SABIC = 14136.5
G2 (1073741792) = 8562.49, p = 1
RMSEA = 0, CFI = NaN, TLI = NaN
> extract.mirt(mod.rasch, what="converged")
[1] TRUE ##output ini menunjukkan bahwa data konvergen
> extract.mirt(mod.rasch, what="secondordertest")
[1] TRUE
> coef(mod.rasch, IRTpars=T, simplify=T)
$items
  a      b g u
b1 1 -1.287 0 1
b2 1 -1.181 0 1
b3 1 -1.117 0 1
... ..
b28 1 -1.383 0 1
b29 1 -1.411 0 1
b30 1 -1.587 0 1
```

Output koefisien sebagaimana di atas, diketahui semua daya beda (a) konstan = 1, parameter *guessing* (tebakan) (c) = 0. Parameter tingkat kesulitan butir soal nomor 1 sebesar $b_1 = -1,287$, $b_2 = -1,181$, $b_3 = -1,117$ dan seterusnya. Unidimensi di lihat dari lambing $u = 1$ (artinya semua butir mengukur hanya satu kemampuan saja, misalnya kemampuan Fisika pada siswa di MTs Muallimat NWDI Pancor sebagaimana di atas).

b. Estimasi Kemampuan Peserta Tes (θ)

```
> fscores(mod.rasch, method="EAP", full.scores=T,
+         full.scores.SE = T)
      F1      SE_F1
[1,] 0.71008655 0.5047549
[2,] -0.82584461 0.3601402
[3,] -0.69408677 0.3660693
... ..
[458,] 0.71008655 0.5047549
[459,] -0.41445539 0.3829210
[460,] -0.26354076 0.3943926
```

Estimasi skor kemampuan seseorang diperoleh dengan menggunakan fungsi "fscores()". Paket "mirt" menyediakan beberapa estimasi kemampuan orang (peserta tes), di mana metode estimasi *default* yang digunakan adalah estimasi *expected a posteriori* (EAP). Selain EAP, paket "mirt" juga menyediakan metode estimasi kemungkinan maksimum (ML), maksimum a posteriori (MAP) yang merupakan estimator model Bayes, estimator kemungkinan tertimbang (WLE; Warm, 1989), dan banyak lagi yang lainnya. Jika "full.scores=T" dihilangkan atau "full.scores=F" ditentukan, maka perkiraan kemampuan orang ditampilkan untuk rangkaian respons unik yang diamati, bukan untuk setiap orang.

Pada output di atas, hasil estimasi kemampuan orang (peserta tes) berada pada kolom F1 dan kesalahan baku di kolom SE_F1. Kemampuan peserta tes nomor urut 1 sebesar $\theta_1 = 0.71008655$ dengan kesalahan baku (SE_F1) = 0.5047549; Kemampuan peserta tes nomor urut 460 sebesar $\theta_{460} = -0.26354076$ dengan (SE_F1)₄₆₀ = 0.3943926.

Hasil estimasi kemampuan orang (θ) pada paket "eRm" berbeda dengan hasil estimasi kemampuan orang (θ) menggunakan paket "mirt". Dalam hal ini terlihat jelas bahwa penggunaan asumsi atau batasan identifikasi pada model yang berbeda dengan metode estimasi parameter item yang berbeda menghasilkan estimasi parameter item yang juga berbeda, meskipun menggunakan kumpulan data yang sama. Dengan demikian, perbedaan estimasi parameter item tercermin dalam estimasi kemampuan seseorang karena nilai parameter item estimasi digunakan untuk menghitung estimasi kemampuan seseorang (peserta tes).

c. Kurva Karakteristik Butir dan Informasi

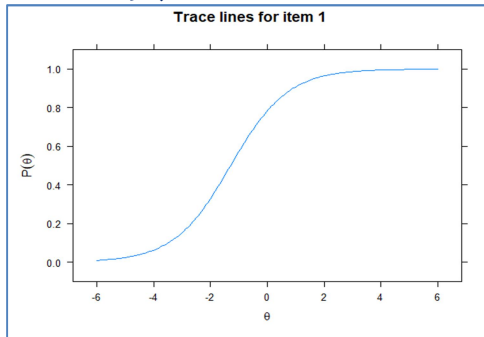
Untuk menampilkan grafik atau kurva karakteristik butir dan kurva fungsi informasi, gunakan script berikut:

```
> itemplot(mod.rasch,1)
> plot(mod.rasch, type = "trace")
> plot(mod.rasch)
> plot(mod.rasch, type="info")
> itemplot(mod.rasch, item=1, type="info")
```

Penjelasan:

Kurva IRF (*Item Response Function*) atau ICC (*Item Characteristics Curve*), digambar menggunakan fungsi "itemplot()". Misalkan kita ingin memunculkan gambar ICC untuk butir nomor 1, maka script yang digunakan "itemplot(mod.rasch,1)". Jika item ICC yang diinginkan berbeda, misalnya item 3, maka "itemplot(mod.rasch, 3)" digunakan.

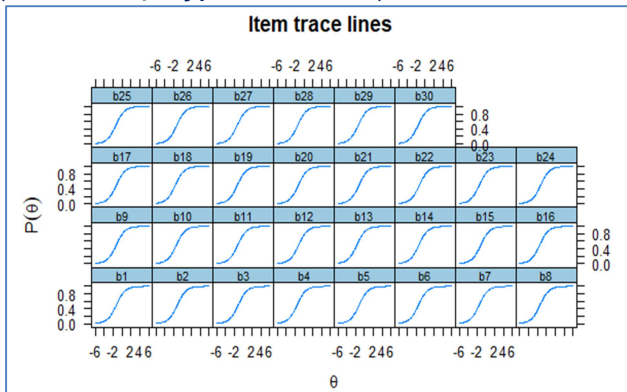
> **itemplot(mod.rasch,1)**



Gambar 38. ICC model Rasch butir nomor 1

Fungsi "plot()" dapat digunakan untuk menghasilkan berbagai grafik atau plot dari objek model Rasch. Misalnya kita ingin menampilkan keseluruhan ICC untuk semua butir, maka script yang digunakan "plot(mod.rasch, type="trace")".

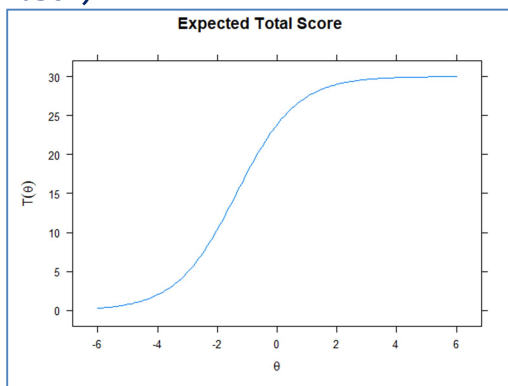
> **plot(mod.rasch, type = "trace")**



Gambar 39. ICC model Rasch seluruh butir

Selanjutnya, Memanggil fungsi, "plot(mod.rasch)," tanpa argumen tambahan apa pun, akan menghasilkan fungsi respons tes (TRF), atau kurva karakteristik tes (TCC). TCC menunjukkan estimasi skor kemampuan terhadap jumlah total skor yang diharapkan. Sebagaimana gambar berikut.

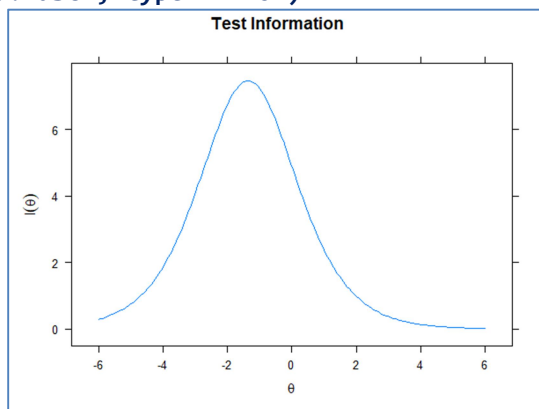
> plot(mod.rasch)



Gambar 40. TCC model Rasch

Untuk menampilkan kurva informasi tes, gunakan script berikut.

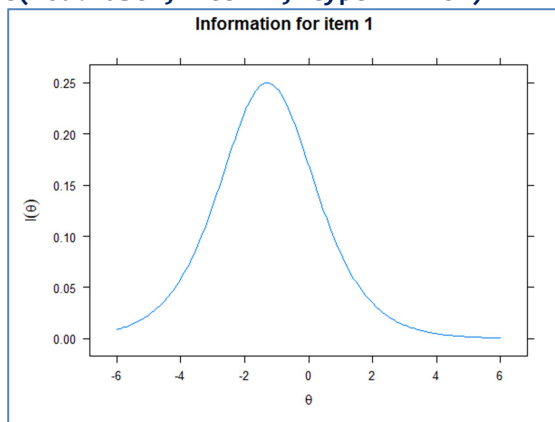
> plot(mod.rasch, type="info")



Gambar 41. Kurva Informasi Tes model Rasch

Jika peneliti lebih tertarik untuk melihat informasi butir nomor tertentu, misalnya butir 1 atau butir yang mana saja, gunakan script berikut.

```
> itemplot(mod.rasch, item=1, type="info")
```



Gambar 42. Kurva Informasi Butir model Rasch pada butir nomor 1

d. Uji Kecocokan Model Data

Untuk tujuan menguji kecocokan model menggunakan paket “mirt”, gunakan script berikut.

```
> M2(mod.rasch)
> itemfit(mod.rasch)
> itemfit(mod.rasch, empirical.plot=3, empirical.CI=.95)
> personfit(mod.rasch, method="EAP")
```

Penjelasan:

Pengujian kesesuaian model-data secara keseluruhan yang disebut M2, yaitu prosedur pengujian informasi terbatas karena menggunakan hingga margin orde kedua dalam data daripada menggunakan informasi lengkap dalam tes. M2 lebih kuat untuk mendeteksi variasi kemiringan (*slope*) atau parameter diskriminasi di seluruh butir ketika membandingkan Rasch atau 1PLM dengan 2PLM, atau ketika struktur sederhana multidimensi bifaktor ada dan model unidimensional cocok dengan data (lihat Xu, Paek, & Xia, 2017).

```
> M2(mod.rasch)
      M2 df      p RMSEA RMSEA_5 RMSEA_95 SRMSR TLI CFI
stats 409.9284 434 0.7910075      0      0 0.01140819 0.04652198 1.005261 1
```

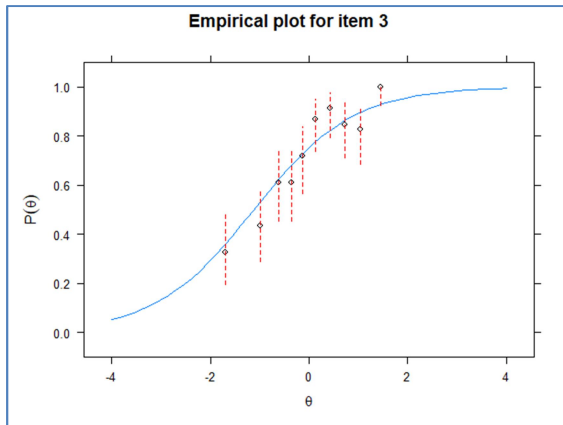
Berdasarkan output di atas, diperoleh p-value sebesar 0,7910075 yang jauh lebih besar dari $\alpha = 0,05$ (tidak signifikan). Hal ini berarti bahwa data secara keseluruhan sesuai dengan model Rasch. Untuk mengetahui apakah data sesuai dengan model pada level butir (setiap butir) digunakan fungsi "itemfit()" sebagaimana berikut.

```
> itemfit(mod.rasch)
  item  S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
1   b1 21.417    19    0.017  0.314
2   b2 24.315    19    0.025  0.184
3   b3 28.382    19    0.033  0.076
...
28  b28 21.330    20    0.012  0.378
29  b29 10.560    20    0.000  0.957
30  b30 13.104    20    0.000  0.873
```

Luaran hasil analisis kecocokan data dengan model (pada masing-masing butir) menggunakan fungsi "itemfit()" di atas, menunjukkan bahwa *p-value* (p.S_X²) untuk semua butir lebih besar dari nilai $\alpha = 0,05$ (tidak signifikan), dengan demikian dapat disimpulkan bahwa setiap butir pada data yang dianalisis cocok/sesuai dengan model Rasch. Selain itu, secara tradisional nilai RMSEA.S_X² juga jauh lebih kecil dari $\alpha = 0,05$ (interpretasinya sama bahwa semua butir sesuai dengan model Rasch).

Fungsi "itemfit()" juga dapat digunakan untuk menguji kecocokan data dengan model secara grafis pada tingkat item, misalnya untuk menampilkan grafik ICC berbasis model dan ICC empiris untuk butir nomor 3 digunakan perintah "empirical.plot=3". Untuk menentukan ICC empiris pada interval kepercayaan CI=.95% ditambahkan perintah "empirical.CI=.95", yang ditunjukkan sebagai garis vertikal putus-putus sebagaimana berikut.

```
> itemfit(mod.rasch, empirical.plot=3, empirical.CI=.95)
```



Gambar 43. ICC Empiris model Rasch butir nomor 3

Paket "mirt" menghasilkan ukuran kecocokan/kesesuaian kemampuan orang (peserta tes) dengan model yaitu dengan menggunakan fungsi "personfit()" (paket "eRm" tidak memiliki kemampuan ini). Outputnya berisi ukuran infit dan outfit dan statistik standar yang sesuai dengan string respons orang dan indeks kecocokan orang lz, yang disebut "Zh" pada kolom terakhir (nilainya dari -1 sampai dengan 1). Secara umum, nilai absolut yang lebih besar dari indeks ini menunjukkan pelanggaran yang lebih parah terhadap independensi lokal untuk pasangan item.

```
> personfit(mod.rasch, method="EAP")
      outfit      z.outfit      infit      z.infit      Zh
1  0.89742087 -0.0842170478 0.89947361 -0.085085986  0.226226458
2  1.04758490  0.4900655472 1.04081260  0.434111706 -0.424082010
3  1.05460939  0.4636461824 1.04999450  0.436137972 -0.410343562
...
198 0.53352568 -0.6561885920 0.54448716 -0.644954354  0.712652570
199 1.02667567  0.6828961194 1.02616283  0.684047435 -0.676489850
200 1.01431097  0.1656123497 0.98207862  0.079260942  0.021217854
[ reached 'max' / getOption("max.print") -- omitted 260 rows ]
```

Untuk output yang lebih ramah/mudah bagi pengguna, hasil "personfit()" dapat diurutkan dari statistik Zh terendah hingga tertinggi menggunakan kode berikut:

```

> p.fit <- cbind(1:460, d1contoh,
+               personfit(mod.rasch, method="EAP"))
> p.fit[order(p.fit[,26]),]

```

	outfit	z-outfit	infit	z-infit	Zh
1	0,89742	-0,08422	0,89947	-0,08509	0,22623
7	1,02630	0,28796	1,02639	0,29380	-0,26041
8	0,97213	-0,32634	0,97851	-0,25175	0,30017
...
69	1,03135	0,33613	1,03342	0,36231	-0,32385
81	1,07838	0,36197	1,04437	0,24910	-0,19139
84	0,97488	-0,01963	1,00317	0,09409	0,02314

B. Analisis Model 1 Parameter Logistik (1 PL)

Analisis butir IRT model 1 PL menggunakan paket “ltm” dan “mirt”, kedua paket tersebut menggunakan estimasi MML (*Marginal Maximum Likelihood*) dan mengasumsikan kemampuan laten peserta tes (θ) terdistribusi secara normal dengan rata-rata (μ) = 0 dan simpangan baku (σ) = 1 atau biasa ditulis $N(0,1)$. Disinilah perbedaan antara estimasi parameter model Rasch yang menggunakan estimasi CML (*conditional maximum likelihood*). Berikut langkah dan script R untuk analisis butir model 1 PL.

1. Analisis Karakteristik Butir Model 1 PL dengan paket “ltm”

a. Estimasi Parameter Butir

```

> library(ltm)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c1PL <- read.table("D10N460B30.39.txt")
> mod.1pl <- rasch(c1PL)
> coef.rasch(mod.1pl)

```

	Dffc1t	Dscrmn
b1	-0.9857776	0.9759354
b2	-1.0605583	0.9759354
b3	-1.0732348	0.9759354
...
b28	-1.5226915	0.9759354
b29	-1.1890349	0.9759354
b30	-1.5983328	0.9759354

Output hasil estimasi karakteristik butir model 1 PL menunjukkan bahwa tingkat kesulitan butir nomor 1 ($b_1 = -0,9857776$), butir nomor 2 ($b_2 = -1,0605583$) dan seterusnya. Nilai daya pembeda butir (a) konstan (sama) yaitu 0,9759. Hal ini membedakan model 1 PL dengan model Rasch, jika model Rasch nilai daya pembeda (a) = 1 untuk semua butir sedangkan model 1 PL konstan (sama) pada semua butir tetapi tidak selalu sama dengan 1. Oleh karena itu, kebanyakan para ahli pengukuran mengatakan bahwa model 1 PL adalah kasus khusus dari model 2PL.

> mod.1pl\$conv

[1] 0

Script “mod.1pl\$conv” digunakan untuk memverifikasi bahwa kalibrasi mencapai konvergensi. Nilai 0 menunjukkan estimasi selesai dalam aturan penghentian yang ditentukan dalam proses estimasi.

b. Estimasi parameter orang (kemampuan peserta tes)

Estimasi parameter kemampuan peserta tes, menggunakan script berikut.

```

> factor.scores(mod.1pl, method="EAP")
Factor-Scores for observed response patterns:
  b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14 b15 b16 b17 b18 b19 b20 b21 b22
1  0  0  0  0  0  0  0  0  0  0  0  0  1  1  0  0  1  1  0  0  0  0
2  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  1  0  1  0  0  1  0
3  0  0  0  0  0  0  0  1  0  1  0  1  0  0  1  0  1  0  1  0  1  1  1
...
27 0  0  0  1  1  0  1  0  0  1  0  0  0  0  0  0  0  0  0  1  1  1  1
28 0  0  0  1  1  0  1  0  0  1  1  1  1  1  0  0  1  1  1  1  1  0  1
29 0  0  0  1  1  1  0  0  0  1  1  1  1  1  0  0  1  1  1  0  1  1  0
  b23 b24 b25 b26 b27 b28 b29 b30 Obs Exp      z1 se.z1
1  0  0  1  0  0  0  0  1  1  0      -2.355 0.401
2  0  1  0  0  0  1  0  1  1  0      -2.207 0.377
3  1  1  1  1  1  1  1  1  1  0      -1.067 0.374
...
27 1  1  1  0  0  1  0  0  1  0      -1.679 0.378
28 1  0  1  1  1  1  1  1  1  0      -0.537 0.373
29 1  0  1  1  1  0  1  1  1  0      -0.794 0.365
[ reached 'max' / getOption("max.print") -- omitted 403 rows ]

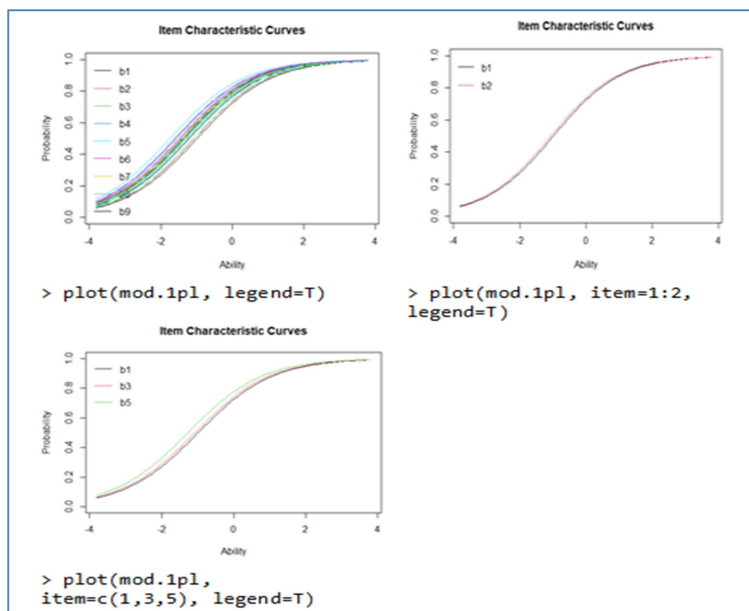
```

Setiap baris dalam output adalah pola respons unik yang diamati dari data, frekuensinya ("Obs"), frekuensi yang diharapkan ("Exp"), estimasi kemampuan EAP ("z1"), dan

kesalahan standarnya ("se. z1"). Secara teoritis, $2^{30} = 1.073.741.824$ pola respons unik yang mungkin, tetapi jumlah pola respons unik yang diamati dalam kumpulan data ini adalah 432. 1PLM juga menjaga korespondensi satu-satu antara estimasi skor laten dan skor mentah seperti pada Model Rasch. Terlihat bahwa kemampuan peserta tes nomor urut 1 $\theta_1 = -2,355$ dengan kesalahan baku (s.e) = 0,401; $\theta_2 = -2,207$ dengan s.e. = 0,377; dan seterusnya.

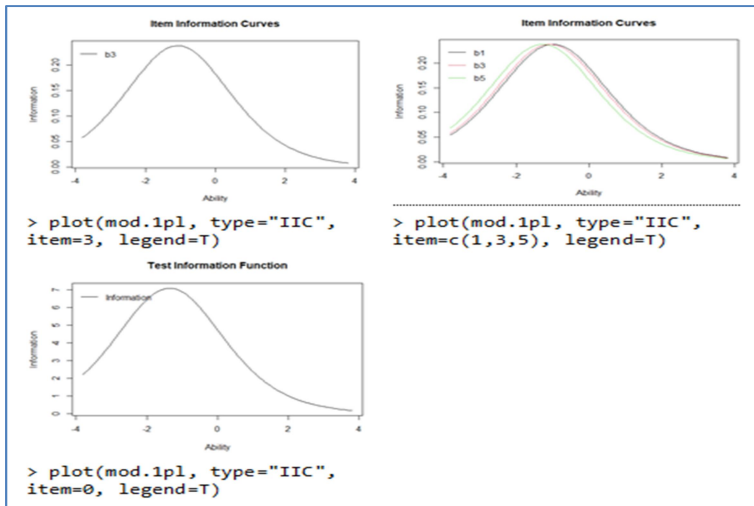
c. Menampilkan grafik ICC, IIC, IRF, dan TRF

Untuk memunculkan grafik atau kurva karakteristik butir dan kurva informasi butir, Anda cukup menuliskan script berikut.



Gambar 44. ICC model IRT 1PL

Selanjutnya, untuk menampilkan kurva informasi butir dan informasi tes, dapat dilihat pada script dan grafik output berikut.



Gambar 45. IIC dan TIF model IRT 1PL

d. Analisis Kecocokan Model

```
> item.fit(mod.1pl)
```

Item-Fit Statistics and P-values

Call:
 rasch(data = c1PL)

Alternative: Items do not fit the model
 Ability Categories: 10

	X^2	Pr(>X^2)
b1	19.1582	0.0140
b2	10.5082	0.2311
b3	15.7709	0.0458
...
b28	11.2126	0.1899
b29	8.6507	0.3726
b30	5.8498	0.6641

Statistik item fit, $Q1$, yang membandingkan proporsi empiris dan proporsi yang dihasilkan model pada sepuluh interval θ dihitung secara default, perhitungan kecocokan item ini berbasis simulasi.

```

> person.fit(mod.1pl)

Person-Fit Statistics and P-values

Call:
  rasch(data = c1PL)

Alternative: Inconsistent response pattern under the estimated model

      b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14 b15 b16 b17 b18 b19 b20 b21 b22
1    0  0  0  0  0  0  0  0  0  0  0  0  1  1  0  0  1  1  0  0  0  0
2    0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  1  0  1  0  0  1  0
3    0  0  0  0  0  0  0  1  0  1  0  1  0  0  1  0  1  0  1  0  1  1
...
28   0  0  0  1  1  0  1  0  0  1  1  1  1  0  0  1  1  1  1  1  0  1
29   0  0  0  1  1  1  0  0  0  1  1  1  1  0  0  1  1  1  0  0  1  1  0
30   0  0  0  1  1  1  0  1  0  0  0  1  0  1  0  0  0  0  0  0  0  1  1
...
      b23 b24 b25 b26 b27 b28 b29 b30      L0      Lz Pr(<Lz)
1    0  0  1  0  0  0  0  1 -14.7572  1.2573  0.8957
2    0  1  0  0  0  1  0  1 -16.0617  1.1518  0.8753
3    1  1  1  1  1  1  1  1 -20.8790 -0.5451  0.2929
...
...
28   1  0  1  1  1  1  1  1 -18.4646  0.0626  0.525
29   1  0  1  1  1  0  1  1 -19.6238  0.0577  0.523
30   1  1  0  0  0  1  0  1 -19.8345  1.0764  0.8591
[ reached 'max' / getOption("max.print") -- omitted 402 rows ]

```

Ada dua pengukuran kecocokan orang yang ditunjukkan oleh output di atas, yaitu l_0 ("L0") dan l_z ("Lz") yang merupakan versi standar dari l_0 . Kolom terakhir ("Pr(<Lz)") adalah probabilitas mengamati nilai l_z saat ini atau nilai yang lebih rendah berdasarkan perkiraan normal. Nilai "Pr(<Lz)" (misalnya., kurang dari .001) atau nilai "Lz" lebih kecil (misalnya., $L_z < 3$) dapat digunakan untuk menandai potensi pola respons menyimpang relatif terhadap 1PLM.

Berdasarkan output tersebut di atas, semua nilai Pr(<Lz) lebih besar dari 0,001 yang berarti parameter orang cocok/sesuai dengan model 1 PL. Untuk lebih meyakinkan peneliti, dilakukan perbandingan model misalnya dengan model 2 PL sebagaimana berikut.

Model-data fit and model comparison

Untuk mengetahui apakah data lebih cocok menggunakan model 1 PL ataukah 2 PL, digunakan analisis varians. Urutan objek yang disediakan penting, di mana model yang lebih kecil terdaftar terlebih dahulu. Pada contoh ini model 1 PL dijadikan sebagai hipotesis nol dan hipotesis alternatif adalah model yang lebih besar – 2PLM. Hasil uji rasio kemungkinan ("LRT," "df," dan "p.value"), AIC, dan BIC sebagaimana berikut.

```

> mod.2pl <- ltm(c1PL~z1, IRT.param=T)
> anova(mod.1pl, mod.2pl)

Likelihood Ratio Table
      AIC      BIC log.Lik  LRT df p.value
mod.1pl 14291.05 14419.12 -7114.52
mod.2pl 14311.66 14559.53 -7095.83 37.39 29 0.136

```

Berdasarkan output tersebut di atas, diperoleh p-value sebesar 0,136 jauh lebih besar dari nilai $\alpha = 0,05$. Sehingga hipotesis null (H_0) diterima atau keputusannya adalah menggunakan model 1 PL lebih baik daripada model 2 PL. Selain menggunakan anava, kita juga bisa menggunakan p-value dari chi-kuadrat yang diboostrop dengan script berikut.

```

> GoF.rasch(mod.1pl, B = 100)
Bootstrap Goodness-of-Fit using Pearson chi-squared
Call:
rasch(data = c1PL)
Tobs: 1101219472
# data-sets: 101
p-value: 0.426

```

Uji kecocokan model-data absolut untuk 1 PL menggunakan pendekatan bootstrap parametrik yang dilakukan dengan uji χ^2 dari Pearson. “GoF.rasch” hanya untuk evaluasi model dari objek yang dibuat dari fungsi “rasch()”. Untuk tujuan ilustrasi, pada contoh di atas jumlah ulangan untuk bootstrap diatur menjadi 100, “B = 100.” Dari hasil tes menunjukkan bahwa p-value (nilai p) yang diperoleh lebih besar dari $\alpha = 0,05$ yaitu p-value = 0,426. Artinya data lebih cocok menggunakan model 1 PL.

C. Analisis Butir Tes Model 2 Parameter Logistik (2 PL)

Estimasi parameter item dengan R menggunakan paket “ltm” dan “mirt” di mana estimasi dilakukan dengan menggunakan estimasi MML. Dalam estimasi MML, sifat laten orang, atau parameter kemampuan (θ) biasanya diasumsikan mengikuti distribusi normal, dan mean dan variansnya ditetapkan sebagai 0 dan 1, masing-masing, untuk menyelesaikan identifikasi model atau skala sifat laten masalah ketidakpastian.

1. Analisis Model 2 PL dengan paket "ltm"

a. Estimasi Parameter butir

```
library(ltm)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c2PL <- read.table("D10N460B30.1.txt")
> mod.2pl <- ltm(c2PL~z1, IRT.param=T)
> mod.2pl$conv
[1] 0
> summary(mod.2pl)
Call:
ltm(formula = c2PL ~ z1, IRT.param = T)
```

Model Summary:

log.Lik	AIC	BIC
-7009.138	14138.28	14386.15

Coefficients:

	value	std.err	z.vals
Dffclt.b1	-0.9989	0.1225	-8.1580
Dffclt.b2	-1.3538	0.2331	-5.8073
Dffclt.b3	-1.0881	0.1669	-6.5209
...
Dffclt.b28	-1.1201	0.1381	-8.1128
Dffclt.b29	-1.2951	0.1766	-7.3336
Dffclt.b30	-1.7903	0.2966	-6.0365
Dscrmn.b1	1.4849	0.2009	7.3896
Dscrmn.b2	0.8219	0.1432	5.7387
Dscrmn.b3	1.0260	0.1562	6.5683
...
Dscrmn.b28	1.3730	0.1915	7.1714
Dscrmn.b29	1.1206	0.1693	6.6202
Dscrmn.b30	0.8411	0.1537	5.4735

Fungsi "summary()" memberikan nilai kemungkinan log pada konvergensi akhir, AIC dan BIC, estimasi parameter item dan kesalahan standar yang sesuai. AIC dan BIC adalah indeks kecocokan model-data deskriptif. Sangat berguna untuk membandingkan model yang berbeda, di mana nilai yang lebih kecil menunjukkan kecocokan model-data yang lebih baik. AIC DAN BIC tidak digunakan dalam aplikasi kalibrasi model tunggal.

Berdasarkan output fungsi "summary()" di atas, diperoleh tingkat kesukaran butir, daya pembeda, kesalahan baku dan nilai z. Karena kebutuhan kita hanya untuk mengetahui parameter tingkat kesulitan dan daya pembeda soal, maka cukup dengan

perintah “coef(mod.2pl)” atau panggil langsung “mod.2pl” yang akan memberikan output seperti yang diinginkan (tingkat kesukaran dan daya pembeda).

```
> coef(mod.2pl)
      Dffclt  Dscrmn
b1 -0.9989477 1.4848505
b2 -1.3538408 0.8219267
b3 -1.0880983 1.0259837
...
b28 -1.1201478 1.3729909
b29 -1.2950673 1.1205913
b30 -1.7903413 0.8411438
```

Output di atas, menunjukkan bahwa tingkat kesulitan butir soal nomor 1 ($b_1 = -0,9989477$) dan daya pembedaya ($a_1 = 1,4848505$); butir nomor 28 (b_{28}) dengan tingkat kesulitan ($b_{28} = -1,1201478$) dan $a_{28} = 1,3729909$, dan seterusnya.

b. Estimasi parameter kemampuan (θ)

Fungsi “factor.scores()” menghasilkan estimasi skor sifat laten atau kemampuan peserta tes. “factor.scores()” menggunakan estimator EAP. EAP dapat dilihat sebagai pilihan alami untuk estimasi kemampuan orang dalam estimasi MML, di mana kemampuan orang (θ) diasumsikan mengikuti distribusi tertentu, seperti distribusi normal.

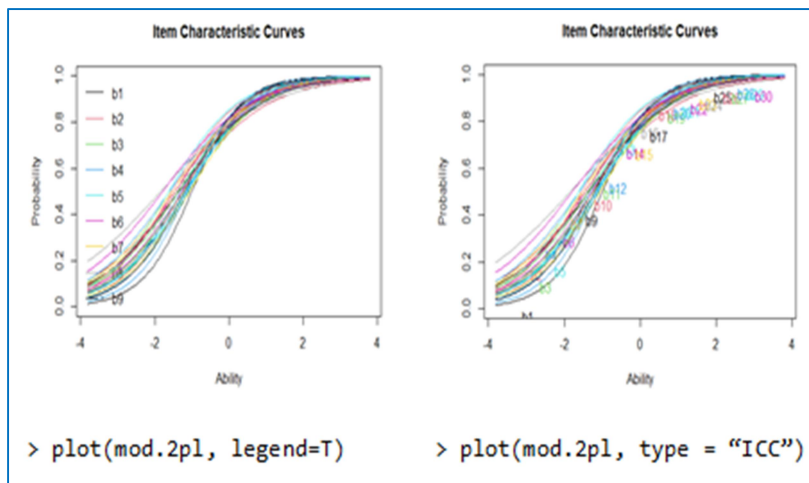
```
> factor.scores(mod.2pl, method="EAP")
Call:
ltm(formula = c2PL ~ z1, IRT.param = T)
Scoring Method: Expected A Posteriori
Factor-Scores for observed response patterns:
      b1  b2  b3  ...  b28  b29  b30  Obs  Exp  z1      se.z1
1     0   0   0   .    0    0    0    1    0  -2,049  0,339
2     0   0   0   .    0    1    0    1    0  -2,115  0,346
3     0   0   0   .    1    0    0    1    0  -1,953  0,345
4     0   0   0   .    1    1    1    1    0  -0,778  0,336
5     0   0   0   .    1    1    1    1    0  -0,172  0,392
.     .   .   .   .    .    .    .    .    .    .      .
27    0   0   1   .    1    1    1    1    0  -1,274  0,315
28    0   0   1   .    0    1    1    1    0  -1,404  0,309
29    0   0   1   .    0    1    1    1    0  -1,201  0,335
reached 'max' /getOption("max.print") -- omitted 394 rows ]
```

Kolom pertama dalam output adalah indeks untuk pola respons unik; hal itu tidak mewakili indeks identifikasi setiap orang. Dalam output, nilai terakhir 29 di kolom pertama berarti ada 29 pola respons unik yang diamati, atau string, dalam data di luar kemungkinan teoritis $2^{30} = 1.073.741.824$ jumlah pola respons

Kolom "Obs" dan "Exp" adalah frekuensi prediksi model teramati. Misalnya, dua peserta tes memiliki pola respons dengan semua nol dalam data ini. Kolom "z1" dan "se.z1" adalah estimasi kemampuan orang (θ) dan kesalahan standar. Dari hasil analisis tersebut, dapat dilihat bahwa kemampuan peserta tes dengan nomor urut 1 sebesar $\theta_1 = -2,049$ dengan kesalahan baku pengukuran $se.z1 = 0,339$. Begitu juga dengan peserta tes nomor urut 27, di mana estimasi kemampuan orang ($\theta_{27} = -1,274$) dan kesalahan baku pengukuran ($se.z1 = 0,315$)

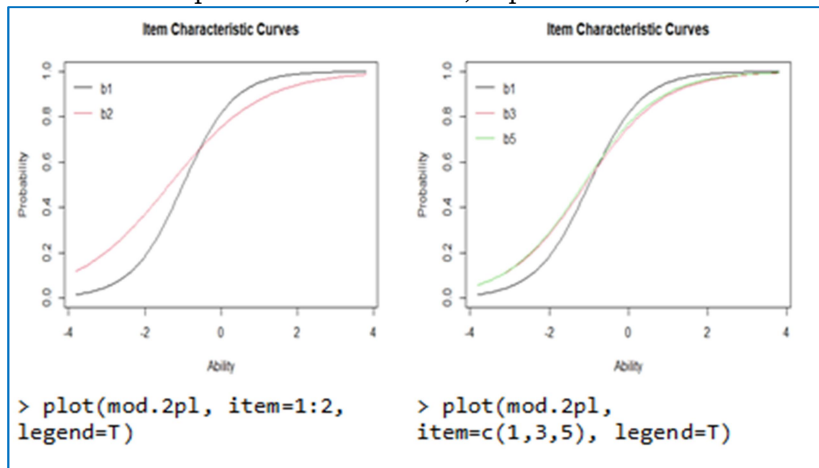
c. Kurva Karakteristik Butir (Model 2 PL)

Fungsi "plot()" digunakan untuk menampilkan ICC untuk satu atau lebih item. Tanpa spesifikasi apapun, semua item digambar dalam satu plot. Argumen "legend=T" menampilkan legenda sebagaimana tampilan ICC berikut.

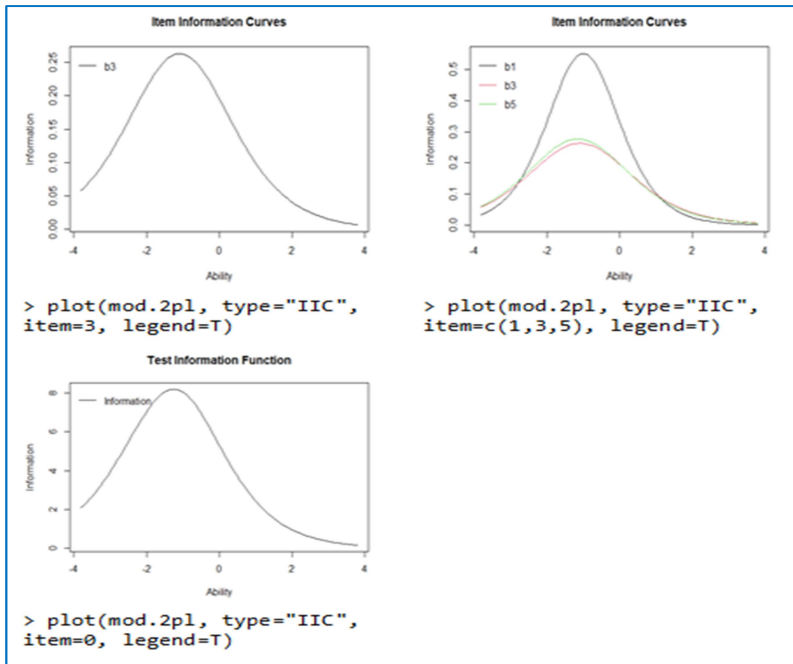


Gambar 46. ICC model IRT 2PL

Untuk menampilkan butir tertentu, seperti contoh berikut.



Gambar 47. ICC model IRT 2PL butir nomor 1 dan 2



Gambar 48. IIC model IRF model IRT 2PL

d. Uji Kecocokan Butir dan Orang dengan Model

Statistik kecocokan yang dihasilkan dalam fungsi “item.fit()” membandingkan pengamatan dengan proporsi menjawab benar item pada sepuluh interval θ dan merangkum perbedaannya. Tesnya adalah statistik Q1 dari Yen (1981, dalam Stone & Zhang, 2003). Proporsi yang diamati dan digunakan dihitung pada titik median di setiap interval θ .

```

> item.fit(mod.2pl)
Item-Fit Statistics and P-values
Call:
ltm(formula = c2PL ~ z1, IRT.param = T)
Alternative: Items do not fit the model
Ability Categories: 10

```

	X ²	Pr(>X ²)
b1	21.1843	0.0067
b2	13.1880	0.1055
b3	9.7998	0.2794
...
b28	14.5153	0.0693
b29	7.3158	0.5030
b30	8.1272	0.4211

Dengan koreksi Bonferroni untuk tingkat kesalahan Tipe I berdasarkan $\alpha = 0,05$, dua item (item 1 dan 21) ditandai secara statistik, yang menunjukkan potensi ketidaksesuaian item tersebut dengan 2PLM. Butir nomor 1 ($\text{Pr}(>X^2) = 0,0067$) lebih kecil dari $\alpha = 0,05$. Butir ini perlu dipertimbangkan untuk diperbaiki atau dibuang (drop)

Selain masalah pengaturan jumlah interval kemampuan, konstruksi interval didasarkan pada estimasi kemampuan, bukan kemampuan sebenarnya yang tidak teramati. Oleh karena itu, uji statistik dan p-value yang dihasilkannya berdasarkan distribusi sampling chi-kuadrat harus ditafsirkan secara kira-kira. Perilaku tes juga terpengaruh secara negatif jika jumlah item dalam tes kecil (misalnya 3 atau 5 butir). Orlando & Thissen (2000) menyarankan agar jumlah butir dalam satu tes lebih dari 40 butir. Namun, menurut penulis jumlah item yang terlalu banyak juga tidak baik bagi peserta tes. Dalam hal ini tingkat kejenuhan, kelelahan dan faktor lainnya perlu dipertimbangkan, jika sedikit butir bisa mencerminkan kemampuan peserta tes, kenapa harus membuat pertanyaan sebanyak mungkin?

Oleh karena itu diperlukan kemampuan penyusunan tes yang memadai, agar tujuan pengukuran dan penilaian hasil belajar tercapai secara maksimal serta memberikan informasi kemajuan belajar yang akurat.

```

> item.fit(mod.2pl, FUN = mean)
Item-Fit Statistics and P-values
Call:
ltm(formula = c2PL ~ z1, IRT.param = T)
Alternative: Items do not fit the model
Ability Categories: 10

      X^2 Pr(>X^2)
b1 20.1567 0.0098
b2 12.3332 0.1369
b3  9.0967 0.3342
...  ...      ...
b28 13.8665 0.0853
b29  6.6837 0.5711
b30  7.6588 0.4675

```

Kecocokan orang dengan Model

Ukuran kecocokan orang dengan L_0 dari (Levine & Rubin, 1979) dan L_z (Drasgow et al., 1985) disediakan dalam fungsi "person.fit()". Ukuran kecocokan orang digunakan untuk mendeteksi pola respons yang menyimpang, atau tidak biasa, relatif terhadap model yang digunakan. Baik L_0 dan L_z didasarkan pada kemungkinan pola respons orang. L_z adalah versi standar dari L_0 . Outputnya mencakup pola respons unik yang diamati dalam data, L_0 , L_z , dan *nilai-p* dari L_z . Kolom terakhir adalah *p-value* atau probabilitas L_z teramati atau nilai yang lebih kecil dari L_z . *p-value* yang kecil (misalnya, kurang dari 0,01) mengindikasikan pola respons yang berpotensi menyimpang secara relatif terhadap model 2PLM.

```

> person.fit(mod.2pl)
Person-Fit Statistics and P-values
Call:
ltm(formula = c2PL ~ z1, IRT.param = T)
Alternative: Inconsistent response pattern under the estimated model

```

	b1	b2	b3	.	b28	b29	b30	L0	Lz	Pr(<Lz)
1	0	0	0	.	0	0	0	-17,0155	0,8721	0,8084
2	0	0	0	.	0	1	0	-17,057	0,6163	0,7312
3	0	0	0	.	1	0	0	-18,4527	0,423	0,6639
4	0	0	0	.	1	1	1	-19,5085	0,1413	0,5562
5	0	0	0	.	1	1	1	-16,0685	0,1753	0,5696
.
28	0	0	1	.	0	1	1	-20,1767	0,6462	0,7409
29	0	0	1	.	0	1	1	-20,8606	-0,4276	0,3345
30	0	0	1	.	0	1	0	-20,3611	-0,4385	0,3305

```

[ reached 'max' /getOption("max.print") -- omitted 393 rows ]

```

Nilai “Pr(<Lz)” untuk semua peserta tes dari output di atas semuanya lebih besar dari 0,05. Artinya kemampuan semua orang (peserta tes) cocok dengan model 2PL.

Model-data fit and model comparison

Untuk mengetahui model mana yang lebih cocok untuk data hasil penelitian atau uji coba tes yang digunakan. Maka digunakan analisis varian (misalnya dalam hal ini model 1 PL dan model 2 PL). Data yang digunakan masih sama dengan data untuk 1 PL.

```

> mod.1pl<-rasch(c2PL)
> anova(mod.1pl, mod.2pl)

```

	AIC	BIC	log.Lik	LRT	df	p.value
mod.1pl	14106.85	14234.92	-7022.43			
mod.2pl	14138.28	14386.15	-7009.14	26.58	29	0.594

Statistik rasio kemungkinan adalah 26,58 dan derajat kebebasannya adalah 29, yang sama dengan perbedaan jumlah parameter model dalam dua model yang dibandingkan. 1PLM memiliki 30 item kesulitan dan satu parameter kemiringan (slope)

dengan total 31 parameter. 2PLM memiliki 30 soal kesulitan dan 30 parameter kemiringan/diskriminasi, dengan total 60. Selisihnya, $60 - 31 = 29$, adalah derajat bebasnya. *p-value* dari uji rasio kebolehjadian = 0,594; dengan demikian, menerima hipotesis nol yaitu soal/tes lebih nyocok menggunakan mode 1PLM daripada 2PLM.

Output juga menunjukkan dua kriteria informasi, AIC dan BIC. Nilai kriteria informasi untuk 1PLM lebih kecil dari pada 2PLM, Sehingga data atau soal yang dianalisis tersebut lebih cocok menggunakan 1 PLM.

2. Analisis Karakteristik Butir IRT 2 PL dengan paket “mirt”

a. Estimasi parameter butir

Estimasi butir dengan paket “mirt()” dengan argument “itemtype =2PL. “SE=T” ditentukan untuk mendapatkan kesalahan standar parameter item. Jika argumen tidak disertakan atau jika “SE=F”, kesalahan baku tidak dihitung secara default. “mirt()” menyediakan beberapa metode penghitungan kesalahan baku, untuk penghitungan kesalahan baku secara default adalah Oakes. Berikut script dan interpretasi hasil estimasi model 2 PL.

```
> library(mirt)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c2PL <- read.table("D10N460B30.1.txt")
> mod.2pl<-mirt(c2PL, model=1, itemtype="2PL", SE=T)
Iteration: 16, Log-Lik: -7009.176, Max-Change: 0.00007

Calculating information matrix...
```

Dengan mengetikkan “mod.2pl” menghasilkan uji konvergensi data yang dianalisis.

```
> mod.2pl
```

Call:

```
mirt(data = c2PL, model = 1, itemtype = "2PL", SE = T)
Full-information item factor analysis with 1 factor(s).
Converged within 1e-04 tolerance after 16 EM iterations.
mirt version: 1.34
M-step optimizer: BFGS
EM acceleration: Ramsay
Number of rectangular quadrature: 61
Latent density type: Gaussian
```

Information matrix estimated with method: Oakes
Second-order test: model is a possible local maximum
Condition number of information matrix = 9.634212

Log-likelihood = -7009.176
Estimated parameters: 60
AIC = 14138.35; AICc = 14156.7
BIC = 14386.22; SABIC = 14195.8
G2 (1073741763) = 8536.03, p = 1
RMSEA = 0, CFI = NaN, TLI = NaN

"extract.mirt()" digunakan untuk melakukan ekstraksi terpisah dari hasil pemeriksaan konvergensi. Ketika outputnya adalah "TRUE", ini menunjukkan bahwa estimasi berjalan lancar dalam kriteria penghentian yang ditentukan (default) dan maxima lokal yang mungkin untuk estimasi akhir

```
> extract.mirt(mod.2pl, what="converged")  
[1] TRUE  
> extract.mirt(mod.2pl, what="secondordertest")  
[1] TRUE
```

Untuk menampilkan parameter butir hasil analisis, digunakan fungsi "coef()" yang menghasilkan parameter daya pembeda (a), tingkat kesulitan butir (b), tebakan (g) dan unidimensi (u), sebagaimana ditunjukkan berikut ini.

```
> coef(mod.2pl, IRTpars=T, simplify=T)  
$items  
      a      b g u  
b1  1.485 -0.999 0 1  
b2  0.823 -1.353 0 1  
b3  1.026 -1.089 0 1  
...  ...      ...  
b28 1.375 -1.119 0 1  
b29 1.121 -1.295 0 1  
b30 0.842 -1.789 0 1
```

Berdasarkan output tersebut di atas, diketahui bahwa daya pembeda (a) untuk butir 1 ($a_1 = 1,485$), tingkat kesulitan butir 1 ($b_1 = -0,999$); daya pembeda butir nomor 29 ($a_{29} = 1,121$) dengan tingkat kesulitan ($b_{29} = -1,295$); parameter tebakan semua butir ($g = 0$) karena model yang dianalisis 2PL, dan parameter unidimensi semua butir sama dengan 1 (unidimensi)

```

> coef(mod.2pl, IRTpars=T)
$b1
      a      b  g  u
par   1.485 -0.999 0  1
CI_2.5 1.090 -1.240 NA NA
CI_97.5 1.880 -0.758 NA NA
...     ...     ...     ...

$GroupPars
      MEAN_1 COV_11
par      0      1
SE      NA      NA

> round (sqrt (diag (extract.mirt(mod.2pl,
+          what="vcov"))), 3)
      a1.1  d.2  a1.5  d.6  a1.9  d.10  a1.13  d.14  a1.17  d.18  a1.21  d.22
0.201 0.169 0.143 0.125 0.156 0.133 0.158 0.143 0.160 0.137 0.162 0.146
a1.25  d.26  a1.29  d.30  a1.33  d.34  a1.37  d.38  a1.41  d.42  a1.45  d.46
0.157 0.142 0.142 0.128 0.163 0.142 0.175 0.154 0.155 0.135 0.157 0.135
a1.49  d.50  a1.53  d.54  a1.57  d.58  a1.61  d.62  a1.65  d.66  a1.69  d.70
0.176 0.164 0.157 0.140 0.162 0.137 0.170 0.153 0.150 0.132 0.175 0.161
a1.73  d.74  a1.77  d.78  a1.81  d.82  a1.85  d.86  a1.89  d.90  a1.93  d.94
0.162 0.142 0.170 0.145 0.160 0.138 0.158 0.138 0.158 0.142 0.153 0.134
a1.97  d.98  a1.101 d.102 a1.105 d.106 a1.109 d.110 a1.113 d.114 a1.117 d.118
0.182 0.158 0.163 0.142 0.157 0.139 0.192 0.166 0.170 0.150 0.154 0.141

```

Nilai pertama dan kedua dari output ("a1.1" dan "d.2") adalah kesalahan standar kemiringan dan intersep untuk item pertama. Nilai kedua dan ketiga ("a1.5" dan "d.6") adalah kesalahan standar kemiringan dan intersep untuk item kedua. Urutan mengikuti kesalahan standar kemiringan terlebih dahulu dan perkiraan parameter intersep untuk item. Label (misalnya, a1.1, d.2, a1.5, d.6) adalah untuk penggunaan internal dalam estimasi model, dan tidak ada perhatian khusus yang diperlukan untuk label tersebut.

b. Estimasi parameter orang (kemampuan peserta tes)

Estimasi parameter kemampuan atau ability atau sifat laten dilakukan dengan memberikan perintah dengan fungsi "fscores()". Estimator kemampuan dengan metode EAP "metode =" EAP ". Paket "mirt" menawarkan lebih banyak opsi untuk estimator kemampuan orang seperti MAP (atau estimator Bayesian) dan WLE. Skor unik diminta dengan "full.scores=T," jika tidak, skor disediakan untuk string respons unik. "full.scores.SE = T" meminta kesalahan standar skor sifat laten.


```

> fscores(mod.2pl, method="EAP", full.scores=T,
+         full.scores.SE = T)
      F1      SE_F1
[1,]  0.621232405 0.4798941
[2,] -0.804912241 0.3432063
[3,] -0.611446165 0.3523680
...
[458,] 0.757989656 0.4999657
[459,] -0.382587854 0.3673149
[460,] -0.359773267 0.3690447

```

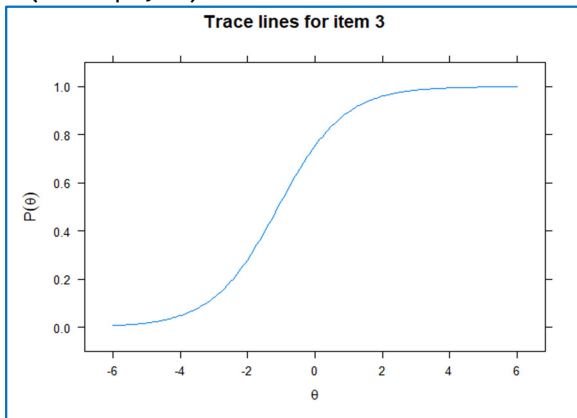
Berdasarkan output di atas, diketahui bahwa kemampuan peserta tes dengan nomor urut 1 (θ_1) sebesar 0.621232405 dengan kesalahan baku (s.e) = 0.4798941; peserta tes dengan nomor urut 458 (θ_{458}) = 0.757989656 dengan kesalahan baku (s.e.) = 0.4999657

c. Kurva Karakteristik Butir (ICC), Kurva Informasi Butir (IIC), IRF dan TCC

```

> itemplot(mod.2pl, 3)

```



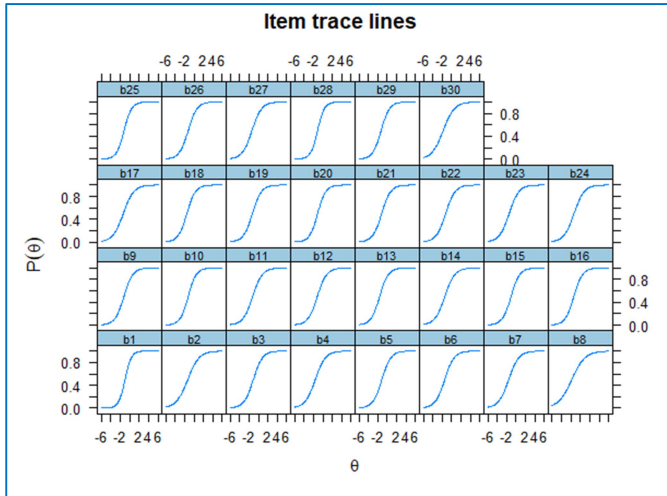
Gambar 49. ICC model IRT 2PL butir nomor 3 dengan paket “mirt”

Untuk meminta ICC semua item, fungsi "plot()" digunakan dengan argumen "type="trace", sebagaimana berikut.

```

> plot(mod.2pl, type = "trace")

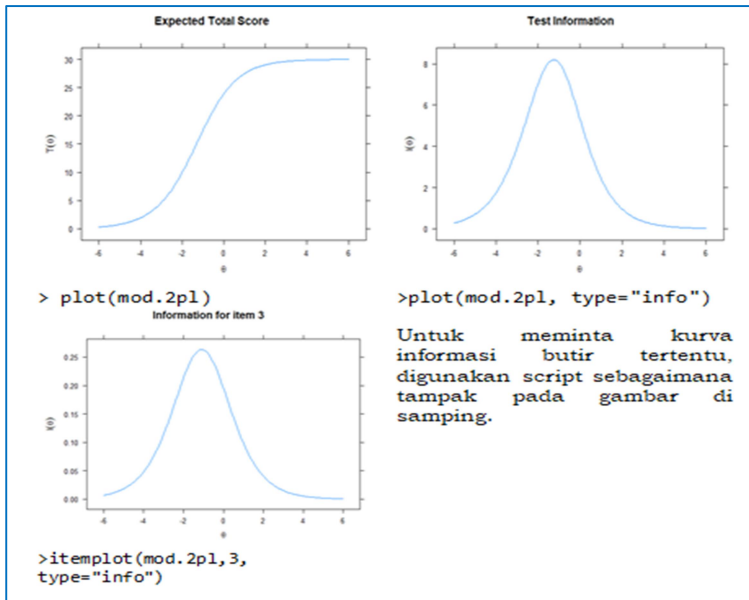
```



Gambar 50. ICC model IRT 2PL dengan paket “mirt”

Berdasarkan grafik semua ICC di atas, terlihat bahwa semua butir tidak memiliki kemiringan yang sama karena setiap butir memiliki diskriminasi yang diestimasi, atau parameter kemiringannya sendiri. Butir dengan estimasi diskriminasi (a) yang lebih rendah memiliki kurva yang lebih datar, sedangkan item dengan estimasi diskriminasi yang lebih tinggi memiliki kurva yang lebih curam.

Script yang digunakan untuk meminta kurva karakteristik tes (TCC) adalah fungsi “plot()” dengan tanpa tambahan argument. Selanjutnya untuk meminta Kurva informasi tes digunakan fungsi "plot()" dengan tambahan argumen "type="info".



Gambar 51. ETS dan Tes Information model IRT 2PL dengan paket "mirt"

d. Analisis Kecocokan Model

Fungsi "M2()" melaporkan statistik M2 kecocokan model-data keseluruhan yang absolut dengan ukuran kecocokan model-data deskriptif seperti RMSEA. Sebagaimana berikut.

```
> M2(mod.2pl)
      M2 df      p RMSEA RMSEA_5 RMSEA_95 SRMSR TLI CFI
stats 390.0627 405 0.6942522      0      0 0.01359908 0.04067649 1.003498 1
```

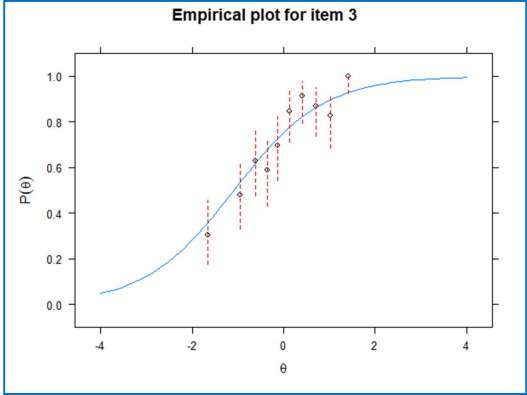
Output menunjukkan bahwa p-value M2 = 0,6942522 jauh lebih besar dari nilai $\alpha = 0,05$. Dengan demikian dapat disimpulkan bahwa seluruh butir (data) cocok dengan model 2 PL. Hasil ini juga sama dengan hasil analisis dengan menggunakan metode RMSEA dan nilai p-value berikut.

```
> itemfit(mod.2pl)
  item  S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
1    b1 17.806    17      0.010 0.401
2    b2 22.938    20      0.018 0.292
3    b3 28.366    19      0.033 0.077
...    ...    ...    ...    ...
28   b28 18.362    18      0.007 0.432
29   b29  9.165    19      0.000 0.971
30   b30 10.194    20      0.000 0.965
```

Berdasarkan output tersebut di atas, dapat dilihat bahwa nilai “p.S_X2” untuk semua butir lebih besar dari nilai $\alpha = 0,05$. Sehingga dapat disimpulkan bahwa seluruh butir cocok dengan model 2PL. Begitu juga dengan menggunakan “RMSEA.S_X2” semua butir lebih kecil dari 0,05 yang juga berarti semua butir cocok dengan model 2PL.

Uji kecocokan butir dengan model juga dapat dilakukan dengan membandingkan ICC berbasis model dengan ICC empiris, sebagaimana berikut.

```
> itemfit(mod.2pl, empirical.plot=3, empirical.CI=.95)
```



Gambar 52. ICC empirical model IRT 2PL dengan paket “mirt”

ICC berbasis model digambarkan dengan garis berwarna biru muda, sedangkan ICC empiris digambarkan dengan titik-titik hitam dengan garis putus-putus warna merah. Pada grafik tersebut di atas terlihat bahwa untuk item 3 ICC empiris mendekati atau berada pada sekitar ICC berbasis model.

Selain menganalisis kecocokan butir dengan model, R juga mampu menganalisis kecocokan orang (kemampuan peserta tes) dengan model. Fungsi yang digunakan adalah “personfit()” yang menghasilkan output yang hampir sama dengan analisis kecocokan orang pada model Rasch. Akan tetapi untuk model 2 parameter, hanya menggunakan zh. Jika nilai Zh < -3, maka perlu diperhatikan kemungkinan adanya pola respons menyimpang. sebagaimana berikut.

```
> personfit(mod.2pl, method="EAP")
      outfit      z.outfit      infit      z.infit      Zh
1  0.9262653 -0.025549332  0.87786218 -0.1570729004  0.251435483
2  1.0449919  0.465085562  1.03906347  0.4180945128 -0.405022697
3  1.0896563  0.668384558  1.08250689  0.6321312407 -0.614998650
...
198 0.5389658 -0.619437961  0.54402811 -0.6616499342  0.714099263
199 1.0188586  0.445003691  1.01814160  0.4383083053 -0.428894857
200 0.9890159  0.111038976  1.02133133  0.1865746260 -0.033459961
[ reached 'max' / getOption("max.print") -- omitted 260 rows ]
```

D. Analisis Butir Tes Model 3 Parameter Logistik (3 PL)

1. Analisis Karakteristik Butir model 3PL dengan paket “ltm”

a. Estimasi parameter butir

```
> library(ltm)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c3PL <- read.table("D10N460B30.1.txt")
> c3PLM <- tpm(c3PL, IRT.param = T)
> c3PLM
Call:
tpm(data = c3PL, IRT.param = T)

Coefficients:
      Gussng      Dffc1t      Dscrmn
b1      0.289      -0.431      2.265
b2      0.000      -1.346      0.823
b3      0.000      -1.090      1.010
...
b28     0.351      -0.395      2.267
b29     0.000      -1.287      1.118
b30     0.283      -1.055      0.987
```

Berdasarkan output di atas, diketahui bahwa tingkat kesulitan butir (“Dffc1t”) nomor 1 ($b_1 = -0,431$), daya pembeda (“Dscrmn”) ($a_1 = 2,265$) dan parameter tebakan (“Gussng”) ($c_1 = 0,289$). Hal ini mengindikasikan bahwa butir satu lebih sulit dari butir 2 ($b_2 = -1,346$), daya pembeda butir 1 juga lebih tinggi daripada daya

pembeda butir 2 ($a_2 = 0,823$). Akan tetapi jawaban tebakan pada butir 1 lebih tinggi daripada butir 2 ($c_2 = 0,000$).

b. Estimasi parameter orang

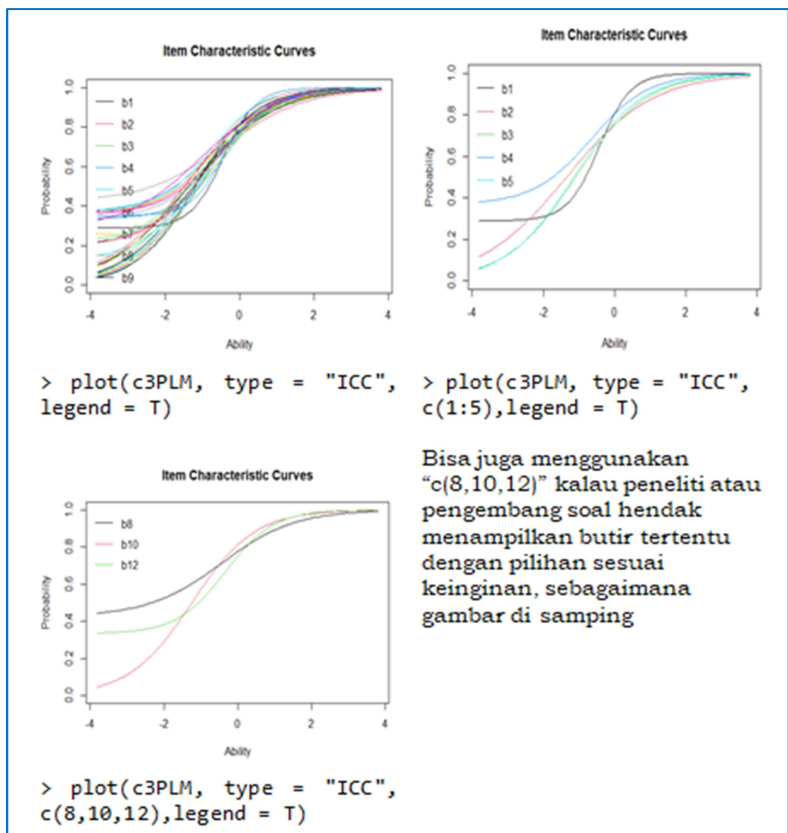
Estimasi parameter kemampuan peserta tes menggunakan script fungsi “factor.scores()” sebagaimana berikut.

```
> factor.scores(c3PLM)
Call:
tpm(data = c3PL, IRT.param = T)
Scoring Method: Empirical Bayes
Factor-Scores for observed response patterns:
  b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14 b15 b16 b17 b18 b19 b20 b21 b22
1  0  0  0  0  0  0  1  0  0  0  0  0  1  1  0  0  1  0  0  1  0  0  1
2  0  0  0  0  0  0  0  1  0  0  0  0  1  1  0  0  0  0  0  0  1  0  1
3  0  0  0  0  0  1  0  0  0  1  1  0  1  0  1  1  0  0  0  0  0  0  0
...
27 0  0  1  0  1  1  1  0  0  1  0  0  0  0  1  0  1  0  0  1  0  0  0
28 0  0  1  1  0  0  1  0  0  1  1  1  1  0  1  0  1  0  0  0  0  0  0
29 0  0  1  1  0  1  0  0  0  1  1  1  1  0  0  1  0  1  1  0  0  0  1
...
  b23 b24 b25 b26 b27 b28 b29 b30 Obs Exp  z1 se.z1
1  0  0  0  1  1  0  0  0  1  0  -2.174 0.486
2  0  0  0  1  0  0  1  0  1  0  -2.152 0.448
3  0  0  0  0  1  1  0  0  1  0  -1.852 0.383
...
27 1  1  0  1  0  1  1  1  1  0  -1.165 0.339
28 1  0  1  0  1  0  1  1  1  0  -1.308 0.347
29 0  1  1  0  1  1  1  1  1  0  -1.088 0.335
[ reached 'max' / getOption("max.print") -- omitted 394 rows ]
```

Luaran tersebut di atas menunjukkan bahwa Kemampuan peserta tes (z_1) untuk peserta tes nomor 1 (θ_1) = -2,174 dengan kesalahan baku ($se.z_1$) = 0,486; peserta tes nomor urut 29 memiliki kemampuan (θ_{29}) = -1,088 dengan kesalahan baku ($se.z_1$) = 0,335. Hal ini mengindikasikan bahwa kemampuan peserta tes nomor urut 29 lebih tinggi daripada peserta tes nomor urut 1.

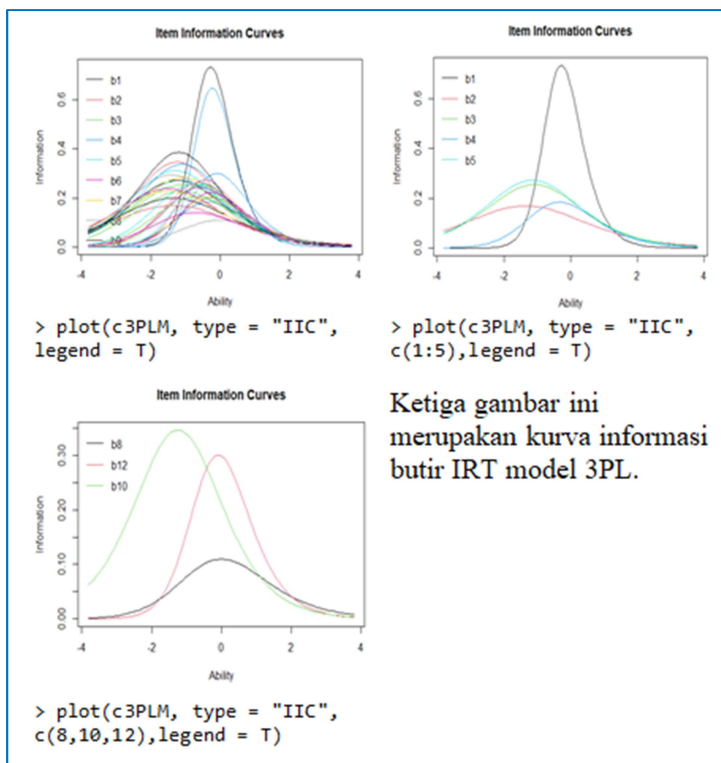
c. Plot ICC dan IIC

Grafik atau plot yang tersedia pada paket “ltm” adalah kurva karakteristik butir (ICC) dan kurva informasi butir (IIC). Script yang digunakan untuk menghasilkan plot ICC dengan menggunakan fungsi “plot()” ditambah dengan argument “type = “ICC”. Untuk argument “legend = T” adalah untuk mempercantik tampilan, jika tidak digunakan, maka nomor urut butir menyatu dengan grafiknya. Bila peneliti atau pengembang soal bermaksud hanya menampilkan beberapa saja dari ICC, maka fungsi “plot()” ditambah argument “c(1:5)” misalnya, untuk menampilkan ICC butir 1 sampai dengan 5.



Gambar 53. ICC model IRT 3PL dengan paket "ltm"

Kurva informasi butir (IIC) dihasilkan dengan menggunakan script yang sama dengan ICC, namun argument "type = "ICC" diganti dengan "type = "IIC". Begitu juga dengan perintah untuk menampilkan butir tertentu caranya sama dengan perintah yang diberikan pada ICC.



Gambar 54. IIC model IRT 3PL dengan paket "ltm"

d. Analisis Kecocokan Butir dan Orang dengan Model

Analisis kecocokan parameter orang (kemampuan peserta tes) dengan model dilakukan dengan menuliskan fungsi "person.fit()" seperti berikut.


```

> person.fit(c3PLM)
Person-Fit Statistics and P-values
Call:
tpm(data = c3PL, IRT.param = T)
Alternative: Inconsistent response pattern under the estimated model

      b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14 b15 b16 b17 b18 b19 b20 b21 b22
1  0  0  0  0  0  0  1  0  0  0  0  0  1  1  0  0  1  0  1  0  0  1
2  0  0  0  0  0  0  1  0  0  0  0  1  1  0  0  0  0  0  0  1  0  1
3  0  0  0  0  0  1  0  0  0  1  1  0  1  0  1  1  0  0  0  0  0  0
...
28 0  0  1  1  0  0  1  0  0  1  1  1  0  1  0  1  0  0  0  0  0  0
29 0  0  1  1  0  1  0  0  0  1  1  1  0  0  1  0  1  1  1  0  0  0
30 0  0  1  1  0  1  0  1  1  1  0  1  0  0  1  1  1  1  1  1  0  0

      b23 b24 b25 b26 b27 b28 b29 b30      L0      Lz Pr(<Lz)
1  0  0  0  1  1  0  0  0 -16.8576  1.0301  0.8485
2  0  0  0  1  0  0  1  0 -17.3470  0.8236  0.7949
3  0  0  0  0  1  1  0  0 -18.9488  0.6263  0.7344
...
28  1  0  1  0  1  0  1  1 -20.2879  0.6361  0.7376
29  0  1  1  0  1  0  1  1 -20.7568 -0.3186  0.375
30  1  1  1  1  0  0  1  0 -19.9770 -0.1750  0.4305
[ reached 'max' / getOption("max.print") -- omitted 393 rows ]

```

Berdasarkan output di atas, diketahui semua nilai “Pr(<Lz)” lebih besar dari -3 sebagaimana juga pada model 2PL yang telah dibahas sebelumnya. Sehingga disimpulkan bahwa semua peserta tes memiliki pola jawaban dan kemampuan yang sesuai atau cocok dengan model 3PL. Selanjutnya, untuk mengetahui apakah butir tes cocok atau sesuai dengan model, dilakukan analisis dengan script (perintah) fungsi “item.fit()” berikut.

```

> item.fit(c3PLM)
Item-Fit Statistics and P-values
Call:
tpm(data = c3PL, IRT.param = T)
Alternative: Items do not fit the model
Ability Categories: 10

      X^2 Pr(>X^2)
b1  10.1745  0.1789
b2  14.2894  0.0463
b3  10.8797  0.144
b21 23.5037  0.0014
b23 16.5081  0.0209
b26 15.7305  0.0277
b28 12.0399  0.0993
b29  5.9680  0.5435
b30  9.3495  0.2285

```

Dengan koreksi Bonferroni untuk tingkat kesalahan Tipe I berdasarkan $\alpha = 0,05$, terdapat 3 butir dari 30 butir yang diujikan ditandai secara statistik, di mana nilai “Pr(>X^2)” yang dihasilkan lebih kecil dari 0,05. Ketiga butir tersebut adalah butir nomor 21, 23, dan 26. Hal ini menunjukkan butir tersebut memiliki potensi

ketidaksesuaian butir dengan model 3PL. Sebaiknya ketiga butir tersebut diperbaiki atau dibuang.

2. Analisis Karakteristik Butir IRT 3 PL menggunakan paket “mirt”

```
> library(mirt)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c3PL <- read.table("D10N460B30.1.txt")
```

a. Estimasi parameter butir

```
> mod.3pl<-mirt(c3PL, model=1, itemtype="3PL", SE=T)
Iteration: 32, Log-Lik: -7000.720, Max-Change: 0.00008
```

Calculating information matrix...

```
> mod.3pl
```

Call:

```
mirt(data = c3PL, model = 1, itemtype = "3PL", SE = T)
```

Full-information item factor analysis with 1 factor(s).
Converged within 1e-04 tolerance after 32 EM iterations.

mirt version: 1.34

M-step optimizer: BFGS

EM acceleration: Ramsay

Number of rectangular quadrature: 61

Latent density type: Gaussian

Information matrix estimated with method: Oakes

Second-order test: model is a possible local maximum

Condition number of information matrix = 41457.2

Log-likelihood = -7000.72

Estimated parameters: 90

AIC = 14181.44; AICc = 14225.83

BIC = 14553.25; SABIC = 14267.62

G2 (1073741733) = 8519.12, p = 1

RMSEA = 0, CFI = NaN, TLI = NaN

```
> extract.mirt(mod.3pl, what="converged")
```

```
[1] TRUE
```

```
> extract.mirt(mod.3pl, what="secondordertest")
```

```
[1] TRUE
```

Parameter butir hasil analisis dapat dihasilkan dengan menggunakan fungsi “coef()” yang menghasilkan parameter daya pembeda (a), tingkat kesulitan butir (b), tebakan (g) dan unidimensi (u). sebagaimana ditunjukkan berikut ini.

```
> coef(mod.3pl, IRTpars=T, simplify=T)
$items
      a      b      g u
b1  2.274 -0.433 0.289 1
b2  0.826 -1.342 0.001 1
b3  1.017 -1.079 0.004 1
...  ...      ...      ...
b28 2.291 -0.394 0.352 1
b29 1.126 -1.268 0.010 1
b30 0.973 -1.113 0.263 1
```

Dari output di atas, diketahui daya pembeda butir 1 ($a_1 = 2,274$) dengan tingkat kesulitan ($b_1 = -0,433$) dan parameter tebakan ($g_1 = c_1 = 0,289$); daya pembeda butir nomor 29 ($a_{29} = 1,126$) dengan tingkat kesulitan ($b_{29} = -1,268$) dan parameter tebakan ($g_{29} = c_{29} = 0,010$). Judul kolom dengan label “**u**” adalah kategori untuk parameter ke empat model 4PL ($u =$ “kecerobohan”).

b. Estimasi Parameter Kemampuan Peserta Tes

Estimasi parameter kemampuan orang (θ) diestimasi menggunakan metode EAP, yang dihasilkan dari fungsi “fscores()” dengan menambahkan argument “method=“EAP””. Kolom berjudul “F1” menunjukkan estimasi sifat laten (θ) untuk setiap individu (karena “full.scores=T” ditambahkan pada fungsi “fscores()”), yang merupakan rata-rata distribusi posterior peserta tes, dan kolom berjudul “SE_F1” menunjukkan standar deviasi distribusi posterior peserta tes, yang diperlakukan sebagai kesalahan baku estimasi.

```
> fscores(mod.3pl, method="EAP", full.scores=T,
+         full.scores.SE = T)
      F1      SE_F1
[1,] 0.693269750 0.4637645
[2,] -0.771015996 0.3430016
[3,] -0.574641928 0.3434446
...  ...      ...
[458,] 0.777660910 0.4714832
[459,] -0.315851386 0.3801335
[460,] -0.396712712 0.3262243
```

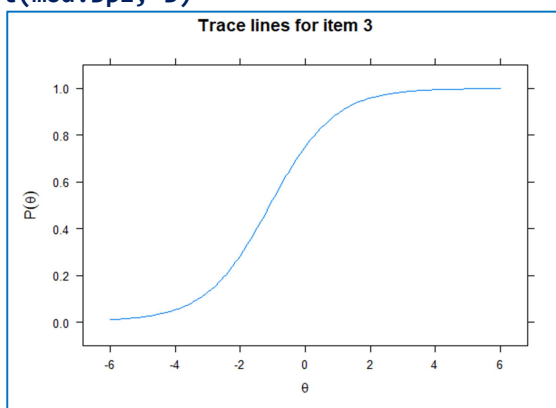
Berdasarkan output tersebut di atas, dapat diketahui bahwa kemampuan peserta tes dengan nomor urut 1 ($\theta_1 =$

0.693269750) dengan kesalahan baku estimasi ($se.F1 = 0.4637645$); peserta tes dengan nomor urut 460 memiliki kemampuan ($\theta_{460} = -0.396712712$) dengan kesalahan baku estimasi ($se.F1 = 0.3262243$); dan seterusnya.

c. Kurva Karakteristik Butir dan Kurva Informasi

Perintah atau script untuk menampilkan kurva karakteristik butir dan kurva informasi model 3PL dengan paket “mirt” sama dengan script sebelumnya (model 1PL dan 2 PL), yaitu menggunakan fungsi “itemplot()”, Misalkan kita ingin menampilkan ICC untuk butir nomor 3, maka perintah atau script yang digunakan seperti berikut.

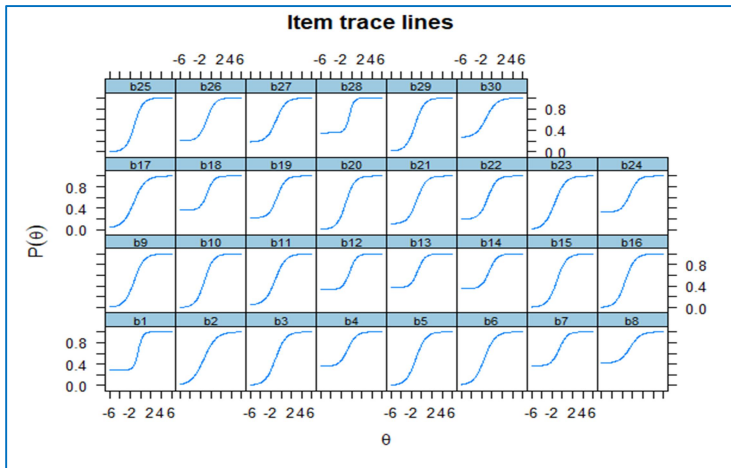
```
> itemplot(mod.3pl, 3)
```



Gambar 55. ICC model IRT 3PL butir nomor 3 dengan paket “mirt”

Perintah atau script untuk menampilkan ICC keseluruhan butir pada satu gambar atau plot secara terpisah (satu-satu) digunakan tambahan argument “type = “trace””.

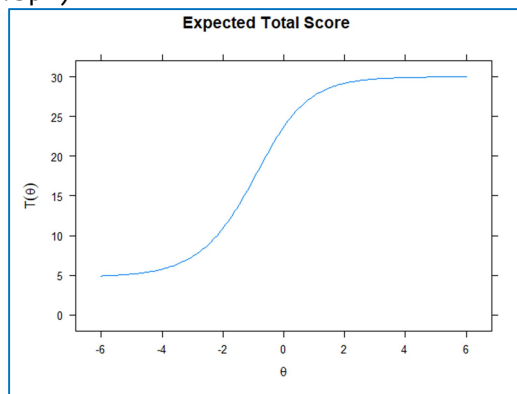
```
> plot(mod.3pl, type = "trace")
```



Gambar 56. ICC model IRT 3PL dengan paket “mirt”

Perintah untuk menampilkan grafik total skor yang diharapkan, cukup dengan mengetikkan fungsi “plot()” tanpa tambahan argument apapun.

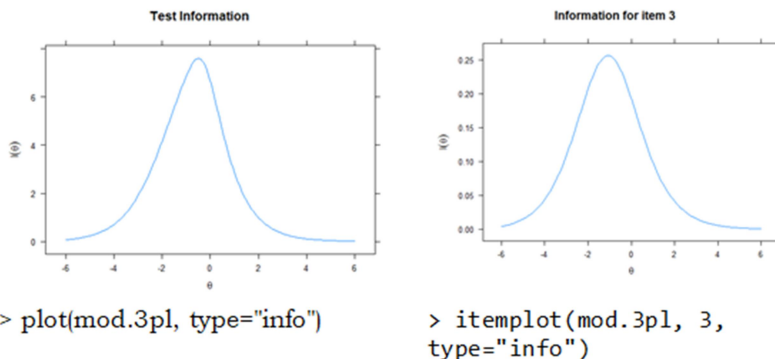
```
> plot(mod.3pl)
```



Gambar 57. ICC Total model IRT 3PL dengan paket “mirt”

Selanjutnya, kurva informasi tes dibuat menggunakan fungsi “plot()” dengan “type=“info”. Pada model 3PL nilai kemiringan butir soal masih memainkan peran penting untuk menentukan besarnya informasi tes di seluruh skala θ , selain itu informasi tes juga dipengaruhi oleh semua parameter butir soal (daya beda

butir, tingkat kesulitan, dan nilai parameter tebakan). Grafik fungsi informasi juga sangat diperlukan oleh peneliti, guru atau pengembang soal untuk mengetahui seberapa banyak informasi yang dapat diperoleh dari tes tersebut.



Gambar 58. Kurva Informasi Tes model IRT 3PL dengan paket “mirt”

Bila kita ingin menampilkan kurva informasi nomor butir tertentu, tambahkan nomor butir yang ingin ditampilkan, misalkan kurva informasi butir nomor 3 sebagaimana berikut.

```

> sum(coef(mod.3pl, IRTpars=T, simplify=T)$items[,3])
[1] 4.792598

```

d. Uji kecocokan model dengan data

Uji kecocokan model dengan paket “mirt” pada model 3PL menggunakan M2 yang dianggap sebagai varian chi-kuadrat. Nilai p (*p-value*) dari M2 pada data yang dianalisis sebesar 0,8636376. Sehingga dapat disimpulkan bahwa data cocok dengan model 3PL.

```

> M2(mod.3pl)

```

	M2	df	p	RMSEA	RMSEA_5	RMSEA_95	SRMSR	TLI	CFI
stats	345.1238	375	0.8636376	0	0	0.009576963	0.04060615	1.007557	1

Selain nilai p (*p-value*), analisis kecocokan menggunakan fungsi “M2()” juga menghasilkan RMSEA. Sebuah studi oleh Xu, Paek, dan Xia (2017) memberikan beberapa wawasan tentang perilaku RMSEA, yang didasarkan pada nilai chi-square M2. Hasil mereka menunjukkan bahwa nilai RMSEA untuk kecocokan yang baik harus mendekati atau sama dengan 0 daripada

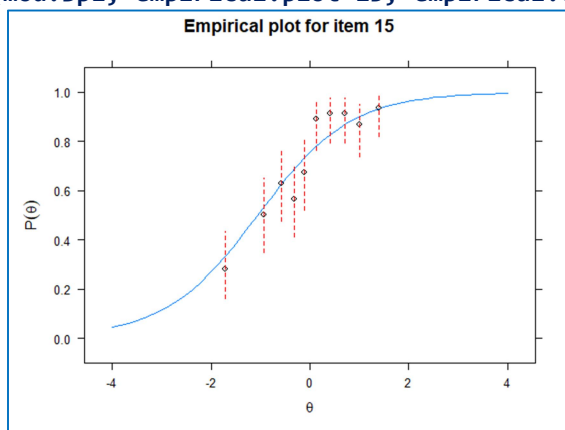
menggunakan nilai ambang batas yang diketahui secara konvensional sebesar 0,05 untuk *goodness of fit*. Pada output di atas nilai RMSEA = 0, sehingga dapat disimpulkan bahwa data yang dianalisis cocok dengan model 3PL.

Analisis kecocokan juga dilakukan pada butir terhadap model menggunakan nilai $p.S_X2$, dengan ketentuan jika $p.S_X2 \geq$ dari 0,05, maka data dikatakan cocok dengan model, begitupun sebaliknya.

```
> itemfit(mod.3pl)
  item  S_X2 df_S_X2 RMSEA.S_X2 p.S_X2
1   b1 15.854    17    0.000  0.534
2   b2 18.984    18    0.011  0.393
3   b3 27.862    18    0.035  0.064
...   ...      ...      ...      ...
28  b28 14.708    18    0.000  0.682
29  b29  9.655    18    0.000  0.943
30  b30 10.090    19    0.000  0.951
```

Pada output fungsi “itemfit()” di atas, terlihat bahwa semua nilai $p.S_X2 \geq$ dari 0,05 dan dengan koreksi Bonferroni di bawah tingkat kesalahan Tipe I sebesar 0,05, tidak ada item yang ditandai secara statistic (tanda bintang). Hal ini menunjukkan bahwa semua butir cocok atau sesuai dengan mode 3PL.

> itemfit(mod.3pl, empirical.plot=15, empirical.CI=.95)



Gambar 59. ICC Empiris model IRT 3PL butir nomor 15 dengan paket “mirt”

Grafik diatas adalah ICC berbasis model (warna biru muda) dan ICC empiris (titik-titik hitam dengan garis merah putus-putus). Terlihat dari grafik tersebut bahwa ICC empiris berada pada sekitar garis ICC berbasis model. Hal ini mengindikasikan dengan kuat bahwa butir 15 (contoh di atas) cocok/sesuai dengan model 3PL.

```
> personfit(mod.3pl, method="EAP")
      outfit      z.outfit      infit      z.infit      Zh
1  0.95686692  0.071531480  0.94471401  0.028868649  0.106180079
2  1.02750130  0.310760177  1.02359863  0.276586953 -0.257806421
3  1.06160466  0.493096056  1.06026048  0.492352669 -0.457667742
...
198 0.58665772 -0.416070811  0.57501231 -0.559919806  0.631646458
199 1.01196378  0.269626011  1.01156817  0.266199446 -0.252943506
200 0.98325113  0.098122711  1.01669385  0.174342409 -0.023908024
[ reached 'max' / getOption("max.print") -- omitted 260 rows ]
```

Statistik kecocokan orang dengan model (Iz) dihitung untuk setiap orang di bawah "Zh" menggunakan fungsi "personfit()". Jika nilai "Zh" lebih kecil dari -3 atau -2, maka hal ini mengindikasikan adanya potensi pola respons yang menyimpang atau tidak cocok dengan model. Pada output "personfit()" di atas, terlihat bahwa semua orang atau peserta memiliki pola jawab yang sesuai atau cocok dengan model 3PL.

E. Analisis Butir Tes Model 4 Parameter Logistik (4 PL)

Model 3PL memiliki parameter asimtot yang lebih rendah untuk mengakomodasi perilaku menebak semu peserta ujian dengan kemampuan sangat rendah. 4PLM menambahkan satu karakteristik lagi ke 3PLM, yang merupakan asimtot atas sebagai parameter bebas. Parameter ke empat ini diistilahkan sebagai parameter "carelessness" atau faktor kecerobohan dan dilambangkan dengan d.

Semua notasi sama seperti pada 3PLM, dengan penambahan parameter ui atau d. Parameter slip atau asimtot atas, ui atau d adalah batas atas item ke-i, atau probabilitas maksimum dari respons yang benar sepanjang skala θ . Parameter slip atau asimtot atas dirancang untuk mengakomodasi kesalahan peserta tes yang berkemampuan tinggi (menjawab salah) karena kecerobohan mereka atau alasan lain. Analisis karakteristik butir model 4PL menggunakan paket "mirt". Berikut script dan penjelasannya.

Estimasi 4PLM menggunakan argumen "itemtype=4PL" dalam fungsi "mirt()". Metode estimasi kesalahan standar default yang

digunakan adalah Oakes ketika argument “SE=T” digunakan. Estimasi 4PLM memiliki kendala mengenai parameter g_i dan u_i , serta masalah konvergensi ketika jumlah sampel dan butir terlalu sedikit (kecil). Direkomendasikan untuk menggunakan ukuran sampel yang besar dengan menggunakan distribusi item prior, khususnya prior untuk *pseudo-guessing* dan parameter slip.

1. Estimasi Parameter Butir

```
> ##R Script untuk IRT 4 PL menggunakan paket mirt
> library(mirt)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c4PL <- read.table("D10N460B30.1.txt")
> mod.4pl<-mirt(c4PL, model=1, itemtype="4PL", SE=T)
Iteration: 67, Log-Lik: -6990.213, Max-Change: 0.00009
```

Calculating information matrix...

```
> mod.4pl
```

Call:

```
mirt(data = c4PL, model = 1, itemtype = "4PL", SE = T)
```

Full-information item factor analysis with 1 factor(s).

Converged within 1e-04 tolerance after 67 EM iterations.

mirt version: 1.34

M-step optimizer: BFGS

EM acceleration: Ramsay

Number of rectangular quadrature: 61

Latent density type: Gaussian

Information matrix estimated with method: Oakes

Second-order test: model is a possible local maximum

Condition number of information matrix = 42453.4

Log-likelihood = -6990.213

Estimated parameters: 120

AIC = 14220.43; AICc = 14306.09

BIC = 14716.17; SABIC = 14335.33

G2 (1073741703) = 8498.1, $p = 1$

RMSEA = 0, CFI = NaN, TLI = NaN

```
> extract.mirt(mod.4pl, what="converged")
```

```
[1] TRUE
```

Cek konvergensi dilakukan dengan fungsi “extract.mirt()” dengan tambahan argument “what=“converged”” atau “what=“secondordertest”” yang menghasilkan “TRUE” pada contoh

di atas, artinya pemeriksaan konvergensi tidak mendeteksi masalah tertentu

```
> extract.mirt(mod.4pl, what="secondordertest")  
[1] TRUE
```

Estimasi titik parameter butir menggunakan versi yang disederhanakan yaitu dengan fungsi “coef()” dan ditambah dengan argument “simplify=T”. “a,” “b,” “g,” dan “u” masing-masing mewakili parameter diskriminasi (atau kemiringan), kesulitan, tebakan semu (atau asimtot bawah), dan slip (atau asimtot atas). Sifat laten atau parameter kemampuan populasi dari distribusi normal ditetapkan dengan rata-rata 0 dan standar deviasi 1.

```
> coef(mod.4pl, IRTpars=T, simplify=T)  
$items  
      a      b      g      u  
b1  2.447 -0.394  0.316  1.000  
b2  1.595 -1.388  0.010  0.874  
b3  1.492 -1.027  0.077  0.925  
...    ...    ...    ...    ...  
b28 2.139 -0.502  0.306  1.000  
b29 1.169 -1.242  0.009  1.000  
b30 0.988 -1.162  0.239  0.999  
$means  
F1  
  0  
$cov  
  F1  
F1  1
```

Output hasil estimasi parameter butir model 4PL tersebut di atas, menginformasikan bahwa butir tes nomor 1 memiliki daya pembeda, tingkat kesulitan, tebakan dan parameter slip masing-masing sebesar ($a_1 = 2,447$, $b_1 = -0,394$, $g_1 = 0,316$, $u_1 = 1,000$), dan seterusnya. Untuk menampilkan parameter butir lengkap dengan interval kepercayaan 95%, digunakan script fungsi “coef()” dengan tanpa penambahan argument “simplify=T”.

```

> coef(mod.4pl, IRTpars=T)
$b1
      a      b      g      u
par  2.447 -0.394 0.316 1.000
CI_2.5 1.240 -0.776 0.121 0.996
CI_97.5 3.654 -0.013 0.510 1.003

```

↓ Dipotong untuk menghemat ruang

```

$b30
      a      b      g      u
par  0.988 -1.162 0.239 0.999
CI_2.5 0.200 -3.571 -0.613 0.958
CI_97.5 1.776 1.246 1.092 1.040

```

```

$GroupPars
      MEAN_1 COV_11
par      0      1
CI_2.5   NA     NA
CI_97.5   NA     NA

```

Jika guru atau peneliti hendak menampilkan kesalahan baku (SE) gunakan argument tambahan “printSE=T” sebagaimana berikut ini.

```

> coef(mod.4pl, IRTpars=T, printSE=T)
$b1
      a      b      g      u
par 2.447 -0.394 0.316 1.000
SE 0.616 0.195 0.099 0.002

```

↓ Dipotong untuk menghemat ruang

```

$b30
      a      b      g      u
par 0.988 -1.162 0.239 0.999
SE 0.402 1.229 0.435 0.021

```

```

$GroupPars
      MEAN_1 COV_11
par      0      1
SE      NA     NA

```

2. Estimasi Kemampuan orang

Estimasi parameter kemampuan orang (θ) pada model 4PL diestimasi menggunakan metode EAP, yang dihasilkan dari fungsi “fscores()” dengan menambahkan argument “method=“EAP”. Kolom berjudul “F1” menunjukkan estimasi sifat laten (θ) untuk setiap individu (karena “full.scores=T” ditambahkan pada fungsi “fscores()”), yang merupakan rata-rata distribusi posterior peserta tes, dan kolom berjudul “SE_F1” menunjukkan standar deviasi distribusi

posterior peserta tes, yang diperlakukan sebagai kesalahan baku estimasi.

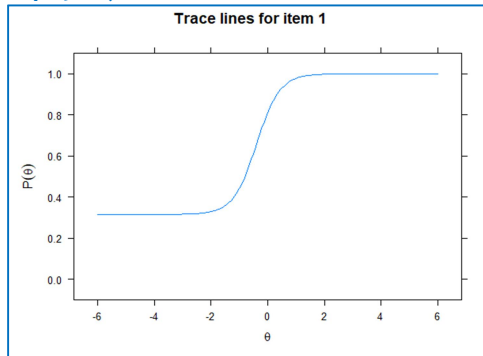
```
> fscores(mod.4pl, method="EAP",
+         full.scores=T, full.scores.SE = T)
      F1      SE_F1
[1,] 0.643253907 0.4913614
[2,] -0.797080769 0.3178918
[3,] -0.595834783 0.2865660
...
[458,] 0.576120906 0.4234430
[459,] -0.397170019 0.3708728
[460,] -0.354704393 0.2985977
```

Contoh ini menggunakan sampel sebanyak 460 orang peserta tes, di mana peserta tes dengan nomor urut 1 memiliki kemampuan (θ_1) sebesar 0.643253907 dengan kesalahan baku (SE_F1) sebesar 0.4913614. Dibandingkan dengan peserta dengan nomor urut 460, maka kemampuan peserta nomor urut 1 jauh lebih tinggi.

3. Kurva Karakteristik Butir dan Kurva Informasi

Perintah atau script untuk menampilkan grafik atau kurva karakteristik butir tertentu dengan menggunakan paket “mirt” pada model 4PL adalah sebagaimana berikut (angka 1 dimaksudkan untuk meminta komputer menampilkan ICC butir 1).

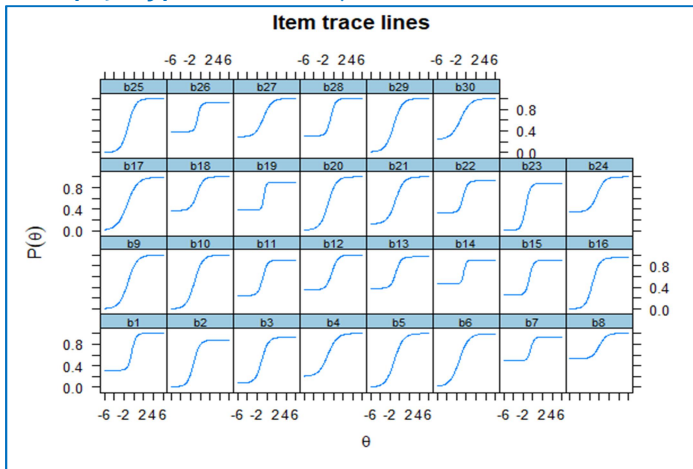
```
> itemplot(mod.4pl, 1)
```



Gambar 60. ICC model IRT 4PL butir nomor 1 dengan paket “mirt”

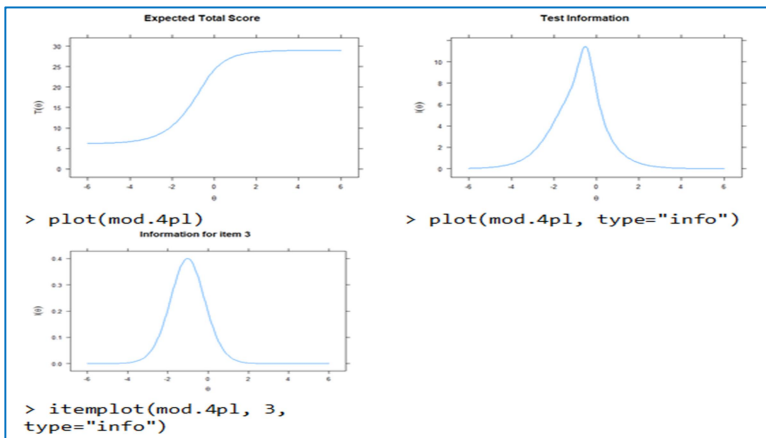
Jika peneliti atau guru bermaksud untuk menampilkan ICC semua butir dalam satu gambar secara terpisah (satu-satu), digunakan tambahan argument “type = “trace”” pada fungsi “plot()”

```
> plot(mod.4pl, type = "trace")
```



Gambar 61. ICC model IRT 4PL dengan paket “mirt”

Gambar grafik kurva karakteristik tes ditampilkan dengan script fungsi “plot()” tanpa tambahan argument.



Gambar 62. TCC model IRT 4PL dengan paket “mirt”

Dibandingkan dengan TCC 3PL, maka TCC 4PL sebagaimana output di atas memiliki ekor kiri yang terangkat pada kisaran kemampuan rendah dan ekor kanan yang rata pada kisaran

kemampuan tinggi karena estimasi bebas dan parameter slip. Jumlah taksiran g_i sama dengan nilai asimtot bawah TCC, dan jumlah taksiran u_i sama dengan nilai asimtot atas TCC.

Untuk menampilkan grafik fungsi informasi tes, gunakan fungsi “plot()” dengan tambahan argument “type=’info’”. Bila hanya ingin menampilkan TCC dengan nomor butir tertentu, cukup beri tambahan perintah dengan angka sesuai dengan nomor butir yang diinginkan, sebagaimana ditunjukkan pada gambar 61 di atas

4. Uji kecocokan Model Data

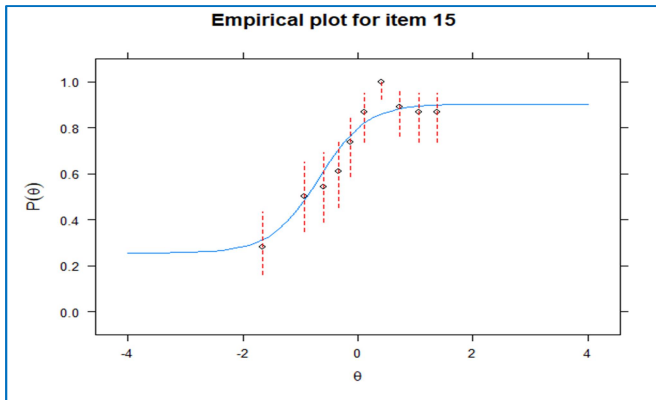
```
> M2(mod.4pl)
      M2 df      p RMSEA RMSEA_5 RMSEA_95 SRMSR TLI CFI
stats 318.2977 345 0.845772      0      0 0.01047482 0.04083704 1.007341 1
```

Uji kesesuaian model-data keseluruhan informasi M2, dihitung menggunakan fungsi “M2()”. Nilai chi-kuadratnya adalah 318.2977 dengan p-value = 0.845772. RMSEA yang didasarkan pada nilai chi-kuadrat M2 nilainya = 0. Hal ini menunjukkan bahwa data sesuai/cocok dengan model 4PL.

```
> itemfit(mod.4pl)
  item  S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
1   b1 18.130    16    0.017 0.316
2   b2 18.734    17    0.015 0.344
3   b3 28.135    17    0.038 0.043
...   ...   ...   ...   ...
28  b28 15.130    17    0.000 0.586
29  b29  9.361    16    0.000 0.898
30  b30 10.217    18    0.000 0.925
```

Pada output fungsi “itemfit()” model 4PL di atas, terlihat bahwa semua nilai $p.S_X2 \geq$ dari 0,05 dan dengan koreksi Bonferroni di bawah tingkat kesalahan Tipe I sebesar 0,05, tidak ada item yang ditandai secara statistic (tanda bintang). Hal ini menunjukkan bahwa semua butir cocok atau sesuai dengan mode 4PL.

```
> itemfit(mod.4pl, empirical.plot=15, empirical.CI=.95)
```



Gambar 63. ICC Empiris model IRT 4PL butir nomor 15 dengan paket “mirt”

Grafik diatas adalah ICC berbasis model (warna biru muda) dan ICC empiris (titik-titik hitam dengan garis merah putus-putus). Terlihat dari grafik tersebut bahwa ICC empiris berada pada sekitar garis ICC berbasis model. Hal ini mengindikasikan dengan kuat bahwa butir 15 (contoh di atas) cocok/sesuai dengan dengan model 4PL

```
> personfit(mod.4pl, method="EAP")
      outfit      z.outfit      infit      z.infit      Zh
1  0.98413312  0.129172527  0.95703722  0.05802142  0.068829381
2  1.02771483  0.333931078  1.02037167  0.26167155 -0.257906301
3  1.02990525  0.267182989  1.03439771  0.30768690 -0.254605952
...
198 0.36392186 -0.757579367 0.54212827 -0.64676488 0.766184046
199 0.99654403 -0.060862200 0.99652347 -0.06313948 0.077232468
200 1.22014987 0.569059402 1.29735523 0.72227715 -0.575124708
[ reached 'max' / getOption("max.print") -- omitted 260 rows ]
```

Statistik kecocokan orang dengan model (lz) dihitung untuk setiap orang di bawah "Zh" menggunakan fungsi "personfit()". Jika nilai "Zh" lebih kecil dari -3 atau -2, maka hal ini mengindikasikan adanya potensi pola respons yang menyimpang atau tidak cocok dengan model. Pada output "personfit()" di atas, terlihat bahwa semua orang atau peserta memiliki pola jawaban yang sesuai atau cocok dengan model 4PL.

DAFTAR PUSTAKA

- Alagumalai, S., Curtis, D. D., & Hungi, N. (Eds.). (2005). *Applied Rasch Measurement: A Book of Exemplars, Papers in Honour of John P. Keeves*. Netherlands: Springer.
- Andrich, D., & Marais, I. (2019). *A Course in Rasch Measurement Theory Measuring in the Educational, Social and Health Sciences*. Singapore: Springer Nature Singapore Pte Ltd.
- Baker, F. B. (2001). *The Basics Of Item Response Theory*, 2nd Edition, Wisconsin: ERIC Clearinghouse on Assessment and Evaluation.
- Baker, F. B., & Kim, S. (2004). *Item Response Theory Parameter Estimation Techniques Second Edition, Revised and Expanded*, New York: Marcel Dekker, Inc.
- (2017). *The Basics of Item Response Theory Using R*, Switzerland: Springer International Publishing AG.
- Bandalos, D. L. (2018). *Measurement Theory and Applications for the Social Sciences*. New York: The Guilford Press.
- Bock, R. D., & Gibbons, R. D. (2021). *Item Response Theory*. Hoboken: John Wiley & Sons, Inc.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. London: Lawrence Erlbaum Associates, Inc.
- Boomsma, A., van Duijn, M. A. J., & Snijders, T. A. B. (Eds.). (2001). *Essays on Item Response Theory*. New York: Springer Science+Business Media New York.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht: Springer Science+Business Media Dordrecht.
- Bulut, O., & Sünbül, O. (2017). Monte Carlo Simulation Studies in Item Response Theory with the R Programming Language. *Journal of Measurement and Evaluation in Education and Psychology*, 8(3): 266-287. DOI: 10.21031/epod.305821
- Bulut, O. (2013). *Between-person and within-person subscore reliability: comparison of unidimensional and multidimensional IRT models*. A Dissertation Submitted To The Faculty Of The Graduate School Of The University Of Minnesota. Diakses dari

https://conservancy.umn.edu/bitstream/handle/11299/155592/Bulut_umn_0130E_13879.pdf, tanggal 07 Nopember 2022.

- Bulut, O., Davison, M. L., & Rodriguez, M. C. (2017). Estimating between-person and within-person subscore reliability with profile analysis. *Multivariate Behavioral Research*, 52(1), 86-104.
- Crocker, L., & Algina, J. (2008). Introduction to Classical and Modern Test Theory, Ohio: Cengage Learning.
- Custer, M. (2015). "Sample Size and Item Parameter Estimation Precision When Utilizing the One-Parameter "Rasch" Model": Paper Presented at the Annual Meeting of the Mid-Western Educational Research Association Evanston, Illinois October 21-24.
- Davies, T. M. (2016). The Book of R : A First Course In Programming And Statistics. San Francisco: No Starch Press.
- de Ayala, R. J. (2009). The Theory And Practice Of Item Response Theory, New York: The Guilford Press.
- de Gruijter, D. N. M., & van der Kamp, L. J. T. (2008). Statistical Test Theory for The Behavioral Sciences. Boca Raton: Chapman & Hall/CRC.
- de Vries, A., & Meys, J. (2012). R For Dummies. England: John Wiley & Sons, Ltd.
- Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38(1), 67-86. <https://doi.org/10.1111/j.2044-8317.1985.tb00817.x>
- Eckes, T. (2015). Introduction to Many-Facet Rasch Measurement: Analyzing and Evaluating Rater-Mediated Assessments 2nd Revised and Updated Edition. Frankfurt am Main: Peter Lang GmbH.
- Embretson, S. E., & Reise, S. P. (2000). Item Response Theory for Psychologists Multivariate Applications Book Series. London: Lawrence Erlbaum Associates, Inc.
- Engelhard, G. J. & Wind, S. A. (2018). Invariansit Measurement With Raters And Rating Scales: Rasch Models for Rater-Mediated Assessments. New York: Routledge.
- Feinberg, R. A., & Rubright, J. D. (2016). Conducting Simulation Studies in Psychometrics. *Educational Measurement: Issues and Practice*, 35(2): 36-49

- Fischer, G. H., & Ivo W. Molenaar, I. W. (Eds.). (1995). *Rasch Models Foundations, Recent Developments, and Applications*, New York: Springer-Verlag.
- Fox, J-P. (2010). *Bayesian Item Response Modeling Theory and Applications*, New York: Springer Science+Business Media, LLC
- Gunawan, M. A. (2022). *Perbandingan Penyetaraan Kernel Secara Klasik Dan Modern Pada Tes Kemampuan Literasi Matematika. Disertasi*. Program Pascasarjana, Penelitian dan Evaluasi Pendidikan, Universitas Negeri Yogyakarta.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing And Validating Test Items*, Oxon: Routledge.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Academic Publishers.
- Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (pp. 147–200). Macmillan Publishing Co, Inc; American Council on Education
- Hambleton, R. K., Swaminathan, H., & Rogers, D. J. (1991). *Fundamentals of Item Response Theory*, London: Sage Publication.
- Hambleton, R. K., van der Linden, W. J., & Wells, C. S. (2010). IRT models for the analysis of polytomously scored data: Brief and selected history of model building advances. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 21–42). Routledge/Taylor & Francis Group
- Hessen, D. J. (2009). *Applying IRT using R*, Lecturer I. Materi Pelatihan Aplikasi Program R Dalam Bidang Psikometri dan Penilaian Pendidikan. Program Pascasarjana Universitas Negeri Yogyakarta.
- Horton, N. J., & Kleinman, K. (2015). *Using R and R Studio for Data Management, Statistical Analysis, and Graphics*. 2nd edition. Boca Raton: Taylor & Francis Group, LLC.
- Kalos, M.H. and Whitlock, P.A. (2008) *Monte Carlo Methods*. 2nd Edition, John Wiley & Sons, Inc., Hoboken. <http://dx.doi.org/10.1002/9783527626212>
- Khine, M. S. (Ed.). (2020). *Rasch Measurement: Applications in Quantitative Educational Research*. Singapore: Springer Nature Singapore Pte Ltd.

- Kolen, M. J. & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking Methods and Practices*. 3rd Edition. New York: Springer Science+Business Media New York.
- Kroese, D.P., Taimre, T. and Botev, Z.I. (2011) *Handbook of Monte Carlo Methods*. John Wiley & Sons, New York. <http://dx.doi.org/10.1002/9781118014967>.
- Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology* <https://online.ucpress.edu/collabra/article/8/1/33267/120491/Sample-Size-Justification>
- Lester, P. E., Inman, D., & Bishop, L. (2014). *Handbook of Tests and Measurement in Education and the Social Sciences*, Third Edition. London: Rowman & Littlefield
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics*, 4(4), 269–290. <https://doi.org/10.2307/1164595>
- Mailund, T. (2017). *Functional Data Structures in R*. Advanced Statistical Programming in R. California: Apress Media, LLC.
- Mair, P. (2018). *Modern Psychometrics with R*. Switzerland: Springer International Publishing AG.
- Mardapi, D. (2012). *Pengukuran Penilaian & Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.
- Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives*, 11(3), 71–101. <https://doi.org/10.1080/15366367.2013.831680>
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50–64. <https://doi.org/10.1177/01466216000241003>
- Paek, I and Cole, K. (2020). *Using R For Item Response Theory Model Applications*. New York: Routledge.
- Rasch, D., Kubinger, K. D., & Yanagida, T. (2011). *Statistics in Psychology Using R and SPSS*. United Kingdom: John Wiley & Sons, Ltd.
- Reise, S. P., & Revicki, D. A. (Eds.). (2015). *Handbook of Item Response Theory Modeling Applications to Typical Performance Assessment*, New York: Routledge.
- Retnawati, H. (2008). *Estimasi Efisiensi Relatif Tes Berdasarkan Teori Respons Butir dan Teori Tes Klasik*. *Disertasi Program Doktor Penelitian dan Evaluasi Pendidikan, Universitas Negeri Yogyakarta*.

- Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana*. Yogyakarta: Nuha Medika.
- Rizzo, M. L. (2008). *Statistical Computing with R*. Boca Raton: Taylor & Francis Group, LLC.
- Rubinstein, R. Y., & Kroese, D. P. (2007). *Simulation and The Monte Carlo Method (2nd Edition)*. New Jersey: John Wiley & Sons, Inc.
- Sahin, A., & Anil, D. (2017). "The Effects of Test Length and Sample Size on Item Parameters in Item Response Theory". *Educational Sciences: Theory & Practice*, 17, 321–335. <http://dx.doi.org/10.12738/estp.2017.1.0270>
- Schmuller, J. (2017). *Statistical Analysis with R for Dummies*. New Jersey: John Wiley & Sons, Inc.
- Smith, R. M. (1999). *Rasch Measurement Models: Interpreting WINSTEPS/BIGSTEPS and Facets Output*. Minnesota: JAM Press.
- Stone, C. A., & Zhang, B. (2003). Assessing Goodness of Fit of Item Response Theory Models: A Comparison of Traditional and Alternative Procedures. *Journal of Educational Measurement*, 40(4), 331–352. <http://www.jstor.org/stable/1435385>.
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*. Cimahi: Trim Komunikata Publishing House.
- Sylva, W. (2020). Dimensionality and validity of the operational performance construct in the aviation industry: a factor analytic approach. *Research Journal of Business and Management (RJBM)*, V.7(4), p.299.321. <http://doi.org/10.17261/Pressacademia.2020.1326>.
- Al-Tarawnah, E. A-W., & Al-Qahtani, M. (2022). The Effect of Test Length on the Accuracy of Estimating Ability Parameter in the Two- and Three-Parameter Logistic Models: Comparison by Using the Bayesian Method of Expected Prior Mode and Maximum Likelihood Estimation. *Journal of Educational and Social Research*, 12(1): 168-187. DOI: <https://doi.org/10.36941/jesr-2022-0015>
- Uyigue, A. V., & Orheruata, M. U. (2019). Test Length and Sample Size for Item-Difficulty Parameter Estimation in Item Response Theory. *Journal of Education and Practice*, 10(30): 72-75, DOI: 10.7176/JEP/10-30-08

- van der Linden, W. J. (2005). *Linear Models for Optimal Test Design*. New York: Springer Science+Business Media, Inc.
- van der Linden, W. J. (Ed.). (2018). *Handbook of Item Response Theory Volume Three Applications*, Boca Raton: Taylor and Francis Group, LLC.
- van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of Modern Item Response Theory*, New York: Springer Science+Business Media New York.
- Watkins, M. W. (2021). *A Step-By-Step Guide To Exploratory Factor Analysis With R And Rstudio*. New York: Routledge.
- Waugh, R. (Ed.). (2010). *Applications Of Rasch Measurement In Education*. New York: Nova Science Publishers, Inc
- Weiss, D. J. (Ed.). (1983). *New Horizons In Testing: Latent Trait Test Theory and Computerized Adaptive Testing*. New York: Academic Press.
- Wood, J. (2017). Logistik IRT Models. Diakses dari https://quantdev.ssri.psu.edu/sites/qdev/files/IRT_tutorial_F_A17_2.html#overview, tanggal 06 November 2022
- Wright, B., & Stone, M. (1999). *Measurement Essentials*, 2nd Edition. Wilmington: Wide Range, Inc.
- Zanon, C., Hutz, C. S., Yoo, H., & Hambleton, R. K. (2016). An Application of Item Response Theory to Psychological Test Development. *Psicol. Refl. Crit*, 29(18): 1-10
<https://doi.org/10.1186/s41155-016-0040-x>.
- Zumel, N., & Mount, J. (2014). *Practical Data Science with R*. New York: Manning Publications, Co.

Lampiran 01. Script Analisis Karakteristik Butir Dikotomus Model Rasch

##ANALISIS BUTIR MODEL RASCH

```
> library(eRm)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> d1contoh <- read.table("D10N460B30.1.txt")
> mod.R<-RM(d1contoh)
> mod.R$conv
> round (cbind(-mod.R$betapar, mod.R$se.beta), 3)
> round (cbind(-confint(mod.R)[,2], -confint(mod.R)[,1]), 3)
> summary(mod.R)
```

##ICC dan IIC

```
> plotICC(mod.R, 1:4)
> plotICC(mod.R, item.subset = c(1,3,5,9), empICC = list("raw"))
> p.R<-person.parameter(mod.R)
> p.R
> round(cbind(p.R$thetapar$NAgroup1, p.R$se.theta$NAgroup1), 3)
> plot(p.R)
> plotPImap(mod.R)
```

##Uji Kecocokan Model

```
> alrt<-LRtest(mod.R)
> alrt
> MLoef(mod.R, splitcr = "median")
> Waldtest(mod.R)
> plotGOF(alrt, conf=list())
> itemfit(p.R)
> plotICC(mod.R, item.subset = 1:4, empICC=list("raw"))
```

nonparameteric test

```
> X<-as.matrix(d1contoh)
> NPtest(X, n=460, method = "T11")
> personfit(p.R)
> plotINFO(mod.R)
> plotINFO(mod.R, type="item")
```

RASCH Model dengan MIRT

```
> library(mirt)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> d1contoh <- read.table("D10N460B30.1.txt")
> mod.rasch<-mirt(d1contoh, model=1, itemtype="Rasch", SE=T)
> mod.rasch
> extract.mirt(mod.rasch, what="converged")
> extract.mirt(mod.rasch, what="secondorderstest")
```

```

> coef(mod.rasch, IRTpars=T, simplify=T)
> coef(mod.rasch, IRTpars=T)
> round(sqrt(diag(extract.mirt(mod.rasch, what="vcov"))),3)
> coef(mod.rasch, printSE=T)
> fscores(mod.rasch, method="EAP", full.scores=T,
+         full.scores.SE = T)
> itemplot(mod.rasch,1)
> plot(mod.rasch, type = "trace")
> plot(mod.rasch)
> plot(mod.rasch, type="info")
> itemplot(mod.rasch, item=1, type="info")

```

Analisis Kecocokan Model

```

> M2(mod.rasch)
> itemfit(mod.rasch)
> itemfit(mod.rasch, empirical.plot=3, empirical.CI=.95)
> personfit(mod.rasch, method="EAP")
> residuals(mod.rasch, type="LD")
> residuals(mod.rasch, type="Q3")
> p.fit <- cbind(1:460, d1contoh, personfit(mod.rasch,
+         method="EAP"))
> p.fit[order(p.fit[,26]),]
> res<-residuals(mod.rasch, type="LD")
(abs(res)-1)/sqrt(2)

> LDX2<-NULL
> for (i in 1:ncol(res)){
  for (j in 1:nrow(res)){
    if (i != j) {LDX2<-rbind(LDX2, cbind(abs(res[j,i]),
                                         res[j,i], colnames(res)[i],
rownames(res)[j]))}
  } }
> data.frame(LDX2[order(LDX2[,1], decreasing=T),][,-1])
> CV<-NULL
for (i in 1:nrow(res)){
  for (j in 1:ncol(res)){
    if (i != j) {CV<-rbind(CV, cbind(abs(res[i,j]), res[i,j],
                                         rownames(res)[i],
colnames(res)[j]))}
  } }
> data.frame(CV[order(CV[,1], decreasing=T),][,-1])

```

Lampiran 02 : Script Analisis Karakteristik Butir Dikotomus Model 1 PL

##ANALISIS BUTIR TES MODEL 1PL

```
> library(ltm)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c1PL <- read.table("D10N460B30.39.txt")
> mod.1pl <- rasch(c1PL)
> coef.rasch(mod.1pl)
> mod.1pl$conv
> factor.scores(mod.1pl, method="EAP")
```

##Kurva Karakteristik Item dan Fungsi Informasi

```
> plot(mod.1pl, legend=T)
> plot(mod.1pl, item=1:2, legend=T)
> plot(mod.1pl, item=c(1,3,5), legend=T)
> plot(mod.1pl, type="IIC", item=3, legend=T)
> plot(mod.1pl, type="IIC", item=c(1,3,5), legend=T)
> plot(mod.1pl, type="IIC", item=0, legend=T)
```

##Analisis Kecocokan Model

```
> item.fit(mod.1pl)
> person.fit(mod.1pl)
> p.fit <- person.fit(mod.1pl, resp.patterns=c1PL)
> p.fit <- cbind(1:460, c1PL, p.fit$p.values)
> p.fit[order(p.fit[,22]),]
```

##Model-data fit and model comparison

```
> mod.2pl <- ltm(c1PL~z1, IRT.param=T)
> anova(mod.1pl, mod.2pl)
> GoF.rasch(mod.1pl, B = 100)
```

##ANALISIS BUTIR IRT 1PL dengan Paket MIRT

```
> spec<-' F = 1-30
> CONSTRAIN=(1-30,a1) '
> mod.1pl<-mirt(c1PL, model=spec, itemtype="2PL", SE=T)
> mod.1pl
> extract.mirt(mod.1pl, what="converged")
> extract.mirt(mod.1pl, what="secondordertest")
> coef(mod.1pl, IRTpars=T, simplify=T)
> coef(mod.1pl, IRTpars=T)
> coef(mod.1pl, IRTpars=T, printSE=T)
> round (sqrt (diag (extract.mirt(mod.1pl, what="vcov"))), 3)
```


Lampiran 03: Script Analisis Karakteristik Butir Dikotomus Model 2 PL

##R Script untuk IRT 2 PL

```
> library(ltm)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c2PL <- read.table("D10N460B30.1.txt")
> mod.2pl <- ltm(c2PL~z1, IRT.param=T)
> mod.2pl$conv
> summary(mod.2pl)
> coef(mod.2pl)
> factor.scores(mod.2pl, method="EAP")
> factor.scores(mod.2pl, method="EAP", resp.patterns = c2PL)
```

##ICC dan IIC

```
> plot(mod.2pl, legend=T)
> plot(mod.2pl, item=1:2, legend=T)
> plot(mod.2pl, item=c(1,3,5), legend=T)
> plot(mod.2pl, type="IIC", item=3, legend=T)
> plot(mod.2pl, type="IIC", item=c(1,3,5), legend=T)
> plot(mod.2pl, type="IIC", item=0, legend=T)
```

##UJI KECOCOKAN MODEL

```
> item.fit(mod.2pl)
> item.fit(mod.2pl, FUN = mean)
> person.fit(mod.2pl)
> person.fit(mod.2pl, resp.patterns=c2PL)
> mod.1pl<-rasch(c2PL)
> anova(mod.1pl, mod.2pl)
```

##ANALISIS IRT 2PL dengan Paket MIRT

```
> library(mirt)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c2PL <- read.table("D10N460B30.1.txt")
```

##Estimasi Parameter Butir dan Orang

```
> mod.2pl<-mirt(c2PL, model=1, itemtype="2PL", SE=T)
> mod.2pl
> extract.mirt(mod.2pl, what="converged")
> extract.mirt(mod.2pl, what="secondordertest")
> coef(mod.2pl, IRTpars=T, simplify=T)
> coef(mod.2pl, IRTpars=T)
> coef(mod.2pl, IRTpars=T, printSE=T)
> round(sqrt(diag(extract.mirt(mod.2pl, what="vcov"))), 3)
> fscores(mod.2pl, method="EAP", full.scores=T, full.scores.SE = T)
```

##ICC dan IIC

```
> itemplot(mod.2pl, 3) #2.2_37
> plot(mod.2pl, type = "trace")
> plot(mod.2pl)
> plot(mod.2pl, type="info")
> itemplot(mod.2pl, 3, type="info")
```

##ANALISIS KECOCOKAN MODEL

```
> M2(mod.2pl)
> itemfit(mod.2pl)
> itemfit(mod.2pl, empirical.plot=3, empirical.CI=.95)
> personfit(mod.2pl, method="EAP")
> residuals(mod.2pl, type="LD")
> residuals(mod.2pl, type="Q3")
> p.fit <- cbind(1:460, c2PL, personfit(mod.2pl,
+                                     method="EAP"))
> p.fit[order(p.fit[,22]),]
> rrr<-residuals(mod.2pl, type="LD")
> (abs(rrr)-1)/sqrt(2)
```

Lampiran 04: Script Analisis Karakteristik Butir Dikotomus Model 3PL

##R Script untuk IRT 3 PL menggunakan paket mirt

```
> library(mirt)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c3PL <- read.table("D10N460B30.1.txt")
> mod.3pl<-mirt(c3PL, model=1, itemtype="3PL", SE=T)
> mod.3pl
> extract.mirt(mod.3pl, what="converged")
> extract.mirt(mod.3pl, what="secondorderstest")
> coef(mod.3pl, IRTpars=T, simplify=T)
> coef(mod.3pl, IRTpars=T)
> coef(mod.3pl, IRTpars=T, printSE=T)
> round (sqrt (diag (extract.mirt(mod.3pl, what="vcov"))) ), 3)
> coef(mod.3pl, printSE=T)
> fscores(mod.3pl, method="EAP", full.scores=T, full.scores.SE = T)
```

##ICC dan IIC

```
> itemplot(mod.3pl, 3)
> plot(mod.3pl, type = "trace")
> plot(mod.3pl)
> plot(mod.3pl, type="info")
> itemplot(mod.3pl, 3, type="info")
> sum(coef(mod.3pl, IRTpars=T, simplify=T)$items[,3])
```

##Model Data FIT

```
> M2(mod.3pl)
> itemfit(mod.3pl)
> itemfit(mod.3pl, empirical.plot=15, empirical.CI=.95)
> personfit(mod.3pl, method="EAP")
> residuals(mod.3pl, type="LD")
> residuals(mod.3pl, type="Q3")
> p.fit <- cbind(1:460,c3PL, personfit(mod.3pl, method="EAP"))
> p.fit[order(p.fit[,32]),]
```

Lampiran 05: Script Analisis Karakteristik Butir Dikotomus Model 4PL

##R Script untuk IRT 4 PL menggunakan paket mirt

```
> library(mirt)
> setwd("D:/Eksend/Kelas10/D10N460/D10N460B30")
> c4PL <- read.table("D10N460B30.1.txt")
> mod.4pl<-mirt(c4PL, model=1, itemtype="4PL", SE=T)
> mod.4pl
> extract.mirt(mod.4pl, what="converged")
> extract.mirt(mod.4pl, what="secondordertest")
> coef(mod.4pl, IRTpars=T, simplify=T)
> coef(mod.4pl, IRTpars=T)
> coef(mod.4pl, IRTpars=T, printSE=T)
> round(sqrt(diag(extract.mirt(mod.4pl, what="vcov"))), 3)
> coef(mod.4pl, printSE=T)
```

##Estimasi Kemampuan orang

```
> fscores(mod.4pl, method="EAP", full.scores=T, full.scores.SE = T)
```

##ItemPlot

```
> itemplot(mod.4pl, 1)
> plot(mod.4pl, type = "trace")
> plot(mod.4pl)
> plot(mod.4pl, type="info")
> itemplot(mod.4pl, 3, type="info")
```

##Model Data Fit

```
> M2(mod.4pl)
> itemfit(mod.4pl)
> itemfit(mod.4pl, empirical.plot=15, empirical.CI=.95)
> personfit(mod.4pl, method="EAP")
> residuals(mod.4pl, type="LD")
> residuals(mod.4pl, type="Q3")
> p.fit <- cbind(1:460,c4PL, personfit(mod.4pl, method="EAP"))
> p.fit[order(p.fit[,32]),]
```

Lampiran 06: Script Analisis Faktor

```
> library(corpcor)
> library(GPARotation)
> library(psych)
> library(MASS)
> library(parallel)
> library(factoextra)
> library(FactoMineR)
> library(MPpsychOR)

> D8P23 <- read.table("D:/Eksend/Kelas8/D8N460/D8N460B30/
+   D8N460B30.23.txt", header = T)
> D8P23
> head(D8P23,5)
> tail(D8P23,5)
> corfact <- cor(D8P23)
> corfact
> round(corfact,3)
> attach(D8P23)
> KMO(D8P23)
> cortest.bartlett(D8P23,n=460)
> f3 <- fa(D8P23,nfactors=3, SMC=TRUE, min.err = .001,
+   max.iter=100, fm="ml", rotate = "none", n.obs = 460)
> print(f3,digits=2)
> parD8P23 <- fa.parallel(D8P23, fa="pc", n.iter=500,
+   ylabel = "Eigenvalues", quant = .50)
> scree(D8P23, pc="TRUE", factors="TRUE", hline="-1",
+   main = "Scree Plot")
> pca <- PCA(D8P23)
> pca
> fviz_eig(pca,addlabels =TRUE, ylim=c(0,70))
```

INDEKS

- Conditional Maximum
Likelihood, 20, 33, 34,
50, 52, 64
- Estimasi Parameter Butir,
29, 32, 45, 50, 74, 114, 4
- Estimasi Parameter
Kemampuan, 56, 108
- fungsi karakteristik butir,
1, 3, 5, 6
- Independensi lokal, 2, 6,
23, 26
- Invariansi, ii, 4
- IRT 1 parameter logistik,
27, 29
- IRT dikotomus, 25
- item characteristic curve, 1,
6
- Item Response Theory*, 1,
124, 125, 126, 10
- Joint Maximum Likelihood,
20, 29, 33
- laten trait, 1
- Layar Console, 13
- Marginal Likelihood Ratio,
20
- Marginal Maximum
Likelihood, 20, 33, 64, 73
- Model 2 Parameter Logistik,
ii, 31, 81
- Model 3 Parameter Logistik,
ii, 40, 100
- model 4 parameter logistik,
25, 44, 45
- Packages, ii, 15, 19
- parameter butir, 4, 20, 23,
27, 30, 31, 33, 34, 35,
39, 41, 45, 46, 50, 52,
53, 64, 91, 92, 100, 106,
110, 115, 116
- R, i, ii, 7, 8, 9, 10, 13, 14,
15, 16, 17, 18, 19, 20,
21, 45, 50, 51, 53, 54,
55, 56, 57, 58, 60, 63,
64, 73, 81, 99, 114, 124,
127, 128, 1, 4, 6, 7, 10,
12
- Rasch, ii, 2, 3, 5, 23, 25,
26, 27, 28, 29, 30, 32,
34, 50, 52, 55, 59, 61,
62, 63, 64, 65, 68, 71,
73, 74, 75, 99, 128, 1
- Rstudio, 13, 14, 15, 16, 50,
51
- script, 15, 20, 50, 51, 52,
56, 57, 58, 59, 62, 63,
68, 69, 70, 73, 74, 75,
80, 91, 97, 101, 103, 105,
108, 109, 114, 116, 119,
120
- teori respons butir, 1, 2, 5,
See
- teori tes, i, 1, 5, *See*
- unidimensi, 1, 6, 21, 22,
30, 39, 93, 107
- variabel laten, 4
- variabel manifes, 4

GLOSARIUM

- Dikotomus : Partisi atau pembagian dari suatu keseluruhan (misalnya dalam himpunan, ruang sampel, dan sebagainya) menjadi dua bagian (subset atau himpunan bagian).
- Respons dikotomus : Jawaban atau respons yang hanya terdiri dari dua kemungkinan, benar (1) atau salah (0).
- Teori Respons Butir : Teori alternatif pilihan yang bertujuan melepaskan diri dari ketergantungan tes yang diberikan dengan sampel peserta tes. Dalam hal ini walaupun soal-soal tersebut dikerjakan oleh siswa yang pandai atau siswa yang kurang pandai, indikasi tingkat kesukaran suatu soal tetap tidak berubah
- Tes Pilihan Ganda : Tes yang disusun dengan pilihan beberapa jawaban yang memuat hanya satu jawaban benar.
- Estimasi Parameter : Perkiraan atau pendugaan atau taksiran terhadap karakteristik populasi (parameter) dengan menggunakan karakteristik sampel (statistik)
- Asumsi : Anggapan yang belum terbukti kebenarannya dan memerlukan pembuktian secara langsung. Pembuktian ini disebut sebagai prasyarat, sehingga uji asumsi juga bisa bermakna uji prasyarat.
- Politomus : Parameter ambang batas kategori (τ_j) memberikan perkiraan jarak psikologis antara titik skala yang tidak dipengaruhi dari parameter item.
- Independensi lokal : Jawaban peserta tes terhadap suatu butir soal tidak bergantung pada jawaban peserta tes butir soal yang lain.
- Unidimensi : merujuk pada sifat laten (latent trait) atau abilitas yang diukur dalam suatu perangkat soal adalah tunggal. Idealnya, setiap butir tes yang dibuat

- hanya mengukur salah satu dari kemampuan peserta tes, bukan mengukur dua atau lebih kemampuan peserta tes.
- Invariansi : Karakteristik butir tetap atau tidak berubah sekalipun subpopulasi peserta tes yang menjawab butir yang sama itu berubah-ubah.
- Tingkat kesulitan : Proporsi antara banyaknya peserta tes yang menjawab butir soal dengan benar dengan banyaknya peserta tes
- Daya beda : Kemampuan suatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang kurang pintar (berkemampuan rendah)
- Guessing (tebakan) : Perilaku menebak peserta tes pada suatu butir tes tertentu yang bisa terjadi jika kemampuan peserta tes rendah mampu menjawab butir tes yang sulit.
- Kecerobohan (*carelessness*) : Perilaku tidak teliti yang menyebabkan peserta yang memiliki kemampuan tinggi tidak mampu menjawab butir soal yang tingkat kesulitannya rendah.
- Uji kecocokan model : hal yang ini mengacu pada seberapa baik model statistik memenuhi serangkaian pengamatan. Statistik uji chi-square digunakan untuk melihat apakah proporsi yang diamati dalam dua atau lebih kategori berbeda secara signifikan dari proporsi apriori atau yang diharapkan secara teoritis.
- Model Rasch : Model yang memberikan informasi apakah pola respons (jawaban) peserta tes memang secara ideal menggambarkan bahwa orang yang mempunyai abilitas tinggi memberikan pola jawaban terhadap item sesuai dengan tingkat kesulitannya.
- Model 1PL : Model teori respons butir yang hanya memuat satu parameter yaitu parameter tingkat kesulitan yang dilambangkan dengan (b).

- Model 2PL : Model teori respons butir yang memuat dua parameter yaitu parameter tingkat kesulitan (b) dan daya beda (a).
- Model 3PL : Model teori respons butir yang memuat tiga parameter yaitu parameter tingkat kesulitan (b), daya beda (a) dan tebakan/guessing (c).
- Model 4PL : Model teori respons butir yang memuat empat parameter yaitu parameter tingkat kesulitan (b), daya beda (a), tebakan/guessing (c) dan kecerobohan/carelessness (u).
- Validitas : Kemampuan alat ukur untuk mengukur apa yang hendak diukur.
- Reliabilitas : Sejauh mana pengukuran dari suatu tes tetap konsisten setelah dilakukan berulang-ulang terhadap subjek dan dalam kondisi yang sama.
- Nilai Eigen : Suatu nilai yang menunjukkan seberapa besar pengaruh suatu variabel terhadap pembentukan karakteristik sebuah vektor atau matriks. eigen value dinotasikan dengan λ .

Biodata Penulis



Muhammad Ali Gunawan Lahir di Pancor-Lombok Timur Provinsi Nusa Tenggara Barat, 12 Maret 1980. Pendidikan: SD Negeri 05 Pancor, MTs Nahdlatul Wathan (NW) Pancor, SMA Negeri 1 Sakra, S1 Pendidikan Matematika P.MIPA STKIP HAMZANWADI Selong, S2 Penelitian dan Evaluasi Pendidikan Universitas Pendidikan Ganesha (UNDIKSHA) Singaraja-Bali, S3

Penelitian dan Evaluasi Pendidikan (PEP-konsentrasi pengukuran dan pengujian) Universitas Negeri Yogyakarta.

Pernah bekerja sebagai staf full-timer pada Proyek Desentralisasi Pendidikan Dasar (DBEP) Asian Development Bank (ADB) Dinas Pendidikan Kabupaten Lombok Timur pada Tahun 2002 sampai dengan 2004. Sejak tahun 2010 sampai dengan tahun 2019 mengabdikan diri sebagai tenaga pengajar pada Program Studi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan (FKIP) Universitas Pekalongan.

Hibah penelitian yang pernah dimenangkan: 1) Kaji tindak pemberdayaan ekonomi masyarakat pesisir melalui swamitramina di desa Tanjung Luar Kabupaten Lombok Timur (Dana Hibah Penelitian Pemkab. Lombok Timur T.A 2007); 2) Analisis Dampak Pemberian Bantuan Asian Development Bank (ADB) Terhadap Peningkatan Mutu Pendidikan Dasar di Kabupaten Lombok Timur (Dana Hibah Penelitian Pemkab. Lombok Timur T.A. 2008), 3) Pemberdayaan petani miskin di lahan marginal melalui kaji tindak agrobisnis pertanian lahan kering berbasis agrosilvopastoral teknik SPB (Agroforestri) di kecamatan Wanasaba (Dana Hibah Penelitian Pemkab. Lombok Timur Tahun 2008); 4) Evaluasi Pelaksanaan Kurikulum Tingkat Satuan Pendidikan (KTSP) Sekolah Menengah Pertama (SMP-MTs) di Kota Pekalongan (Dana Hibah Penelitian Pemkot. Pekalongan T.A. 2011), 5) Faktor-Faktor Kesulitan Belajar dan Dampaknya Terhadap Motivasi Berprestasi Pada Mata Kuliah Statistik Matematika I (Dana Hibah Penelitian Dosen Pemula DP2M Dikti Tahun 2013); 6) Peningkatan kualitas dan hasil produksi penyulingan minyak Nilam dengan metode penguapan langsung pada UMKM Nilam Sari desa Sukoharjo kec. Kandang Serang kabupaten Pekalongan (Dana Hibah Lembaga Ilmu Pengetahuan Indonesia_LIPI. T.A. 2014).

Buku yang pernah diterbitkan di antaranya: 1) Statistik Penelitian Untuk Pendidikan (2013); 2) Statistik Penelitian Bidang Pendidikan, Psikologi dan Sosial: Dilengkapi dengan Contoh Manual dan SPSS (2015), 3) Panduan Praktis Analisis Instrumen Penilaian Menggunakan Item Response Theory (IRT) (2020); 4) Statistika Terapan S1 PGSD & S2 Dikdas: Dilengkapi Contoh Kasus, Analisis Berbasis Software SPSS dan Pembuatan Instrumen (2020); 5) Penelitian Tindakan Kelas: Landasan Filosofis dan Praktis (2022); 6) Penyetaraan Tes Literasi Matematika Secara Klasik, Modern dan Kernel Dengan R (Mahasiswa, Guru dan Pengembang Tes).

BIODATA PENULIS



Heri Retnawati, lahir di Gunungkidul, 3 Januari 1973, sekarang mengajar di UNY setelah menempuh pendidikan sarjana dan pasca sarjana di Universitas Negeri Yogyakarta. Bidang keahlian: Asesmen pendidikan matematika dan implementasinya untuk perbaikan pembelajaran. Penelitian-penelitian yang terkait dengan pembelajaran dan asesmen banyak dilakukan. Beberapa judul buku telah dituliskan terkait dengan pembelajaran matematika, asesmen pendidikan, teori pengukuran, dan juga metode penelitian.

BIODATA PENULIS



Ari Setiawan., Lahir di bawah kaki gunung lawu, tawangmangu karanganyar, menghabiskan masa studi sd sampai SMA di kampong halaman lalau tahun 2000 hijrah ke jogja melanjutkan Pendidikan di UIN dan UNY. Lulusan doktor bidang penelitian dan evaluasi Pendidikan tahun 2017. Saat ini aktif sebagai dosen Pascasarjana PEP UST, TEKNIK, PGSD dan Juga PVKK dan PBSI. Selain itu aktif menulis buku dan artikel jurnal. Saat ini mengemban amanah sebagai ketua Asosiasi kolaborasi dosen lintas negara (CeL KODELNO, juga aktif di BAN SM DIY dan Juga Majelis DIKDASMEN. Email: ari.setiawan@ustjogja.ac.id dan Hp. 081228153789.

Biodata Penulis



Amalia Fitri, M.Pd. Lahir di Pekalongan Provinsi Jawa Tengah, 28 Januari 1986. Pendidikan: MIS NU Samborejo, SMP Negeri 1 Tirto, SMA Negeri 1 Pekalongan, S1 Pendidikan Matematika Universitas Negeri Semarang (UNNES), S2 Pendidikan Matematika Universitas Negeri Semarang (UNNES). Sejak tahun 2009 sampai dengan sekarang mengabdikan diri sebagai tenaga pengajar pada Program Studi Pendidikan

Matematika Fakultas Keguruan dan Ilmu Pendidikan (FKIP) Universitas Pekalongan.

Hibah penelitian yang pernah diperoleh: 1) Faktor-Faktor Kesulitan Belajar dan Dampaknya Terhadap Motivasi Berprestasi Pada Mata Kuliah Statistik Matematika I (Dana Hibah Penelitian Dosen Pemula DP2M Dikti Tahun 2013); 2) Penerapan Project based Learning berbasis Penilaian Portofolio terhadap Prestasi Belajar Mahasiswa Mata Kuliah Matematika Ekonomi (Dana Hibah Penelitian Dosen Pemula DP2M Dikti Tahun 2015); 3) Penerapan Flipped Classroom Berbantuan Video Pembelajaran terhadap Kemampuan Penalaran Siswa Kelas VII (Dana Hibah Penelitian Dosen Pemula DP2M Dikti Tahun 2017).

Buku yang pernah diterbitkan, di antaranya, 1) Penelitian Tindakan Kelas: Landasan Filosofis dan Praktis (2022); 2) Penyetaraan Tes Literasi Matematika Secara Klasik, Modern dan Kernel Dengan R (Mahasiswa, Guru dan Pengembang Tes) (2022).