PENAFSIRAN

HASIL ANALISIS ITEMAN 4.30

Oleh: Ali Muhson

FAKULTAS EKONOMI

UNIVERSITAS NEGERI YOGYAKARTA

2020

Introduction

This report provides the results of a classical item and test analysis by the computer program Iteman Version 4.3 (Assessment Systems Corporation, 2013) for Latihan ITEMAN. The output is divided into three sections:

1. Specifications
2. Summary statistics
3. Item-by-item results.

The statistical output is also recorded in a comma-separated value (CSV) file of the same name.

Specifications

The Windows paths for the input files used in this analysis were:

D:\@MyData\Mengajar\Evaluasi\Iteman\Anabut.txt

The Windows paths for the output files produced by this analysis were:

Hasil1
Hasil1.csv
Hasil1 Scores.csv

Table 1 presents the specifications and basic information concerning the analysis. This provides important documentation of the setup of the program for historical purposes.

Table 1: Specifications

Specification	Value	Specification	Value
Number of examinees	60	Total Items	50
Scored Items	50	Pretest Items	0
Multiple Choice Items	50	Polytomous Items	0
Number of domains	1	External scores	No
Minimum P	0.00	Maximum P	1.00
Minimum item mean	0.00	Maximum item mean	15.00
Minimum item correlation	0.00	Maximum item correlation	1.00
ITEMAN 3.0 Header	Yes	Exclude omits from option statistics	No
Number of ID columns	4	ID begins in column	1
Responses begin in column	5	Omit character	0
Not Admin character	N	Produce quantile tables	Yes
Correct for spuriousness	Yes	Produce quantile plots	Yes
Save data matrix	No	Include omit codes in matrix	N/A
Include Not Admin codes in matrix	N/A	Include scaled scores for	Total Score
Scaling function	N/A	Scaled score setting 1	N/A
Scaled score setting 2	N/A	Dichotomous Classification	No
Classify based on	N/A	Cutpoint	N/A
Low group label	Low	High group label	High
Data is delimited by	N/A	Test for DIF	No
Group status is in column	N/A	Ability levels for DIF	N/A

Group 1 code	N/A	Group 2 code	N/A
Group 1 label	N/A	Group 2 label	N/A

Summary statistics

Table 2 presents the summary statistics of the test, for the scored items. Definitions of these statistics are found in the Iteman manual.

Table 2: Summary statistics

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
Scored Items	50	29.000	5.035	15	39	0.581	0.167
Scaled Total	50	0.000	0.000	0.000	0.000	-	-

Keterangan:

- *Item menunjukkan ada 50 butir soal yang dianalisis*
- *Mean menunjukkan rata-rata skor yang diperoleh peserta tes adalah 29,00*
- *SD menunjukkan standar deviasi skor yang diperoleh peserta tes adalah 5,035*
- *Min/Max Score adalah skor minimal dan maksimal yang diperoleh peserta tes*
- *Mean P menunjukkan rata-rata tingkat kesukaran soal adalah 0,581 (tingkat kesukaran soal dalam kategori sedang)*
- *Mean Rpbis menunjukkan rata-rata tingkat daya beda soal adalah 0,167 (soal termasuk tidak baik daya bedanya)*

Table 3 presents a reliability analysis of the tests. Alpha (also known as KR-20) is the most commonly used index of reliability, and is therefore used to calculate the standard error of measurement (SEM) on the raw score scale. Also presented are three configurations of split-half reliability, first as uncorrected correlations, and then as Spearman-Brown (S-B) corrected correlations. This is because an uncorrected split-half correlation is referenced to a "test" that only contains half as many items as the full test, and therefore underestimates reliability.

Table 3: Reliability

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0.661	2.931	0.525	0.423	0.355	0.689	0.594	0.524

Keterangan:

- *Alpha adalah angka koefisien reliabilitas Alpha Cronbach*
- *SEM adalah Standard Error of Measurement atau kesalahan standar pengukuran*
- *Split-Half (Random) adalah angka koefisien reliabilitas teknik belah dua secara random*
- *Split-Half (First-Last) adalah angka koefisien reliabilitas teknik belah dua (belahan awal dan akhir)*

- **S-B (Random)** *adalah angka koefisien reliabilitas teknik Spearman-Brown secara random*
- **S-B (First-Last)** *adalah angka koefisien reliabilitas teknik Spearman-Brown belah dua awal dan akhir*
- **S-B (Odd-Even)** *adalah angka koefisien reliabilitas teknik Spearman-Brown belah dua ganjil dan genap*

Table 4 presents the item statistics and flags for the item(s) that were flagged during the analysis

Table 4: Summary Statistics for the Flagged Items

Item ID	P / Item Mean	R	Flag(s)
2	0.700	0.003	K
5	0.417	0.010	K
6	0.233	0.107	K
7	0.233	-0.061	K, LR
8	0.683	-0.043	K, LR
9	0.483	0.034	K
10	0.783	-0.075	K, LR
13	0.600	0.187	K
14	0.450	0.028	K
15	0.683	0.155	K
21	0.450	0.111	K
27	0.600	0.081	K
31	0.305	-0.051	K, LR
32	0.700	0.025	K
35	0.183	-0.061	K, LR
43	0.700	0.070	K
44	0.814	0.082	K
45	0.633	0.000	K
47	0.367	-0.097	K, LR
50	0.797	0.052	K

Keterangan:

- **Item ID** *adalah identitas butir soal (nomor butir soal)*
- **P/Item Mean** *adalah tingkat kesukaran butir*
- **R** *adalah koefisien korelasi point biserial atau daya beda butir*
- **Flag(s)** *adalah keterangan, misalnya K = Key Error (kunci jawaban yang diberikan salah), LR = Low Rpbis (daya beda rendah), LP = Low Proportion correct (tingkat kesukaran rendah), dsb*

Figure 1 displays the distribution of the raw scores for the scored items across all domains.
Table 5 displays the frequency distribution for total score shown in Figure 1.

Figure 1: Total score for the scored items
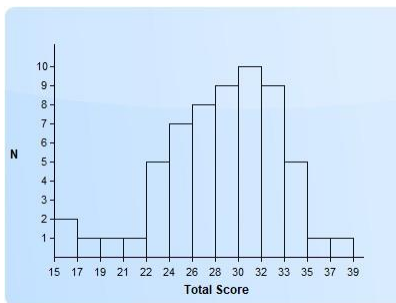
Table 5: Frequency Distribution for Total Score

Range	Frequency
14 to 16	2
17 to 18	1
19 to 20	1
21 to 22	1
23 to 24	5
25 to 26	7
27 to 28	8
29 to 30	9
31 to 32	10
33 to 34	9
35 to 36	5
37 to 38	1
39	1

Figure 2 displays the distribution of the P values for the dichotomously scored items (correct/incorrect). Table 6 displays the frequency distribution of the P values shown in Figure 2.

Figure 2: P values for the scored items
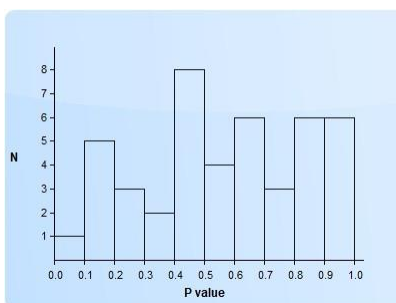
Table 6: Frequency Distribution for the P values

Score	Frequency
0.0 to 0.1	1
0.1 to 0.2	5
0.2 to 0.3	3
0.3 to 0.4	2
0.4 to 0.5	8
0.5 to 0.6	4
0.6 to 0.7	6
0.7 to 0.8	3
0.8 to 0.9	6
0.9 to 1.0	6

Figure 3 displays the distribution of the Point-Biserial Correlations for the dichotomously scored items (correct/incorrect).   Table 7 displays the frequency distribution of the Point-Biserial correlations shown in Figure 3.

Figure 3: Rpbis for the scored items
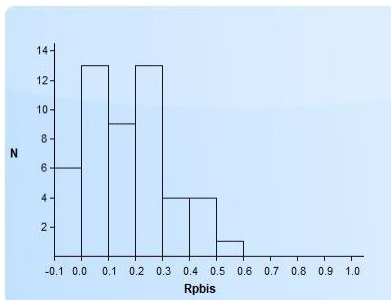
Table 7: Frequency Distribution for the Rpbis

Score	Frequency
-0.1 to 0.0	6
0.0 to 0.1	13
0.1 to 0.2	9
0.2 to 0.3	13
0.3 to 0.4	4
0.4 to 0.5	4
0.5 to 0.6	1
0.6 to 0.7	0
0.7 to 0.8	0
0.8 to 0.9	0
0.9 to 1.0	0

Figure 4 displays the scatterplot of P (difficulty) by Rpbis (discrimination) for the dichotomously scored items (correct/incorrect).
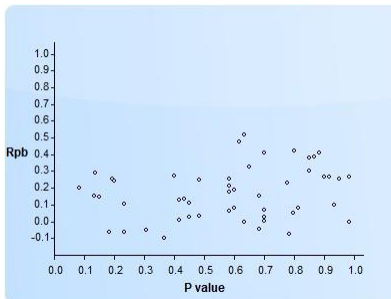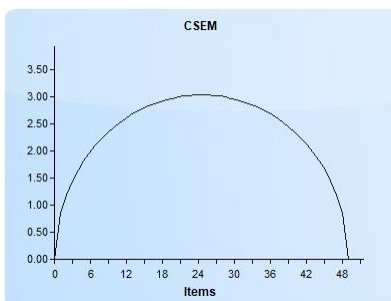
Figure 4: P by Rpbis

Figure 5 displays a graph of the Conditional Standard Error of Measurement (CSEM) Formula IV.

Figure 5: CSEM

Item-by-item results

The following section presents the item-by-item results of the analysis.   Each item has several tables and a figure.   The figure, called a quantile plot, shows the proportion of examinees selecting each option, for consecutive segments of the examinees as ranked by score.   The key thing to evaluate in this figure is that the line for the correct answer has a positive slope (goes up from left to right), which means that examinees with higher scores tend to answer correctly more often.   Conversely, the lines for the incorrect options, called distractors, should have a negative slope.   Note, however, that the use of a small number of groups (e.g., 3 or fewer) oversimplifies the graph, so that items which are very difficult or very easy (that is, discriminating in only the top or bottom 20% of examinees) might appear to have poor quantile plots and classical statistics.   For such items, item response theory presents significant advantages in analysis

There are four tables presented for each item.

1. Item information table: records the information supplied by the control file (or Iteman 3 header) for this item.
2. Item statistics table: overall item statistics.
3. Option statistics: detailed statistics for each item, which helps diagnose issues in items with poor statistics.
4. Quantile plot data: the values used to create the quantile plot.

The item statistics table presents overall item statistics in the first row of numbers.   The two most important item-level statistics for dichotomously scored (correct/incorrect) items are the P value and the point-biserial correlation, which represent the difficulty and discrimination of the item, respectively.   For polytomously

scored (rating scale or partial credit) items, the difficulty is represented by the mean (average) item score, while the discrimination is represented by a Pearson r correlation.

The P value is the proportion of examinees that answered an item in the keyed direction.   P ranges from 0 to 1.   A high value (0.95) means that an item is easy, a low value (0.25) means that the item is difficult. The point-biserial correlation (Rpbis) is a measure of the discriminating, or differentiating, power of the item. Rpbis ranges from -1 to 1.   A negative Rpbis is indicative of a bad item as lower scoring examinees are more likely than higher scoring examinees to respond in the keyed direction.

For rating scale or partial credit items, the mean item score ranges from the minimum to the maximum of the scale.   For example, if the item has a rating scale of 1 to 5, the possible range for the mean is 1 to 5. The Pearson r is similar to the Rpbis in that it ranges from -1 to 1, with a positive r indicating that the item correlates well with total score.

The option statistics table presents statistics for each individual option (alternative).   The key thing to examine in this portion of the table is that no distractors have a higher Rpbis than the correct answer.   That indicates that higher scoring examinees are selecting the incorrect answer, which therefore might be arguably correct.

The quantile plot data table simply presents the values calculated to create the quantile plot.   Because it contains the same information, the quantile plot itself presents a useful picture of the item's performance, but this table can be used to examine that performance in detail to help diagnose possible issues.
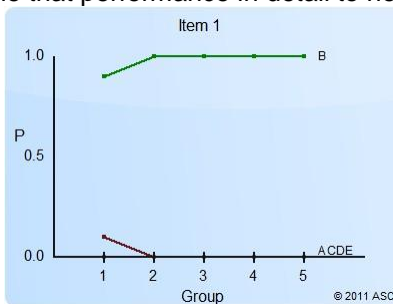
Item information

Seq.	ID	Key	Scored	Num Options	Domain	Flags
4	4	A	Yes	5	1	

Keterangan:

- *Seq. adalah the item sequence number (nomor urut butir)*
- *ID adalah identitas butir soal (nomor butir soal)*
- *Key adalah kunci jawaban*
- *Scored adalah status penskoran (ya atau tidak)*
- *Num. Options adalah jumlah alternatif jawaban.*
- *Domain adalah aspek yang diukur (misalnya test tersebut mengukur beberapa aspek)*
- *Flags adalah keterangan, misalnya K = Key Error (kunci jawaban yang diberikan salah), LR = Low Rpbis (daya beda rendah), LP = Low Proportion correct (tingkat kesukaran rendah), dsb*

Item statistics

N	P	Total Rpbis	Total Rbis	Alpha w/o
60	0.867	0.388	0.612	0.645

Keterangan:

- *N adalah jumlah peserta tes*
- *P adalah tingkat kesukaran butir*
- *Total Rpbis adalah daya beda butir yang diukur dengan korelasi point-biserial*
- *Total Rbis adalah daya beda butir yang diukur dengan korelasi biserial*
- *Alpha w/o adalah koefisien reliabilitas instrumen (koefisien alpha) jika butir tersebut dihilangkan*

Penafsiran:

*Berdasarkan pada tabel **item statistics** menunjukkan bahwa tingkat kesukaran butir nomor 4 adalah 0,867 sehingga butir tersebut termasuk soal yang mudah. Koefisien daya beda sebesar 0,388 sehingga daya bedanya baik*

Kriteria:

Untuk menentukan tingkat kesukaran (P) didasarkan pada kriteria berikut:
- *Lebih dari 0,70 → Mudah*
- *0,30 – 0,70 → Sedang*
- *Kurang dari 0,30 → Sulit*

Untuk menentukan daya beda butir didasarkan pada kriteria berikut:
- *Lebih dari 0,30 → daya beda baik*
- *0,20 – 0,30 → daya beda cukup baik*
- *Kurang dari 0,20 → daya beda tidak baik*

Option statistics

Option	N	Prop.	Rpbis	Rbis	Mean	SD	Color	
A	52	0.867	0.388	0.612	29.865	4.653	Maroon	**KEY**
B	6	0.100	-0.321	-0.549	23.500	4.506	Green	
C	0	0.000	--	--	--	--	Blue	
D	1	0.017	-0.004	-0.011	28.000	0.000	Olive	
E	1	0.017	-0.274	-0.847	18.000	0.000	Gray	
Omit	0							
Not Admin	0							

Keterangan:

- *Option* adalah alternatif jawaban
- *N* adalah jumlah peserta tes yang memilih alternatif jawaban
- *Prop.* adalah proporsi peserta tes yang memilih alternatif jawaban
- *Rpbis* adalah daya beda butir yang diukur dengan korelasi point-biserial untuk masing-masing alternatif jawaban
- *Rbis* adalah daya beda butir yang diukur dengan korelasi biserial untuk masing-masing alternatif jawaban
- *Mean* adalah rata-rata skor peserta tes yang memilih alternatif jawaban
- *SD* adalah standar deviasi skor peserta tes yang memilih alternatif jawaban
- *Color* adalah warna alternatif jawaban yang tertuang di plot kuartil
- *Key* adalah kunci jawaban

Penafsiran:

*Berdasarkan pada tabel **option statistics** menunjukkan bahwa kunci jawaban untuk soal nomor 4 adalah A, sedangkan proporsi peserta tes yang menjawab masing-masing alternatif jawaban B=0,100, C=0,000, D=0,017, dan E=0,017. Oleh karena proporsi yang menjawab alternatif C, D, dan E kurang dari 0,05 maka alternatif jawaban tersebut bukanlah pengecoh yang efektif, sedangkan alternatif jawaban B adalah pengecoh yang efektif karena proporsinya lebih dari 0,05.*

Kriteria:

Untuk menentukan efektif tidaknya alternatif jawaban sebagai pengecoh didasarkan pada kriteria minimal ada 5 persen peserta tes yang menjawab alternatif jawaban tersebut.

Quantile plot data

Option	N	0-20%	20-40%	40-60%	60-80%	80-100%	Color	
A	52	0.500	0.900	0.857	1.000	1.000	Maroon	**KEY**
B	6	0.400	0.100	0.071	0.000	0.000	Green	
C	0	0.000	0.000	0.000	0.000	0.000	Blue	
D	1	0.000	0.000	0.071	0.000	0.000	Olive	
E	1	0.100	0.000	0.000	0.000	0.000	Gray	