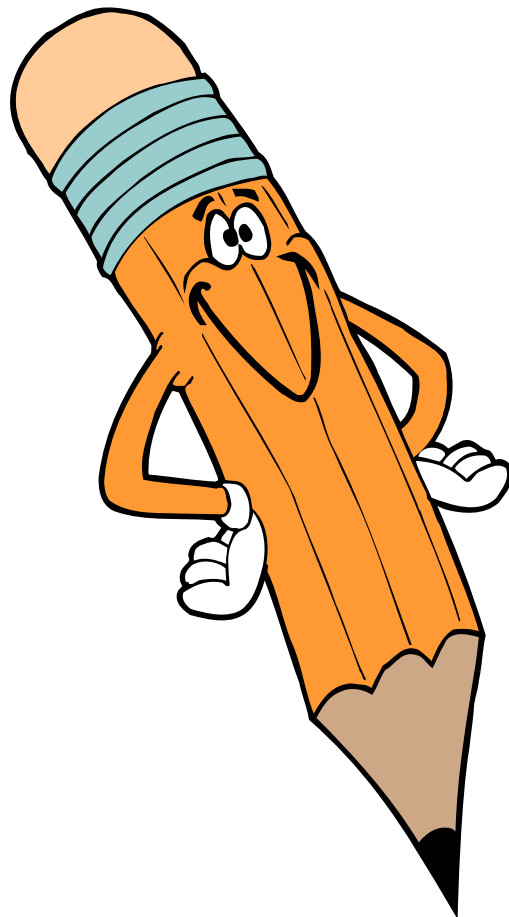


UNDERSTANDING YOUR STATISTICAL ANALYSIS



Interpreting Your Statistical Analysis
© 1997, BYU Testing Services
265 HGB, Provo, UT 84602-2700
(801) 378-6147
All Rights Reserved

Prepared for use by the faculties of:

Brigham Young University
BYU Hawaii Campus
BYU Idaho Campus
Central Michigan University
Salt Lake Community College
Utah Valley State College
Weber State University

INTRODUCTION

This guide provides information to help you interpret the Statistical Analysis portion of your test results. It will cover the major aspects of the analysis in an easy to understand format to help you assess your test results. It will not cover in detail how individual items are calculated or the full statistical meaning behind them. If this guide proves to be ineffective for your particular needs, refer to a statistics handbook.

The Statistical Analysis is made up of several parts which may vary depending on the type of test. Tests can either be 1) administered and/or graded in the Testing Center, or 2) administered in the classroom and graded in the Testing Center (also known as “BATCH TEST”).

Either test may consist of an Objective part (regular multiple choice, True/False, or any type of questions which can be answered on a bubble sheet) and/or a Non-Objective part (essay, short answer, fill in the blank, or questions which must be graded by a teacher or TA). In this guide, all parts of the analysis will be shown and described. In the exclusive case of batch tests, the instructor can choose the type of processing desired. Thus, most parts of the analysis that apply to batch tests are optional and are requested at the time of submission.

For the purpose of explaining the Statistical Analysis, a sample report has been included for the fictional, general education class *Dieting for Fun 123* (DIET 123). This test was “taken” by more than 2,000 students in the Testing Center and was comprised of both objective and non-objective parts.

As a standard format for the analysis, all pages have a heading which shows the Department Code for the class (DIET 123 in this case), Test Number, Test Form and the time and date the report was generated. Under the header, the title for the report is printed (Frequency Distributions, Item Analysis, etc.)

KEY FOR QUESTION EVALUATION

S.C.O.U.T. STATISTICAL ANALYSIS FOR DIET 123 TEST=1 FORM=A 10:04:27 07 DEC 1996

KEY FOR QUESTION EVALUATION

- A. QUESTION DISCRIMINATES VERY WELL BETWEEN STUDENTS SCORING HIGH ON THIS EXAM AND THOSE DOING POORLY.
- B. QUESTION DISCRIMINATES WELL BETWEEN STUDENTS SCORING HIGH ON THIS EXAM AND THOSE DOING POORLY.
- C. ADEQUATE DISCRIMINATION CAPABILITY IS SHOWN BY THIS QUESTION.
- D. SOME DISCRIMINATION IS SHOWN, BUT POSSIBLE AMBIGUITIES EXIST IN THIS QUESTION.
- E. A QUESTION ANSWERED CORRECTLY BY EVERYONE (THUS, NO DISCRIMINATION BETWEEN HIGH AND LOW SCORING STUDENTS IS SHOWN), OR THERE IS AN INVERSE DISCRIMINATION INDICATING DEFINITE AMBIGUITIES. IN THIS CASE, POSSIBLE REVISION OF THIS QUESTION MIGHT BE APPROPRIATE.

KEY FOR QUESTION EVALUATION

If an Item Analysis has been requested, a *Key For Question Evaluation* sheet will accompany the analysis. This explains the evaluation given each question in the objective part of the test. The key states:

Evaluation	Meaning
A	Most students who answered this question correctly did well on the exam while most students who answered it incorrectly did poorly.
B	Many students who answered this question correctly did well on the exam but some who also answered it correctly did poorly.
C	Some students who did well on the exam answered this question correctly while others who also did well on the exam answered it incorrectly. Thus, this question by itself may not be a good indication of a student's mastery of the materials.
D	There is a definite mix between the students who answered this question correctly and their total test scores, implying possible ambiguities in the question itself. Revision for this question is recommended.
E	The question was either answered correctly or incorrectly by all students. If this result was not conscientiously expected, revision is strongly recommended.

A less "formal" interpretation for these evaluations would be to consider them as "letter grades." An **A** means it was an excellent question while an **E** means it was a poor question in most circumstances. It is advisable to thoroughly analyze the Question Evaluation report to improve test quality.

FREQUENCY DISTRIBUTIONS

S.C.O.U.T STATISTICAL ANALYSIS FOR DIET 123 TEST=1 FORM=A 10:04:27 07 DEC 1996

FREQUENCY DISTRIBUTIONS											
TOTAL SCORE				# OBJECTIVE QSTNS CORRECT				NON-OBJECTIVE SCORE			
%SCORE	FREQ	CUM PCT	CUM FQ	NUMBER	FREQ	CUM PCT	CUM FQ	SCORE	FREQ	CUM PCT	CUM FQ
99	2	99.91	2309	34	17	99.26	2294	42	3	99.87	2308
98	1	99.87	2308	33	54	96.93	2240	41	22	98.92	2286
97	2	99.78	2306	32	124	91.56	2116	40	33	97.49	2253
96	8	99.44	2298	31	161	84.60	1955	39	63	94.76	2190
95	9	99.05	2289	30	198	76.03	1757	38	76	91.48	2114
94	20	98.18	2269	29	187	67.94	1570	37	95	87.36	2019
93	18	97.40	2251	28	209	58.89	1361	36	145	81.09	1874
92	20	96.54	2231	27	195	50.45	1166	35	133	75.34	1741
91	26	95.41	2205	26	183	42.54	983	34	141	69.23	1600
90	36	93.86	2169	25	174	35.01	809	33	151	62.70	1449
89	52	91.61	2117	24	140	28.95	669	32	163	55.65	1286
88	52	89.36	2065	23	119	23.80	550	31	153	49.03	1133
87	62	86.67	2003	22	96	19.65	454	30	128	43.49	1005
86	54	84.34	1949	21	83	16.05	371	29	119	38.34	886
85	61	81.70	1888	20	77	12.72	294	28	146	32.02	740
84	75	78.45	1813	19	79	9.30	251	27	104	27.52	636
83	68	75.51	1745	18	43	7.44	172	26	109	22.80	527
82	79	72.09	1666	17	49	5.32	123	25	75	19.56	452
81	80	68.63	1586	16	36	3.76	87	24	82	16.01	370
80	77	65.30	1509	15	23	2.77	64	23	63	13.28	307
79	74	62.09	1435	14	19	1.95	45	22	54	10.95	253
78	75	58.85	1360	13	16	1.25	29	21	45	9.00	208
77	61	56.21	1299	12	9	0.87	20	20	19	7.31	169
76	84	52.57	1215	11	8	0.52	12	19	35	5.80	134
75	66	49.72	1149	10	4	0.35	8	18	24	4.76	110
74	76	46.43	1073	9	3	0.22	5	17	21	3.85	89
73	63	43.70	1010	8	4	0.04	1	16	18	3.07	71
72	59	41.15	951	6	1	0.00	0	15	12	2.55	59
71	49	39.03	902					14	13	1.99	46
70	70	36.00	832					13	7	1.69	39
69	54	33.67	778					12	6	1.43	33
68	55	31.29	723					11	6	1.17	27
67	65	28.47	658					10	6	0.91	21
66	50	26.31	608					9	1	0.87	20
65	59	23.76	549					8	1	0.82	19
64	46	21.77	503					7	2	0.74	17
63	41	19.99	462					6	4	0.56	13
62	47	17.96	415					5	4	0.39	9
61	34	16.49	381					4	2	0.30	7
60	31	15.15	350					3	2	0.22	5
59	23	14.15	327					2	1	0.17	4
58	27	12.98	300					1	3	0.04	1
57	26	11.86	274					0	1	0.00	0
56	35	10.34	239								
55	26	9.22	219								
54	15	8.57	198								
53	15	7.92	183								
52	17	7.18	166								
51	16	6.49	150								
50	21	5.58	129								
49	17	4.85	112								
48	13	4.28	99								
47	10	3.85	89								
46	7	3.20	82								
45	8	2.87	74								
44	8	2.64	64								
43	2	2.03	61								
42	3	1.70	54								
41	7	1.51	47								
40	7	1.30	35								
39	3	0.95	32								
38	9	0.74	23								
37	5	0.69	18								
36	8	0.61	10								
35	3	0.52	7								
33	2	0.30	5								
32	1	0.17	4								
31	2	0.09	2								
29	2	0.00	0								

TEST STATISTICS FOR FIRST TAKES

5 N= 2311
 6 MEAN= 72.67
 7 S.D.= 13.10
 8 MEAN%= 72.6651
 9 MAXIMUM BINARY SCORE= 34
10 AVERAGE ELAPSED TIME (MIN)=89.42
 11 AVERAGE NON-OBJECTIVE SCORE CODED IN= 29.65

FREQUENCY DISTRIBUTIONS

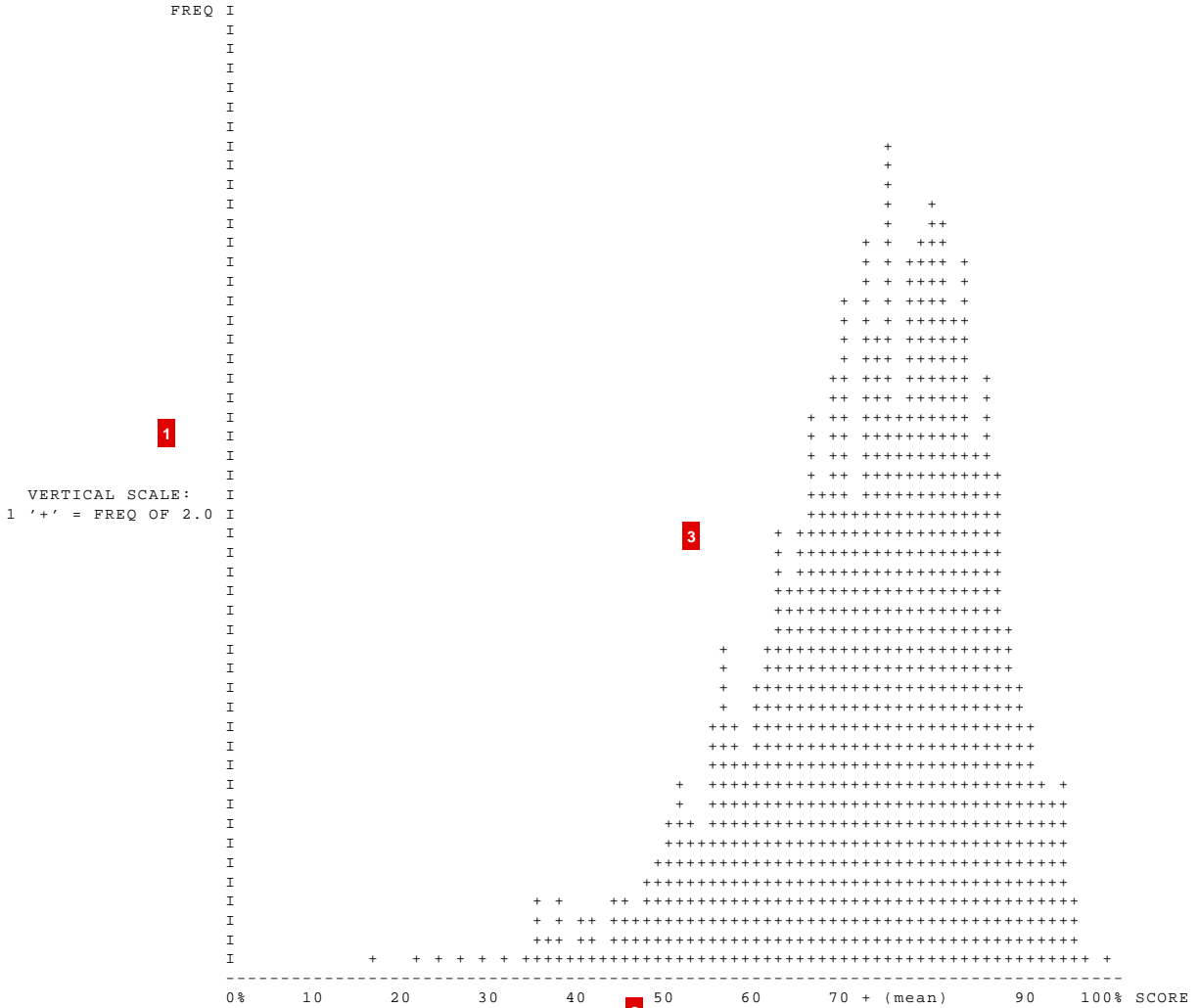
The *Frequency Distributions* report provides the test scores for all parts of the exam. In this example, there are three main categories: Total Score, Number of Objective Questions Correct, and the Non-Objective score.

- Item 1:** The first column of each category ranks the scores (either in raw or percentage format) from high to low. Only the scores achieved by students are printed. In this example, no students achieved 100%, 34% or below 29% for a total score and no student scored below a 6 on the objective part. Therefore, these scores don't appear in their respective columns.
- Item 2:** The *Frequency* (FREQ) columns shows the number of students who achieved the same score. In this case, only one student scored 98% while 62 students achieved 87% as their total score.
- Item 3:** The *Cumulative Percentage* (CUM PCT) columns refer to the percentage of students who scored lower than the indicated score. (**NOTE:** Regularly, cumulative data IS shown as the number, or the percentage at or below the given sample; in these reports, the cumulative values show only those *below* and *not* those at or below.) In this example, the students who scored 90% total on this test did better than 93.86% of the students and those who achieved 36 points in the non-objective test scored higher than 81.09% of the students.
- Item 4:** The *Cumulative Frequency* (CUM FREQ) columns refer to the number of students who scored lower than the score indicated. Taking the students who scored a total of 90%, they did better than 2169 of the students who took the test.
- Item 5:** *N* is the total number of students who took the test (2,311 in this example).
- Item 6:** The *Mean* is the arithmetic average of weighted scores (72.67 in this example).
- Item 7:** The *Standard Deviation* size (13.10 in this example).
- Item 8:** *Mean%* is the arithmetic average of percentage scores (72.6651% in this example).
- Item 9:** The *Maximum Binary Score* is the highest unweighted score possible in the objective part of the examination (34 in this example).
- Item 10:** *Average Elapsed Time* shows the average time it took the students to complete the examination in minutes (89.42 min. in this example).
- Item 11:** The *Average Non-Objective Score Coded In* shows the average score of the non-objective part of the test (29.65 in this example).

DIAGRAM OF TOTAL TEST SCORES

S.C.O.U.T. STATISTICAL ANALYSIS FOR DIET 123 TEST-1 FORM=A 10:04:27 07 DEC 1996

DIAGRAM OF TOTAL TEST SCORES



N=2311 MEAN=72.67 S.D.=13.10 MEAN%=72.6651 MAXIMUM BINARY SCORE=34 AVERAGE ELAPSED TIME (MIN)=89.42

DIAGRAM OF TOTAL TEST SCORES

This diagram is a graphic distribution of total test scores. To better understand this distribution, a definition of some essential terms and their relation to this report will be necessary.

Mean - Arithmetic average of the weighted test scores, or the sum of all scores divided by the number of students who took the test.

Median - The middle score, where half the scores are above it and the other half are below it, or where the cumulative percentile is 50 (approx. 76 in this example).

Mode - The most frequently obtained score, or the largest number in the frequency column (76 in this example since 84 people received it).

Normal Distribution - Distribution where the mean, median, and mode are all equal. A normal distribution graph is a perfectly symmetrical bell curve.

Standard Deviation - Unit of measure indicating the dispersion of a frequency distribution by describing how it is centered around its mean. On a normal distribution curve, 34.13% of the population under the curve is found between the mean and one standard deviation unit away from the mean. Therefore, approx. $\frac{2}{3}$ of the scores would lie under the curve between one standard deviation away from each side of the mean, about 95% of the scores would lie between two standard deviations away from each side of the mean and close to 99% of them would lie between three standard deviations away from each side of the mean.

The explanation of the diagram is as follows:

- Item 1:** The *Y-Axis* represents the frequency. The scale for the axis (which varies greatly for each graph) is marked on its left side (each “+” represents 2 people in this example).
- Item 2:** The *X-Axis* represents the percentage score which ranges from 0% to 100%.
- Item 3:** The graph itself which represents the distribution of all test applicants.
- Item 4:** *N* is the total number of students who took the test (2,311 in this example).
- Item 5:** The *Mean* is the arithmetic average of weighted scores (72.67 in this example).
- Item 6:** The *Standard Deviation* size (13.10 in this example).
- Item 7:** *Mean%* is the arithmetic average of percentage scores (72.6651% in this example).
- Item 8:** The *Maximum Binary Score* is the highest unweighted score achieved in the objective part of the examination (34 in this example).
- Item 9:** *Average Elapsed Time* shows the average time it took the students to complete the examination in minutes (89.42 min. in this example).

ITEM ANALYSIS

S.C.O.U.T. STATISTICAL ANALYSIS FOR DIET 123 TEST=1 FORM=A 10:04:27 07 DEC 1996

		ITEM ANALYSIS																						
ITEM	FREQ	PRCT	FREQ	PRCT	FREQ	PRCT	FREQ	PRCT	FREQ	PRCT	FREQ	PRCT	FREQ	PRCT	FREQ	PRCT	FREQ	PRCT	FREQ	PRCT	BLNK	WRNG		
VAL	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	DISC	NUMB	
1	169	7.3	42	1.8	31	1.3	1994	86.3	75	3.2												13.7		
c		-34.5		-12.6		-22.2	**	46.4		-15.4													317	
2	73	3.2	1923	83.2	53	2.3	100	4.3	160	6.9												2	16.8	
c		-22.3	**	48.8		-14.3		-34.8		-19.4													388	
3	166	7.2	115	5.0	17	0.7	1844	79.8	169	7.3													20.2	
c		-10.0		1635		-5.2	**	31.8		-23.6													467	
4	2177	94.2	35	1.5	8	0.3	22	1.0	69	3.0													5.8	
c	**	37.3		-19.6		-6.8		-12.5		-27.7													134	
5	97	4.2	1817	78.6	15	0.6	31	1.3	349	15.1	1	0.0										1	21.4	
B		-26.0		56.6		-12.3		-12.8		-43.1		-3.0											494	
6	377	14.6	1680	72.7	79	3.4	101	4.4	110	4.8												4	27.3	
B		-35.6	**	53.2		-13.2		-16.4		-24.1													631	
7	10	0.4	2	0.1	2291	99.1	7	0.3	1	0.0													0.9	
D		-35.6		-3.5	**	14.4		-9.0		-2.5													20	
8	86	3.7	8	0.3	1036	44.8	98	4.2	1081	46.8												2	55.2	
c		-16.4		-9.7	**	36.0		-13.0		-23.0													1275	
9	9	0.4	48	2.1	212	9.2	2024	87.6	16	0.7												2	12.4	
c		-9.4		-16.0		-29.4	**	37.7		-13.7													287	
10	736	31.8	1245	53.9	12	0.5	306	13.2	12	0.5													16.1	
c		-35.3	**	48.3		-16.2		-16.8		-10.3													1066	
11	119	5.1	456	19.7	434	18.8	942	40.8	356	15.4												4	59.2	
c		-12.5		-10.1		-20.9	**	38.2		-10.4													1369	
12	771	33.4	69	3.0	289	12.5	1180	51.1	2	0.1													48.9	
c		-17.4		-15.0		-16.3	**	32.1		3.0													11.1	
13	46	2.0	1432	62.0	106	4.6	49	2.1	678	29.3													38.0	
c		-9.9	**	45.7		-8.8		-18.0		-35.9													879	
14	1548	67.0	181	7.8	49	2.1	331	14.3	200	8.7												2	33.0	
c	**	39.5		-13.8		-8.9		-31.4		-8.8													763	
15	3	0.1	2207	95.5	34	1.5	21	0.9	46	2.0													4.5	
c		-3.4	**	34.2		-27.8		-11.4		-18.1													104	
16	70	3.0	111	4.8	62	2.7	1608	69.6	459	19.9												1	30.4	
B		-18.0		-19.7		-16.6	**	62.0		-35.5													703	
17	24	1.0	1726	74.4	7	0.3	225	9.7	329	14.2													25.3	
D		-8.5	**	24.1		-11.1		-14.5		-13.6													585	
18	111	4.8	1947	84.2	215	9.3	36	1.6	1	0.0												1	15.8	
D		-20.8	**	50.3		-41.8		-11.4		-10.5													364	
19	17	0.7	9	0.4	276	11.9	47	2.0	1962	84.9													15.1	
c		-13.2		-9.4		-28.1		-19.1	**	37.7													349	
20	244	10.6	297	12.9	38	1.6	760	32.9	972	42.1													67.1	
B		-10.3		-13.4		-10.2	**	50.6		-30.1													1551	
21	28	1.2	4	0.2	2212	95.7	25	1.1	41	1.8												1	4.3	
D		-10.7		-1.5	**	24.6		-9.6		-20.7													99	
22	126	5.5	35	1.5	1980	85.7	15	0.6	155	6.7													14.3	
c		-31.6		-13.7	**	38.4		-16.9		-13.0													331	
23	1287	55.7	287	12.4	16	0.7	13	0.6	708	30.6													44.3	
c	**	47.2		-32.6		-12.8		-4.8		-24.5													1024	
24	1372	59.4	188	8.1	231	10.0	517	22.4															3	40.6
c	**	37.7		-21.0		-22.6		-14.2															939	
25	26	1.1	22	1.0	9	0.4	37	1.6	2125	92.0	12	0.5	3	0.1	11	0.5	10	0.4	35	1.5	21	8.0		
B		-17.2		-17.2		-11.1		-30.8	**	57.9		-14.4		-2.8		-16.1		-19.5		-22.7			186	
26	557	24.1	4	0.2	1629	70.5	3	0.1			20	0.9	1	0.0	72	3.1			12	0.5	13	29.5		
B		-50.3		-8.7	**	60.9		-7.7		-9.1		-4.6		-14.6						-15.8			682	
27	76	3.3	4	0.2	99	4.3	1	0.0	4	0.2	18	0.8	9	0.4	2050	88.7			37	1.6	13	11.3		
B		28.3		-9.0		-24.1		-4.6		-10.1		-16.0		-7.6	**	52.6				-22.3			261	
28	10	0.4	67	2.9	51	2.2	19	0.8	3	0.1	2090	90.4	14	0.6	27	1.2	2	0.4	11	0.5	17	9.6		
c		-9.1		-224.		-19.4		-12.6		-3.7	**	48.2		-17.1		-17.1		-7.7		-17.2			221	
29	28	1.2	15	0.6	43	1.9	28	1.2	43	1.9	21	0.9	54	2.3	30	1.3	10	0.4	2023	87.5	16	12.5		
B		-10.2		-17.6		-17.0		-25.1		-26.0		-21.4		-17.8		-22.9		-16.2	**	62.0			288	

ITEM ANALYSIS

This analysis gives specific information about each objective question as follows:

Item 1: This column shows the question number at the top and, at the bottom, the evaluation given according to the *Key for Question Evaluation* sheet (in this example, question 1 received a *C* and question 20 a *B*).

Item 2: The middle columns show details on how the question was answered. These columns are labeled A-J and represent the available choices for each question. Each cell is divided into four parts: Frequency (upper left-hand corner), Key Answer (lower left-hand corner), Percentage (upper right-hand corner), and Discrimination Index (lower right-hand corner.) These parts describe the following:

Frequency - Number of students who selected this choice as their answer (2,024 students chose *D* as an answer to question 9 in this example).

Correct Answer - Denotes the “correct answer” according to the key when two asterisks (**) are present (*A* was the correct answer to question 4 in this example).

Percentage - The percent of students who selected this choice as an answer (32.9% of the students chose *D* as the answer to question 20 in this example).

Discrimination Index - The coefficient of selective efficiency (ranging from -100 to 100). A value of -100 indicates all individuals choosing this response had lower total test scores than those answering differently. A value of 100 indicates all individuals choosing this response had higher scores than those answering differently. A value of 0.0 indicates those choosing this response had the same scores as those who answered differently. This means all students chose the same answer, thus no discrimination shown. It is desirable to maximize the value of this statistic in a positive direction for correct responses and in a negative direction for incorrect responses. In this example, response *E* to question 20 has a discrimination index of -30.1 while response *D* in the same question has 50.6.

[For further explanation, see [Educational Measurement](#), 2nd edition, edited by Robert Thorndike, published by the American Council of Education, 1971, p. 142.]

Item 3: This column shows the number of students who failed to answer or left the answer sheet blank for a particular question in the upper left-hand corner (13 for question 27), the percentage of students who answered this question incorrectly in the upper right-hand corner (48.9% for question 12), and the actual number of students who answered this question incorrectly in the bottom right-hand corner (1,131 for question 12).

TEST RELIABILITY ANALYSIS

S.C.O.U.T. STATISTICAL ANALYSIS FOR DIET 123 TEST=1 FORM=A 10:04:27 07 DEC 1996

TEST RELIABILITY ANALYSIS

1	2	3	4	5	6
SOURCE	DF	SS	MS	F-RATIO	VAR
MEAN	1	44607.57			
INDIV	2310	1632.9888	0.7069	5.0132	0.0166
7 ITEMS	33	2213.1798	67.0661	475.609	0.029
ITMxEXM	76230	10749.2613	0.141		
TOTAL	78574	59203			

8

HOYT RELIABILITY (KR-20)= 0.8005

KR-21= 0.7596

9

MEAN= 72.67

10

S.D.= 13.10

11

SKEWNESS= -0.7849

12

MAX PTBIS=0.7892

TEST RELIABILITY ANALYSIS

This analysis has several items which describe the overall reliability of the examination. Items 1 through 7 are only steps to calculating the remaining items.

Item 1: The *Source* of variation.

Item 2: The *Degrees of Freedom* (DF).

Item 3: The *Sum of Squares* (SS).

Item 4: The *Mean Squares* (MS).

Item 5: The *F-Ratio* (F-RATIO).

Item 6: The *Variance* (VAR).

Item 7: These rows contain statistics associated with the interaction between the individuals who took the exam and each item on the exam.

Item 8: The Hoyt Reliability (Kuder-Richardson, formula 20) and the KR-21 measure inter-item consistency. They are computed by using statistical models which determine what proportion of the total variance on the test scores is contributed by the covariance between the items. The statistics would be less than accurate with the introduction of anything which causes the examinee's responses to a given item to be inconsistent (covary) with his/her responses to other items. The statistics assume that the items are homogeneous measures of a single trait and are inappropriate for use with instruments measuring more than one trait. The KR-21 formula assumes every measure to be a dichotomous item, while the Hoyt coefficient (KR-20) does not make that assumption (e.g. some items can be of a Lickert type). The KR-20 will always be greater than or equal to the KR-21. The single case which equates the two coefficients occurs when all item difficulties are equal. Otherwise, KR-20 will be less than the KR-21 and consequently "underestimate" reliability. When in doubt, the KR-21 is the "conservative" measure.

The values of the KR-20 and KR-21 formulas range from 0 to 1. National test scores (such as GMAT, GRE, ACT, etc.) usually receive a value of 0.95 for the KR-20 and a slightly lower value for the KR-21. The KR-20 value for better than average tests should lie between 0.80 to 0.85 with a slight decrease for the KR-21 value.

Item 9: The *Mean* is the arithmetic average of weighted scores (72.67 in this example).

Item 10: The *Standard Deviation* size (13.10 in this example).

- Item 11:** The *Skewness* of the distribution. A negative value indicates the distribution is skewed to the left. A positive value indicates a skew to the right.
- Item 12:** The *Maximum Point-Biserial* (MAX PTBIS) correlation computed for this test. Points-Biserial are derived by the Maximum Point-Biserial to obtain the discrimination index used in the item analysis. For an explanation of the point-biserial statistic, please refer to a statistics handbook.