

## Obtaining Item Analysis Statistics for Cognitive Items Using SPSS Dropdown Menus

In this document I explain how to use SPSS to obtain item analysis statistics for cognitive items using dropdown menus. See the document “Obtaining Item Analysis Statistics for Cognitive Items Using SPSS Syntax” for instructions on using syntax.

The data for this example are from the study described in the article “Self-efficacy beliefs in college statistics courses<sup>1</sup>.” The cognitive items used for this example are the 14 statistics tasks reported in the article and shown at the end of this document. The item data are from the posttest that was administered at the end of a statistics class.

Although the items are multiple choice, the data for each item were scored as 0 (incorrect) or 1 (correct). It is these dichotomous values that are contained in the dataset. There are no missing data. The SPSS dataset “Stats Task.sav” contains the item-level data (t1-t14) as well as the total score (tottask).

For cognitive items that are already scored as 0 (incorrect) and 1 (correct), the item difficulty is equal to the item mean. In SPSS, the mean can be obtained from the **DESCRIPTIVES** command.

For classroom tests, discrimination is often obtained by dividing the data into upper and lower groups based on total scores. Discrimination indices are then obtained as the difference between the proportion of correct answers for the upper group minus that for the lower group (see p. 124 in the book).

However, when using computer packages such as SPSS, the best way to obtain discrimination indices is to obtain the correlation of each item score with the total score. As explained in the book (see p. 137), the *corrected* item-total correlation should be used, in which each item is correlated with the sum of all the other items *except itself*. This is because an item will always correlate perfectly with itself, so including it in the sum score will artificially inflate the item-total correlation. If the **RELIABILITY** command is used to obtain item-total correlations, the corrected values will be obtained by default.

### Obtaining Difficulty and Discrimination Indices with the Reliability Command

The most efficient way to obtain difficulty and discrimination indices in SPSS is to use the reliability command, which will also output the item means and other descriptive statistics if these are requested.

To obtain these values, choose **Scale** and then **Reliability Analysis** from the **Analyze** menu.

---

<sup>1</sup> Finney, S.J., & Schraw (2003). Contemporary Educational Psychology, 28, 161–186

The screenshot shows the 'Reliability Analysis' dialog box. On the left, a list of items (t1 through t13) is shown, with t1 selected. A blue arrow button points from this list to the 'Items:' box on the right. The 'Items:' box contains a list of the same items (t1 through t6). Below the 'Items:' box is a 'Ratings:' box, which is currently empty. A second blue arrow button points from the 'Items:' box to the 'Ratings:' box. At the top right of the dialog is a 'Statistics...' button. At the bottom left, there is a 'Model:' dropdown menu set to 'Alpha' and a 'Scale label:' text box. At the bottom center are five buttons: 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

Click on the **Statistics** button and choose **Item** and **Scale If Item Deleted** from the menu under **Descriptives For**. Choosing **Item** provides the item-level means and standard deviations. Choosing **Scale If Item Deleted** provides the corrected item-total correlations.

Click on **Continue** and then on **OK**.

Reliability Analysis: Statistics

Descriptives for

☒ Item

☐ Scale

☒ Scale if item deleted

Inter-Item

☐ Correlations

☐ Covariances

Summaries

☐ Means

☐ Variances

☐ Covariances

☐ Correlations

ANOVA Table

☒ None

☐ F test

☐ Friedman chi-square

☐ Cochran chi-square

Interrater Agreement: Fleiss' Kappa

☐ Display agreement on individual categories

☐ Ignore string cases

☒ String category labels are displayed in uppercase

Asymptotic significance level (%): 95

Missing

☒ Exclude both user-missing and system missing values

☐ User-missing values are treated as valid

☐ Hotelling's T-square

☐ Tukey's test of additivity

☐ Intraclass correlation coefficient

Model: Two-Way Mixed

Type: Consistency

Confidence interval: 95 %

Test value: 0

Continue Cancel Help

This will result in the output below.

The item-level descriptive statistics are printed first, as shown below.

### Item Statistics

	Mean	Std. Deviation	N
t1	.24	.431	103
t2	.40	.492	103
t3	.51	.502	103
t4	.52	.502	103
t5	.54	.501	103
t6	.26	.442	103
t7	.88	.322	103
t8	.30	.461	103
t9	.38	.487	103
t10	.62	.487	103
t11	.53	.501	103
t12	.62	.487	103
t13	.35	.479	103
t14	.36	.482	103



**Difficulty values**

The values .24, .40, etc. are the difficulty values for the 14 items. Recall that the difficulty index is really an “easiness” index because it indicates the proportion of respondents who obtained a correct answer. This set of items is rather difficult as the highest difficulty value is .88 (for item t7) and half of the items had difficulty values less than .50.

The difficulty of the items is likely due, at least in part, to the fact that the distractors for these items were based on common misconceptions.

The corrected item-total correlations, or discrimination values, are printed next; these are shown in the box in the table below.

### Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
t1	6.29	5.424	.007	.497
t2	6.14	4.883	.229	.445
t3	6.02	4.686	.315	.420
t4	6.01	5.265	.047	.493
t5	5.99	4.774	.274	.432
t6	6.27	5.357	.035	.492
t7	5.65	5.210	.211	.456
t8	6.23	5.259	.072	.484
t9	6.16	5.250	.061	.489
t10	5.91	5.002	.176	.459
t11	6.00	4.627	.345	.412
t12	5.91	4.728	.311	.423
t13	6.18	5.093	.139	.469
t14	6.17	5.048	.159	.464

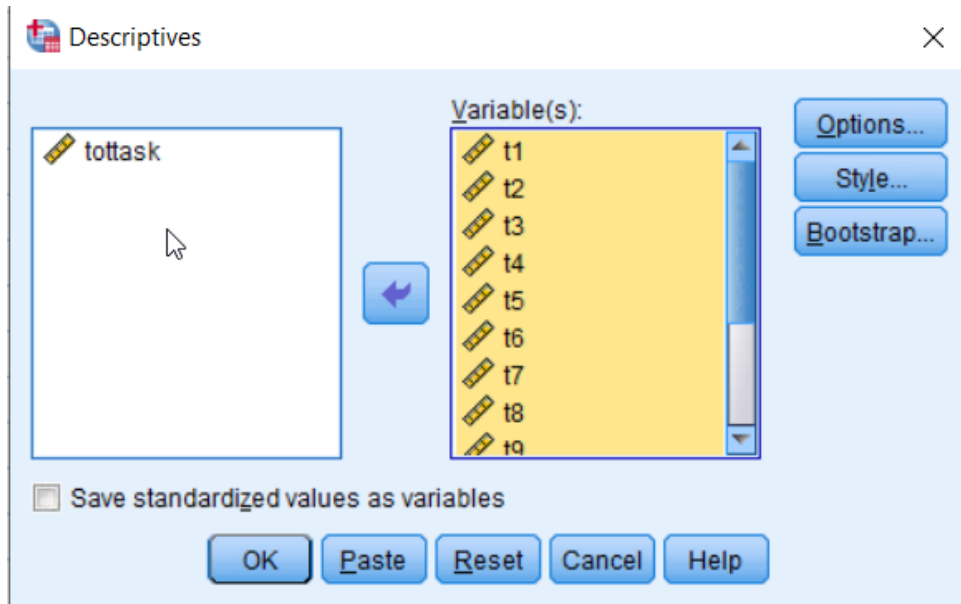


**Discrimination values**

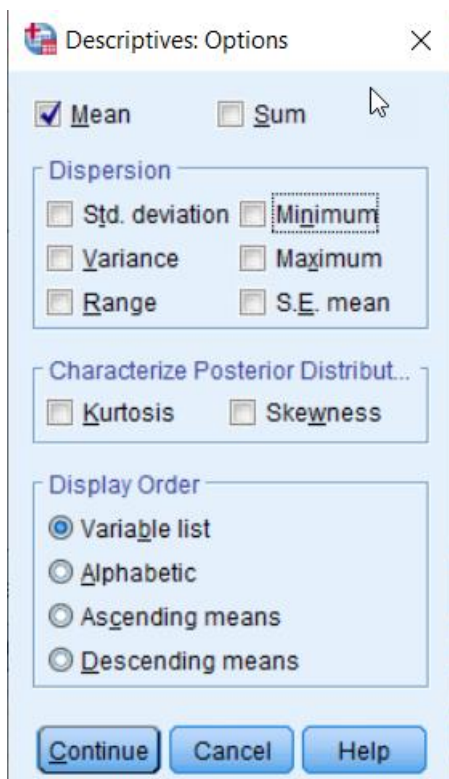
Many of the items (t1, t4, t6, t8, t9) have extremely low discrimination indices, indicating items that do not differentiate well among those with different levels of task knowledge. This is likely due to two things: the relatively high difficulty of the tasks, and the diversity of the tasks, which cover a wide variety of material. This diversity of content, along with the low discrimination values for some items, suggests that the scale may not be unidimensional.

## Obtaining Item Difficulty Statistics with the *DESCRIPTIVES* Command

For cognitive items that are already scored as 0 (incorrect) and 1 (correct), the item difficulty is equal to the item mean. In SPSS, the mean can be obtained from the **DESCRIPTIVES** command.



Under **Options** choose the mean.



This will result in the following output:

**Descriptive Statistics**

	N	Mean
t1	103	.24
t2	103	.40
t3	103	.51
t4	103	.52
t5	103	.54
t6	103	.26
t7	103	.88
t8	103	.30
t9	103	.38
t10	103	.62
t11	103	.53
t12	103	.62
t13	103	.35
t14	103	.36
Valid N (listwise)	103	

Note that these values are identical to those obtained using the **RELIABILITY** command.

## Statistics Tasks

Please complete the following problems to the best of your ability. Take your time and think through each problem and select the best answer. If you have no idea what the correct answer is you should select the option "I don't know" instead of simply skipping the problem.

1. A researcher is interested in the number of miles driven by tourists in the U.S.A. over Labor Day weekend. What scale of measurement is used to measure miles?
  - a. nominal
  - b. ordinal
  - c. interval
  - d. ratio
  - e. I don't know.
  
2. A researcher conducted a study to see the effects of tutoring on exam performance. A sample of students was administered an exam both before and after a tutoring session. The researcher used the appropriate statistical test and found that the probability value (p-value) equaled .02. Select the correct interpretation of this value.
  - a. The probability that the null hypothesis (no difference between pre-tutoring and post-tutoring exam scores) is true equals .02.
  - b. The probability that the alternative hypothesis (there is a difference between pre-tutoring and post-tutoring exam scores) is true equals .02.
  - c. The probability of observing a difference between pre- and post-tutoring exam scores at least this large equals .02, given that the null hypothesis is true.
  - d. The probability of finding this same difference between pre-and post-tutoring exam scores in future studies equals .98.
  - e. I don't know.
  
3. Which of the following distributions is positively skewed?
  - a. mean=15, median=20, mode=22
  - b. mean=45, median=40, mode=39
  - c. mean=30, median=31, mode=31
  - d. mean=12, median=12, mode=12
  - e. I don't know.

4. A researcher is interested in knowing if people who voted in the last presidential election have significantly higher IQ scores than people who did not vote in the last presidential election. The researcher takes a random sample of 1500 voters and 1500 non-voters and has each of them complete an IQ test. Choose the correct statistical procedure/test to use to answer the researcher's question.
- One-sample z-test
  - One-sample t-test
  - Independent samples t-test
  - Repeated measures t-test
  - Pearson's correlation coefficient
  - Regression
  - I don't know.
5. Referring to the research scenario in question 4, what would be the correct conclusion if the researcher rejected the null hypothesis?
- There is not a significant difference between the IQ scores of voters and non-voters.
  - Non-voters have significantly higher IQ scores than voters.
  - Voters have significantly higher IQ scores than non-voters.
  - None of the above conclusions are correct.
  - I don't know.
6. From the following list select *all* the factors that influence the power of a statistical test.  
This means you may select more than one response.
- effect size (treatment effect)
  - alpha level
  - sample size
  - if the test is directional or nondirectional (one vs. two tailed test)
  - I don't know
7. The standard deviation of Sociology test scores from group of freshmen enrolled in introductory Sociology equals 6. This tells us that
- they all received very low scores on the test
  - the distribution of the test scores is positively skewed
  - the typical distance of the test scores from the mean test score is 6 points
  - the distance between each pair of test scores is 6 points
  - if another sample were taken, its average would be 6 points less than this sample
  - I don't know.

8. If Dr. Baker conducts a study and rejects the null hypothesis when, in fact, the null hypothesis is true, in which category would her statistical decision fall?
- Correct decision ( $1-\alpha$ )
  - Type I error ( $\alpha$ )
  - Correct decision ( $1-\beta$ )
  - Type II error ( $\beta$ )
  - I don't know.
9. If the standard error for a sampling distribution equals 3, which statement best describes this value?
- The typical distance between the sample means and the population mean is 3 points.
  - The typical distance between the sample values and the sample mean is 3 points.
  - The typical distance between the sample standard deviation and the population standard deviation is 3 points.
  - The typical distance between the population mean and the population standard deviation is 3 points.
  - I don't know.
10. Statistical procedures that summarize the data in a sample are called\_\_\_\_\_ while statistical procedures that are used to make educated guesses about the population are called \_\_\_\_\_.
- inferential, measures of variability
  - sampling distributions, descriptive
  - sampling distributions, measures of variability
  - descriptive, inferential
  - inferential, descriptive
  - I don't know.
11. It is reported that the median salary for employees at a local library equals \$25,000. Which statement below is the best interpretation of this value?
- The arithmetic average salary received by the employees equals \$25,000.
  - No employees received a salary of \$25,000.
  - Half of the employees received salaries higher than \$25,000.
  - \$25,000 was the salary most frequently received by the employees.
  - I don't know.

12. A researcher knows that the average age for all grant recipients in the United States is 47.5. She is interested in examining if there is a significant difference between the average age of male and female grant recipients. She, therefore, selects a random sample of male and female grant recipients from the United States and finds the average age for males is 50.2 while the average age for females is 43.7. Which of the following values from above is a population parameter?
- a. 47.5
  - b. 50.2
  - c. 43.7
  - d. all of the above
  - e. I don't know.
13. The measure of central tendency that can be computed for any type of data is
- a. the median
  - b. the mode
  - c. the mean
  - d. all of the above
  - e. I don't know.
14. A sampling distribution is different from a population distribution in that
- a. the mean of the sampling distribution is larger than the mean of the population distribution
  - b. a sampling distribution is a distribution of statistics while a population distribution is a distribution of scores
  - c. a sampling distribution is skewed while a population distribution is symmetric
  - d. all of the above
  - e. I don't know.