

Obtaining Item Analysis Statistics for Cognitive Items Using SAS

In this document I explain how to use SAS to obtain item analysis statistics for cognitive items.

The data for this example are from the study described in the article “Self-efficacy beliefs in college statistics courses¹.” The cognitive items used 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 SAS dataset “task.sas7bdat” 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.

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 SAS, the easiest 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 **proc corr** is used to obtain item-total correlations, the corrected values will be obtained by default.

Obtaining Difficulty and Discrimination Indices with *Proc Corr*

The most efficient way to obtain difficulty and discrimination indices in SAS is to use **proc corr**, which will output the item means and other descriptive statistics by default. To obtain the item-total correlations, include the options **alpha** and **nomiss**.

Inclusion of the **alpha** option provides the corrected item-total correlation and alpha-if-item-deleted indices. Recall that the alpha-if-item-deleted index is the value of coefficient alpha that would be obtained if a given item were deleted (see p. 137 in the book).

The **nomiss** option causes missing data to be deleted listwise rather than pairwise (the latter is the default method). In listwise deletion, any case with missing values for any of the variables in the analysis will be treated as missing (and excluded from analyses), whereas in pairwise deletion cases are included in the analysis for any pair of variables for

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

which they have complete data. Listwise deletion is required for computation of item-total correlations and alpha-if-item-deleted values.

The following syntax will output the difficulty and discrimination values for the 14 statistics tasks. By default, the correlation matrix will also be printed. To suppress this, use the option **nocorr**.

```
proc corr data = task alpha;
var t1-t14;
run;
```

The output from this analysis is printed below.

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
t1	103	0.24272	0.43082	25.00000	0	1.00000
t2	103	0.39806	0.49189	41.00000	0	1.00000
t3	103	0.51456	0.50223	53.00000	0	1.00000
t4	103	0.52427	0.50185	54.00000	0	1.00000
t5	103	0.54369	0.50052	56.00000	0	1.00000
t6	103	0.26214	0.44195	27.00000	0	1.00000
t7	103	0.88350	0.32240	91.00000	0	1.00000
t8	103	0.30097	0.46092	31.00000	0	1.00000
t9	103	0.37864	0.48742	39.00000	0	1.00000
t10	103	0.62136	0.48742	64.00000	0	1.00000
t11	103	0.53398	0.50128	55.00000	0	1.00000
t12	103	0.62136	0.48742	64.00000	0	1.00000
t13	103	0.34951	0.47915	36.00000	0	1.00000
t14	103	0.35922	0.48212	37.00000	0	1.00000



The item means are the difficulty values.

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 item discrimination values (corrected item-total correlations) are shown in the table below. Note that there are two sets of columns: one for “raw variables” and one for “standardized variables.” The “standardized variables” columns are for situations in which the variable scores are in standardized form, such as z-scores. If this is not the case, as in this example, the values in the “raw variables” columns should be used.

Cronbach Coefficient Alpha with Deleted Variable				
Deleted Variable	Raw Variables		Standardized Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
t1	0.007020	0.497277	0.013761	0.494291
t2	0.229336	0.444871	0.226023	0.441041
t3	0.315359	0.420350	0.327375	0.413977
t4	0.046621	0.493151	0.046334	0.486413
t5	0.273802	0.432253	0.257436	0.432769
t6	0.035082	0.491868	0.030804	0.490182
t7	0.210575	0.456205	0.207988	0.445744
t8	0.072131	0.484392	0.057594	0.483665
t9	0.060937	0.488578	0.063043	0.482331
t10	0.176199	0.459085	0.190619	0.450241
t11	0.345485	0.411754	0.340060	0.410513
t12	0.310753	0.422838	0.308170	0.419189
t13	0.139254	0.468537	0.126505	0.466574
t14	0.158703	0.463609	0.169500	0.455668

The discrimination values are shown in the “correlation with total” column. These are the corrected item-total correlations.

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 Indices Using Proc Mean

If only the difficulty indices are needed, these can be obtained from **proc means**. Although **proc univariate** could also be used, it produces a large amount of output and there is no option to obtain only the mean.

The syntax below will output the mean for each item and limit the number of decimal places to two.

```
proc means data = task maxdec=2 mean;  
var t1-t14;  
run;
```

This syntax will result in the output below. Note that the values obtained are the same as those obtained from **proc corr**, although the number of decimal places has been limited to 2 (**maxdec=2**).

The MEANS Procedure

Variable	Mean
t1	0.24
t2	0.40
t3	0.51
t4	0.52
t5	0.54
t6	0.26
t7	0.88
t8	0.30
t9	0.38
t10	0.62
t11	0.53
t12	0.62
t13	0.35
t14	0.36

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, it's 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 (α)
 - Correct decision ($1-\beta$)
 - Type II error (β)
 - 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.