**NANTEZA ANGELLAH**

**REG NO: 2023/DSCE/00106/SS**

**Assignment**

**What statistical hypothesis test can be applied to two samples of categorical data with the same categories to determine if the frequency distributions are significantly different?**

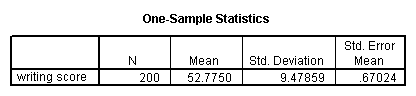
**One sample t-test**

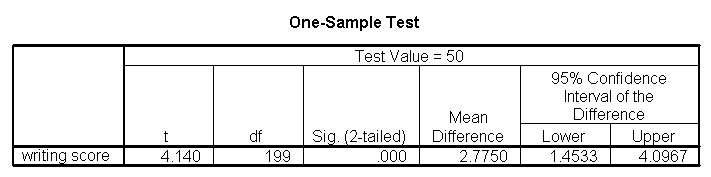
A one sample t-test allows us to test whether a sample mean (of a normally distributed interval variable) significantly differs from a hypothesized value.  For example, using the [hsb2 data file](https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsb2-3.sav), say we wish to test whether the average writing score (**write**) differs significantly from 50.  We can do this as shown below.

**t-test**

**/testval = 50**

**/variable = write.**





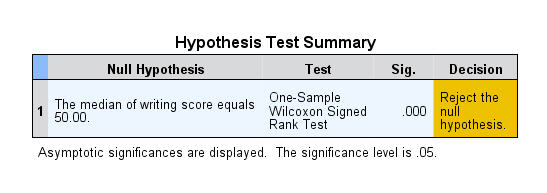
The mean of the variable **write** for this particular sample of students is 52.775, which is statistically significantly different from the test value of 50.  We would conclude that this group of students has a significantly higher mean on the writing test than 50.

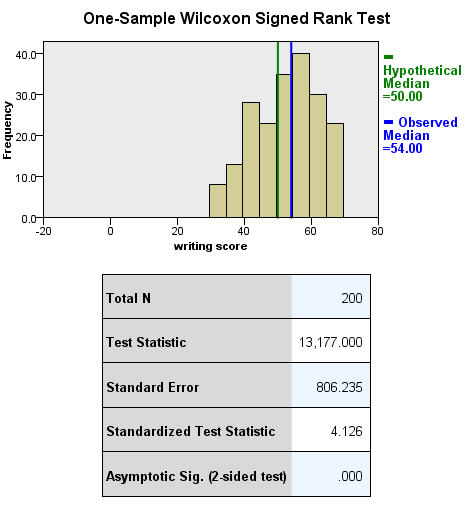
**One sample median test**

A one sample median test allows us to test whether a sample median differs significantly from a hypothesized value.  We will use the same variable, **write**, as we did in the [one sample t-test](https://stats.oarc.ucla.edu/spss/whatstat/#1sampt) example above, but we do not need to assume that it is interval and normally distributed (we only need to assume that **write** is an ordinal variable).

**nptests**

**/onesample test (write) wilcoxon(testvalue = 50).**



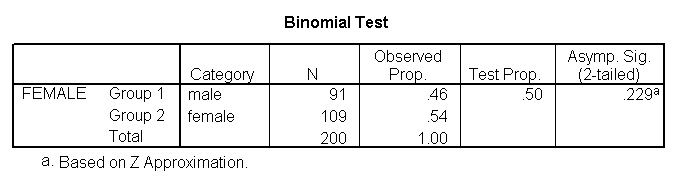


**Binomial test**

A one sample binomial test allows us to test whether the proportion of successes on a two-level categorical dependent variable significantly differs from a hypothesized value.  For example, using the [hsb2 data file](https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsb2-3.sav), say we wish to test whether the proportion of females (**female**) differs significantly from 50%, i.e., from .5.  We can do this as shown below.

**npar tests**

**/binomial (.5) = female.**



The results indicate that there is no statistically significant difference (p = .229).  In other words, the proportion of females in this sample does not significantly differ from the hypothesized value of 50%.

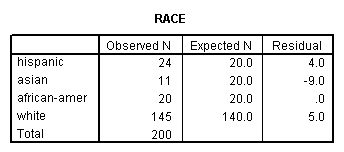
**Chi-square goodness of fit**

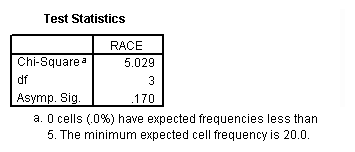
A chi-square goodness of fit test allows us to test whether the observed proportions for a categorical variable differ from hypothesized proportions.  For example, let’s suppose that we believe that the general population consists of 10% Hispanic, 10% Asian, 10% African American and 70% White folks.  We want to test whether the observed proportions from our sample differ significantly from these hypothesized proportions.

**npar test**

**/chisquare = race**

**/expected = 10 10 10 70.**





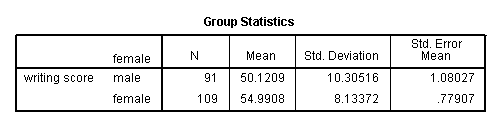
These results show that racial composition in our sample does not differ significantly from the hypothesized values that we supplied (chi-square with three degrees of freedom = 5.029, p = .170).

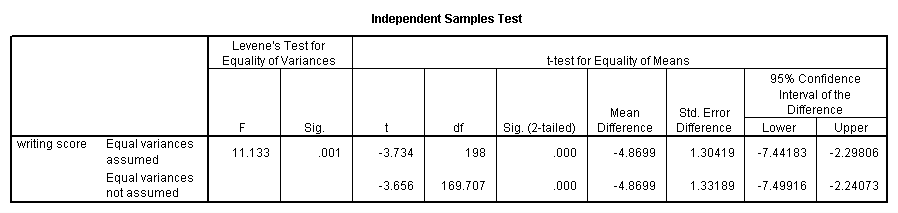
**Two independent samples t-test**

An independent samples t-test is used when you want to compare the means of a normally distributed interval dependent variable for two independent groups.  For example, using the [hsb2 data file](https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsb2-3.sav), say we wish to test whether the mean for **write** is the same for males and females.

**t-test groups = female(0 1)**

**/variables = write.**





Because the standard deviations for the two groups are similar (10.3 and 8.1), we will use the “equal variances assumed” test.  The results indicate that there is a statistically significant difference between the mean writing score for males and females (t = -3.734, p = .000).  In other words, females have a statistically significantly higher mean score on writing (54.99) than males (50.12).

**See also**

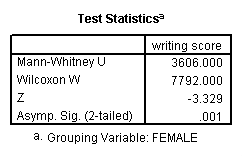
* [SPSS Learning Module: An overview of statistical tests in SPSS](https://stats.oarc.ucla.edu/spss/modules/an-overview-of-statistical-tests-in-spss/)

**Wilcoxon-Mann-Whitney test**

The Wilcoxon-Mann-Whitney test is a non-parametric analog to the independent samples t-test and can be used when you do not assume that the dependent variable is a normally distributed interval variable (you only assume that the variable is at least ordinal).  You will notice that the SPSS syntax for the Wilcoxon-Mann-Whitney test is almost identical to that of the independent samples t-test.  We will use the same data file (the [hsb2 data file](https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsb2-3.sav)) and the same variables in this example as we did in the [independent t-test example](https://stats.oarc.ucla.edu/spss/whatstat/#2ittest) above and will not assume that **write**, our dependent variable, is normally distributed.

**npar test**

**/m-w = write by female(0 1).**



The results suggest that there is a statistically significant difference between the underlying distributions of the **write** scores of males and the **write** scores of females (z = -3.329, p = 0.001).

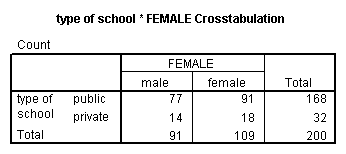
**Chi-square test**

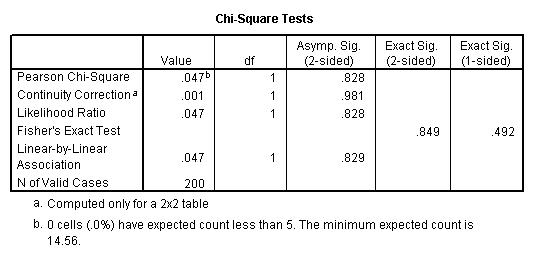
A chi-square test is used when you want to see if there is a relationship between two categorical variables.  In SPSS, the **chisq** option is used on the **statistics** subcommand of the **crosstabs** command to obtain the test statistic and its associated p-value.  Using the [hsb2 data file](https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsb2-3.sav), let’s see if there is a relationship between the type of school attended (**schtyp**) and students’ gender (**female**).  Remember that the chi-square test assumes that the expected value for each cell is five or higher. This assumption is easily met in the examples below.  However, if this assumption is not met in your data, please see the section on Fisher’s exact test below.

**crosstabs**

**/tables = schtyp by female**

**/statistic = chisq.**





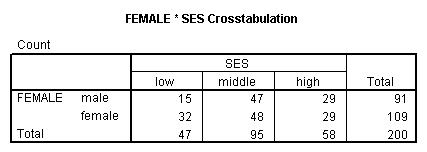
These results indicate that there is no statistically significant relationship between the type of school attended and gender (chi-square with one degree of freedom = 0.047, p = 0.828).

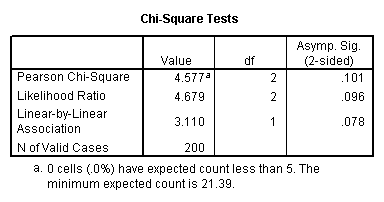
Let’s look at another example, this time looking at the linear relationship between gender (**female**) and socio-economic status (**ses**).  The point of this example is that one (or both) variables may have more than two levels, and that the variables do not have to have the same number of levels.  In this example, **female** has two levels (male and female) and **ses** has three levels (low, medium and high).

**crosstabs**

**/tables = female by ses**

**/statistic = chisq.**





Again we find that there is no statistically significant relationship between the variables (chi-square with two degrees of freedom = 4.577, p = 0.101).

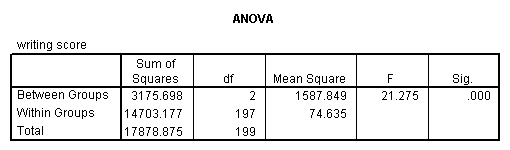
**Fisher’s exact test**

The Fisher’s exact test is used when you want to conduct a chi-square test but one or more of your cells has an expected frequency of five or less.  Remember that the chi-square test assumes that each cell has an expected frequency of five or more, but the Fisher’s exact test has no such assumption and can be used regardless of how small the expected frequency is. In SPSS unless you have the SPSS Exact Test Module, you can only perform a Fisher’s exact test on a 2×2 table, and these results are presented by default.  Please see the results from the chi squared example above.

**One-way ANOVA**

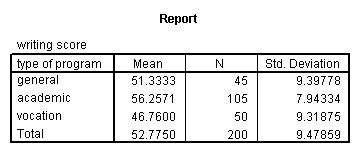
A one-way analysis of variance (ANOVA) is used when you have a categorical independent variable (with two or more categories) and a normally distributed interval dependent variable and you wish to test for differences in the means of the dependent variable broken down by the levels of the independent variable.  For example, using the [hsb2 data file](https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsb2-3.sav), say we wish to test whether the mean of **write** differs between the three program types (**prog**).  The command for this test would be:

**oneway write by prog.**



The mean of the dependent variable differs significantly among the levels of program type.  However, we do not know if the difference is between only two of the levels or all three of the levels.  (The F test for the **Model** is the same as the F test for **prog** because **prog** was the only variable entered into the model.  If other variables had also been entered, the F test for the **Model** would have been different from **prog**.)  To see the mean of **write** for each level of program type,

**means tables = write by prog.**



From this we can see that the students in the academic program have the highest mean writing score, while students in the vocational program have the lowest.