

ANOVA

Does truancy in Year 11 impact total GCSE score?

We've just used a two-sample t test to compare differences in GCSE means across two categories in a binary variable, **s1gender**. Now suppose we'd like to see if there are significant differences in mean GCSE scores based on truancy in Year 11. To operationalize truancy, our independent variable, we'll use **s1tru**, which is a categorical variable with three categories. Because we want to compare the significance of differences in mean GCSE score in an independent variable with more than two categories, we'll need to use an ANOVA. An ANOVA works just like a t test, in that it analyses the significance of differences in means between categories in a variable with regard to sample variation. However, ANOVA tests are used when variables have more than two categories. Because we are curious about the impact one independent variable with several categories has on our dependent variable, we'll be running a one-way ANOVA. If we were looking into the influence of two independent variables with several categories on a dependent variable, we would use a two-way ANOVA.

Before we begin, we should run frequency analysis on **s1tru** to see what condition this variable is in (e.g. look for missing data and so on). Click on **Analyze, Descriptive Statistics**, and then **Frequencies**. Move **s1tru** over to the **Variable(s)** text box and then click **OK**.

Your output should look like this:

Statistics

Truancy in year 11

| | | |
|---|---------|-------|
| N | Valid | 14003 |
| | Missing | 0 |

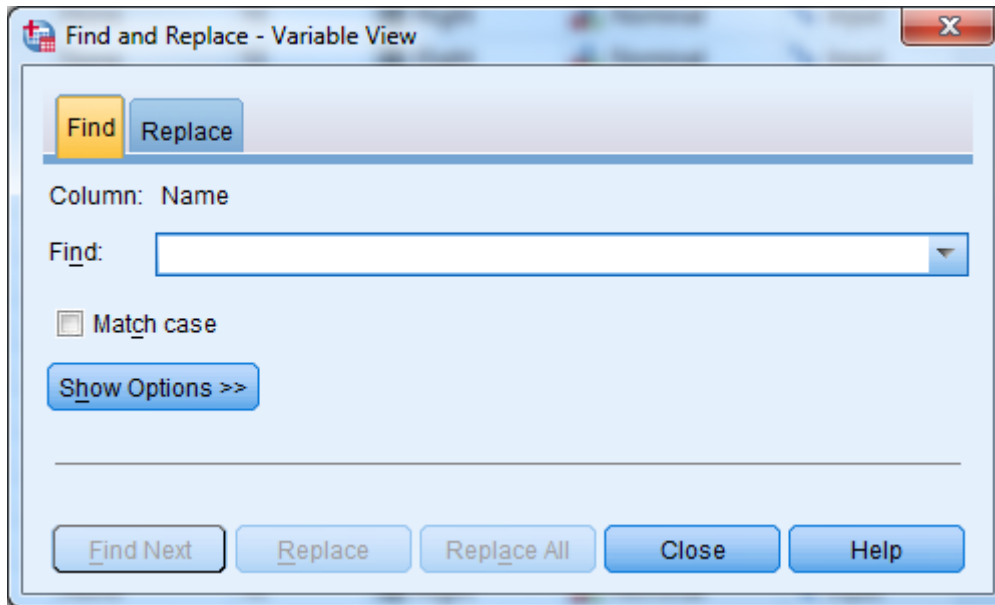
Truancy in year 11

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|--------------------------|-----------|---------|---------------|--------------------|
| Not answered | 171 | 1.2 | 1.2 | 1.2 |
| Persistant truancy | 339 | 2.4 | 2.4 | 3.6 |
| Valid Occasional truancy | 3692 | 26.4 | 26.4 | 30.0 |
| No truancy | 9801 | 70.0 | 70.0 | 100.0 |
| Total | 14003 | 100.0 | 100.0 | |

In the **Truancy in year 11** table above, we can see that there are a total of 171 cases in which a young person didn't answer the question about truancy. These unanswered cases are effectively missing data, as we do not know whether or not these 171 people were ever truant. Therefore, we can code these responses as missing.

Luckily, this is really easy to do.

Find **Variable View** on the **Data Editor** window of SPSS. Click to highlight any cell under **Name**, the first column in **Variable View**. When a cell is highlighted, hit **Ctrl + F**. You should now see a dialogue box that looks like this:



This box will help you locate **s1tru** (or any other variable you'd like to find) quickly and easily. Just type **s1tru** into the text box and click **Find Next**. SPSS will find **s1tru** in the list of variables and bring it up for you.

Once you've found **s1tru**, follow its row over to the **Missing** column and click to open the **Missing Value** dialogue box. Here, select **Discrete Missing Values** and enter in -9.00, the numerical label for all the unanswered cases in the variable. (You can check this by clicking on the **Values** cell in the **s1tru** row.) Click **OK**.

Now, you should see -9.00 in the **Missing** cell for **s1tru** – it's been coded as missing in **s1tru** and won't be used in our analyses.

When you've recoded the missing values in **s1tru**, go to **Analyze, Compare Means**, and then **One-Way ANOVA**.

Enter **s1gcseptsnew** in the **Dependent List** box and **s1tru** in the **Factor** box. Click **OK**.

You should get the following output:

ANOVA

ks4 pts score on new basis not capped

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|-------|--------------|---------|------|
| Between Groups | 24398087.792 | 2 | 12199043.896 | 752.941 | .000 |
| Within Groups | 224055564.576 | 13829 | 16201.863 | | |
| Total | 248453652.368 | 13831 | | | |

Because an ANOVA test works like a t test, we are again looking for a p-value of less than 0.05 in order to say that there is a statistically significant difference in GCSE scores across the three categories of **s1truan**. In the table above, we can see that the significance level for this ANOVA is 0.000, well below the 0.05 threshold. This means that there in fact differences in GCSE scores between young people who were persistently truant, those who were occasionally truant, and those who were never truant.

Summary

You've just run a One-Way ANOVA to determine that there were statistically significant differences in mean GCSE score between the categories of s1truan. Next, suppose we wanted to find out exactly what the differences in mean were between all the categories of s1truan. To answer this question, we'll use a Post-Hoc Test.

*****Note: as we are making changes to a dataset we'll continue using for the rest of this section, please make sure to save your changes before you close down SPSS. This will save you having to repeat sections you've already completed!**