***Features of SPSS:***

* **Generalized Spatial Association Rule (GSAR):**

One of the new GeoSpatial Modeling Wizard options allows you to build a Time Series model using geomapping information. The idea is to map events taking place in space over slices of time. For instance, a lot of urban crime is at night, but suburban breaking-and-entering crimes tend to happen during the workday.

* **Spatio-Temporal Prediction (STP):**

This is another new menu in the GeoSpatial Modeling Wizard. This technique allows you to create linear models when data has been collected over a long period of time at different locations.

* **Temporal Causal Modeling (TCM):**

A whole new Forecasting menu. It uses a wizard like environment to help you add the best predictors to your Time Series models.

* **Completely redesigned web reports:**

Version 23 brings with it the new Web Report with a lot more interactivity. And because it’s web based, you don’t have to worry about the recipient having a copy of SPSS.

* **A wider range of R programming options:**

The combination is really proving powerful, so SPSS now allows you to call SPSS from R.

* **Compare Subgroups Plot:**

Another bit of big news in this release is that there are a ton of new programmability plug-ins in the menus. IBM has written these for you so you don’t have to know any Python. In fact, you don’t really have to know where they came from except that you have to select Install Python when you install Version 23. As an example, there is a nifty plot in the Graphs menu that automatically chooses appropriate graphic based on the Level of Measurement of the variables.

* **Split into Files:**

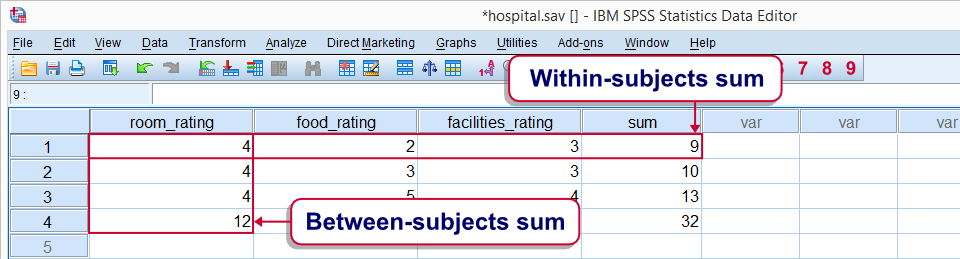
Another one of the Python plugin macros. It makes it super easy to create files for each category in a categorical file for instance, you may want to create a file for new customers and a separate file for established customers.

* **Create Dummy Variables:**

Another great Python plug-in. This one creates true/false variables for each category in a categorical variable. This is a requirement in Regression. Many people have been doing this manually for years, but this plug-in makes it easier.

***Functionalities of SPSS:***

This tutorial walks you through SPSS' main statistical functions. They are mainly used with COMPUTE and IF. Note that these are all within-subjects functions (or “horizontal functions”). For between-subjects (or “vertical”) functions, see AGGREGATE. We recommend you follow along by downloading and opening hospital.sav.



* **SPSS Statistical Funcions - Missing Values:**

SPSS statistical functions only return system missing values if all their input values are missing values. If a single input value is valid, the output value will be valid too. This holds for all functions we'll cover in this tutorial. Remember that the opposite holds for SPSS numeric functions: the latter only return a valid value if all their input values are valid.

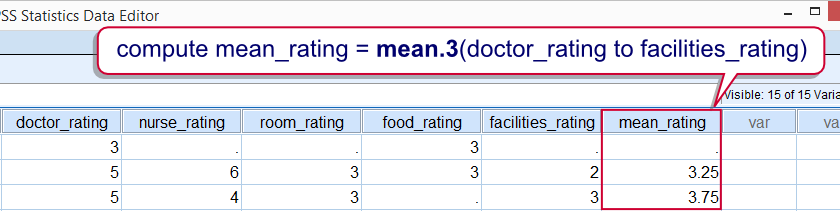
* **SPSS Statistical Funcions - Dot Operator:**

A minimal number of valid input values can be specified for statistical functions. This is done by suffixing the function with a period followed by the required number of valid values. For example

**compute mean\_v = mean.3(v1 to v5).**

means “Compute mean\_v only for cases having at least 3 valid values over v1 to v5. Cases with fewer valid values must get a system missing value on mean\_v.”

The dot operator can be used with all functions covered in this tutorial. Don't overlook it. Although it's little known among SPSS users, it's a terrific time saving feature.



* **Data Preparation:**

We'll use only the last 5 variables in our data.\* The functions we'll demonstrate on them may return incorrect values if we fail to specify user missing values. We'll therefore do a quick check by running FREQUENCIES with the syntax below. Note the TO keyword in step 5.

\*1. Specify folder where data are located. 🡺 cd 'd:/temp'.

\*2. Open data file. 🡺 get file 'hospital.sav'.

\*3. Show values and value labels in output. 🡺 set tnumbers both.

\*4. Inspect frequencies. 🡺 frequencies doctor\_rating to facilities\_rating.

* **SPSS MEAN Function:**

Means over variables are returned by SPSS MEAN function. If missing values are present, the sum of the valid values is divided by the number of valid values. The syntax below shows how to compute within-subjects means.

\*Compute mean\_rating as mean over all 5 ratings.

compute mean\_rating = mean(doctor\_rating to facilities\_rating).

* **SPSS SUM Function:**

SPSS SUM function returns the sum over a number of variables. In the presence of missing values, the sum over all valid values is returned. Keep in mind that the result may be somewhat misleading in this case.\* The syntax below computes the within-subjects sum over our rating variables.

\*Compute sum over 5 ratings.

compute sum\_rating = sum(doctor\_rating to facilities\_rating).

* **SPSS MIN and MAX Function:**

SPSS MIN and MAX functions return the minimum and maximum values

compute min\_rating = min(doctor\_rating to facilities\_rating).

compute max\_rating = max(doctor\_rating to facilities\_rating).

* **SPSS SD Function:**

The standard deviation over a number of variables is returned by SPSS SD function. Keep in mind that we're referring to the within-subjects standard deviation here.\*Computing within-subjects standard deviations comes in handy in survey research for detecting straightliners: respondents who give the same answer to all questions will have a standard deviation of zero over these questions. This may be an indication that the questions weren't answered seriously, in which case you may want to exclude such cases from analysis.

SPSS SD Function Syntax Example:

\*1. Compute within-subjects standard deviation over rating variables.

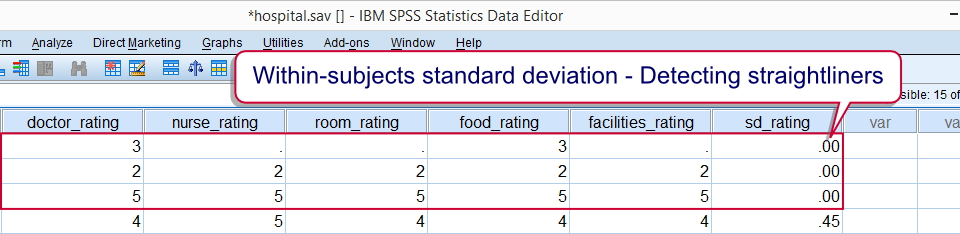
compute sd\_rating = sd(doctor\_rating to facilities\_rating).

\*2. Move straightliners to top of file.

sort cases by sd\_rating.

\*3. Delete straightliners from data.

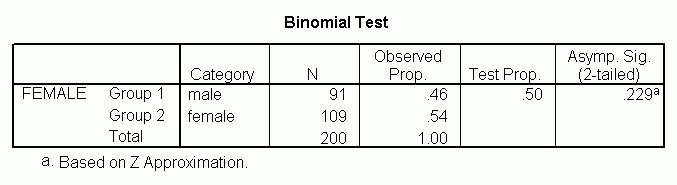
select if sd\_rating > 0.



***Different type of experiments can be performed in SPSS:***

* **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, 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.



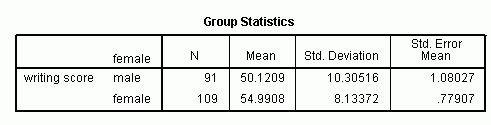
The results indicate that there is no statistically significant difference (p = .229). In other words, the proportion of females in this sample does not significantly differ from the hypothesized value of 50%.

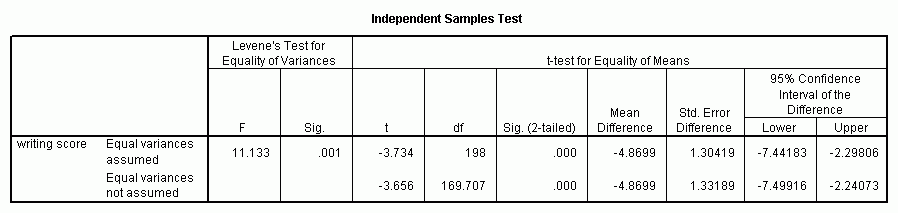
* **Two independent samples t-test**

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, 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.

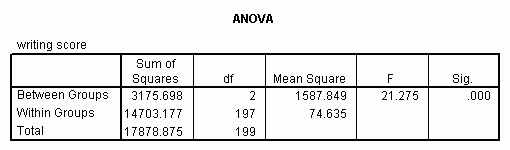




* **One-way ANOVA**

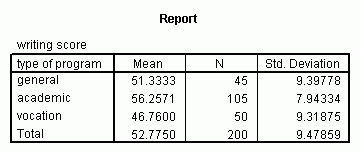
A one-way analysis of variance (ANOVA) is used when you have a categorical independent variable (with two or more categories) and a normally distributed interval dependent variable and you wish to test for differences in the means of the dependent variable broken down by the levels of the independent variable. For example, using the hsb2 data file, 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.

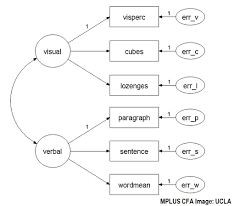


From this we can see that the students in the academic program have the highest mean writing score, while students in the vocational program have the lowest.

***Different type of research can be conducted by using SPSS***

* **Factor Analysis**

Factor analysis is used to examine underlying factors within variables or items. It searches for patterns and correlation within the different items or variables and creates new factors out of similar variables. You want to examine the influence of intelligence and hours of study on the grade.



* **Descriptive Statistics**

Descriptives statistics contain ‘frequencies’, ‘descriptives’ and ‘explore’. You’ll need these functions to describe your data. If you, for example, want to know how many men and how many women your research contains, you’ll need the ‘frequencies’ function.



* **ANOVA (analysis of variance)**

The ANOVA compares means (just like the t-test) and is used to test hypotheses. The ANOVA has the power to compare 2 or more groups. When do you use the ANOVA?The ANOVA is used when you want to compare the means of 2 or more groups. 