

# SPSS Lab 9

## Exam Practice: Parametric and Non-Parametric Tests

### *Demo 1*

We will carry out a complete statistical analysis of the HIV.sav data. The file contains the results of Phylogenetic (evolutionary) analysis of the HIV viral sequences isolated from 8 different patients at different times during a longitudinal study.

### Description of Data

Several studies have shown a direct correlation between CD4+ T cell decline and the proliferation of X4 strains. X4 strains are therefore linked to progression to AIDS. The variable “X4” indicates whether the sequence contains a mutation indicative of the X4 phenotype. Recombinant sequences can be identified by an average breakpoint position (avg\_bkp) larger than 0. The recombination process in HIV uses the mechanism of copy-choice replication, in which the viral polymerase switches between different RNA templates during transcription to give rise to mosaic genomes. Mosaic genomes and genes can be identified through a computational sliding window method like the one implemented in Sliding MinPD. The results shown here were obtained by Sliding MinPD. Every single virus sampled from a patient is identified by a sequence id (seq). The time when the virus was sampled is shown in variable “time”.

- Hypothesis

Create a scatterplot of only the recombinant sequences of patient 1 ( patient =1 and avg\_bkp>0) . You will need to filter out all rows except those for patient 1 and with bkp>0. Go to Graphs\Legacy Dialogs\Scatter\Dot... Select “Simple Scatter” and click on Define. Move avg\_bkp to the Y-axis, time to the x-axis, and X4 to “Set Markers by”. Click Ok. Try to create a similar scatterplot using Graphs\Interactive\Scatterplots... (use help to do it right). Which graph do you prefer? Now run a few other scatterplots for the other patients (or all patients) and see if you can find a pattern, find some difference between the X4 recombinants and the non-X4 recombinants. You should see a difference! Now it’s time to state your hypothesis:

- a. Write down a 1-sided and a 2-sided hypothesis
- b. Although the sample size is small we will use a t-test. Should we use an independent samples or a paired samples test?

- Prepare Data

In order to carry out a comparison analysis of the average X4 breakpoint position with the non-X4 breakpoint position with need to summarize the data such that for every patient we have two rows, one with the average breakpoint position for non-X4 recombinant sequences, and the other with the average breakpoint position for X4 recombinant sequences. Recombinant sequences are those have a breakpoint larger than 0.

- a. Start by filtering out the rows of non-recombinant sequences

- b. In order to summarize the data in the form described above, go to Data\Aggregate...Click on the help button to find out how to obtain the data in the required format. Remember to select “Create a new dataset...” so that you do not lose your initial data. (Hint: Move patient and X4 to “break variables” and avg\_bkp to “summaries of variables”)
  - c. Once you have a summary of the breakpoints per patient, you still need to modify the data in order to carry out a paired t test. Run the Data/Restructure wizard to restructure the rows into columns as needed by the paired t test. Click on help to guide you in the process. (Hint: Patient is the identifier, X4 the Index. Later: group by Index.)
- Run the test  
First let's test if the data is normally distributed! Create a new field called diff, which contains the difference between the X4 and the non-X4 average breakpoints. Go to Analyze\Descriptive statistics\Explore and select diff as the dependant variable and click on “Plots..” and select “Normality Plots with Tests”.
    - a. What do the results mean? Is the data normally distributed?
    - b. Now run a paired t-test to test your hypothesis. What is the 2-sided (2-tailed) p-value. What is the 1-sided p-value? What do the results mean?
  - Run an appropriate non-parametric test!  
Describe the results. Which method (parametric or non-parametric) is the preferred method in this case and why?

### ***Lab Assignment 1***

Repeat the analysis from above, but this time for the variable “recsignal” which measures the intensity of the recombination signal in all sequences (recombinants and non-recombinants).