IRootLab Tutorials

Mann-Whitney “U”-test per wavenumber

Julio Trevisan

30/Jan/2012

[Creative Commons Licence](http://creativecommons.org/licenses/by-nc-sa/3.0/)  
This document is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](http://creativecommons.org/licenses/by-nc-sa/3.0/).

[Introduction 1](#_Toc315696749)

[Conventions 1](#_Toc315696750)

[Tutorial 1](#_Toc315696751)

[The dataset 1](#_Toc315696752)

[Preparation 3](#_Toc315696753)

[One-step way 5](#_Toc315696754)

[Alternative way 7](#_Toc315696755)

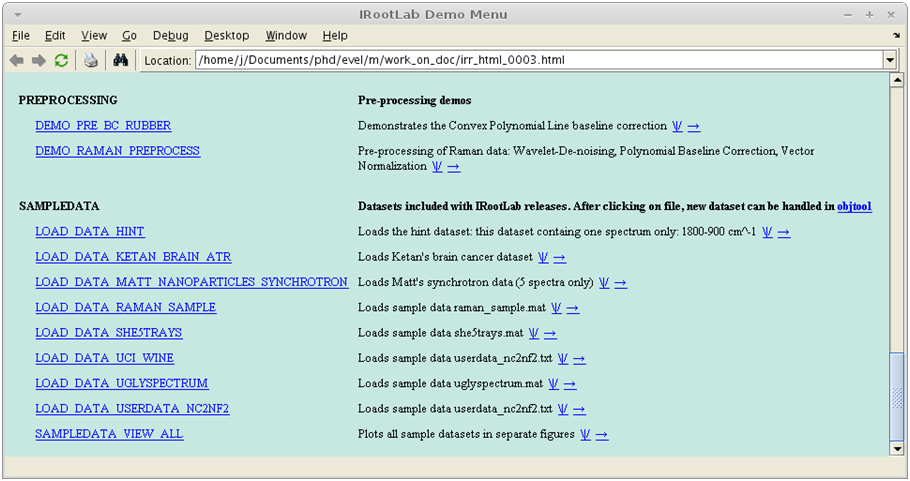
# Introduction

This tutorial shows how to perform a Mann-Whitney “U”-test per wavenumber and get a curve similar to a loadings curve.

# Loading data

This tutorial uses Ketan’s Brain data[1], which is shipped with IRootLab.

1. At MATLAB command line, enter browse\_demos
2. Click on “LOAD\_DATA\_KETAN\_BRAIN\_ATR”
3. Click on “objtool” to launch objtool



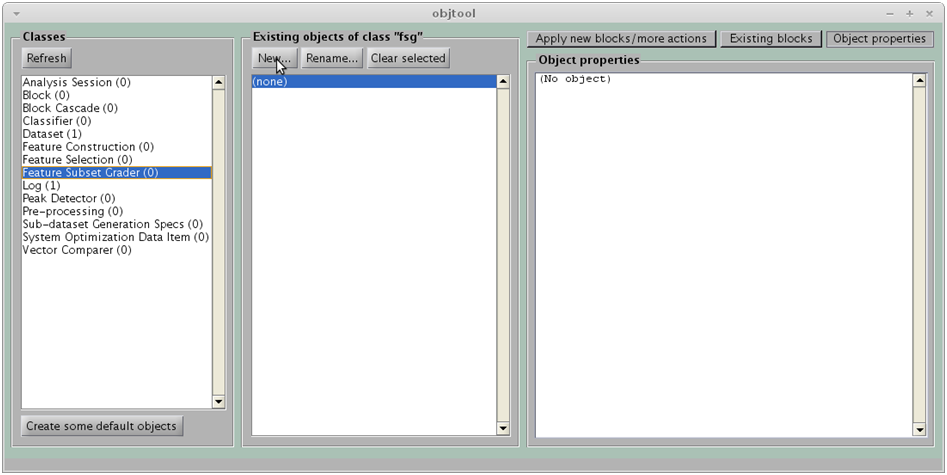
**2**

**3**

# Preparation

This step creates a Feature Subset Grader (FSG) object. A FSG object evaluates features (wavenumbers) according to some criterion (in our case, the criterion will be the U-test).

1. Click on Feature Subset Grader
2. Click on New…



**1**

**2**

1. Click on U-test
2. Click on OK

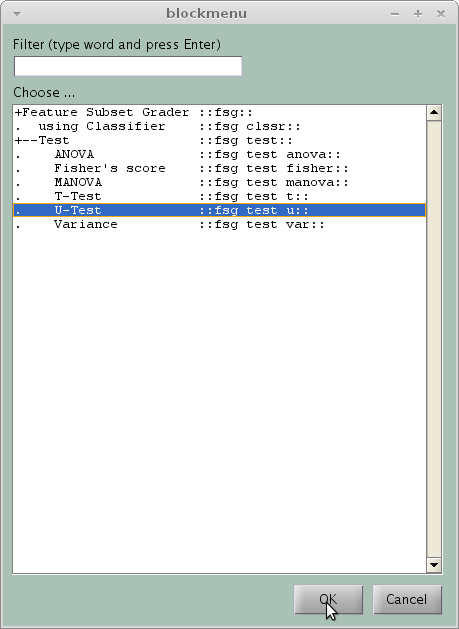
**Note** –You can use another test instead of the U-test. However, the U-test is theoretically more appropriate than, for example, the T-test, because the data variables probability distributions may be skewed, bimodal etc. So, the U-test is appropriate because it is non-parametric: it does not try to guess any parameters of an assumed distribution. In opposition, the T-test assumes normal (Gaussian) distribution.

Fisher’s score is the between-class variance divided by the within-class variance;

ANOVA is the multi-class equivalent of the T-test;

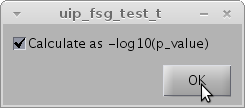
Variance calculates the variance of each wavenumber;

MANOVA is not appropriate for the context. It is a multivariate test, whereas we are interested in univariate measures here.

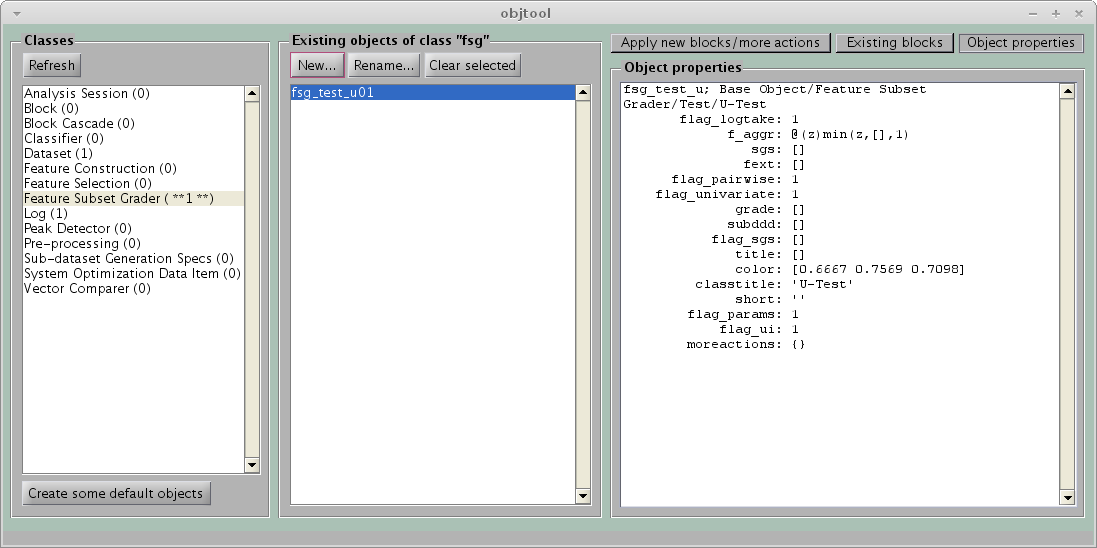


1. Click on OK

**Note** - Because the checkbox is checked, the curves obtained will have –log10(*p-value*) in the y-axis, rather than the p-value itself. This form is convenient because it transforms the *p*-value into a “significance measure”. The lower the p-value, the higher the significance.



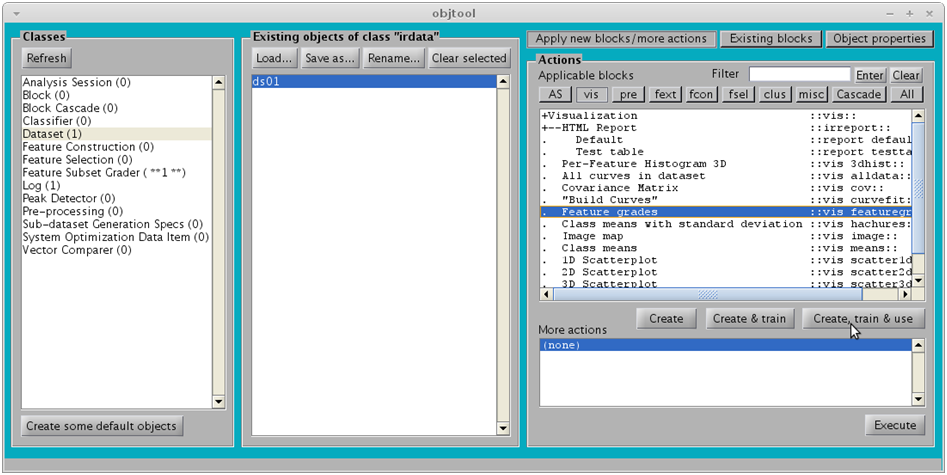
An object called fsg\_test\_u01 should appear in the middle panel:



# One-step way

This way is quicker, but will draw the curve only (see also **Alternative way** below).

1. Click on Dataset
2. Click on Apply new blocks/more actions
3. Click on Feature grades
4. Click on Create, train & use



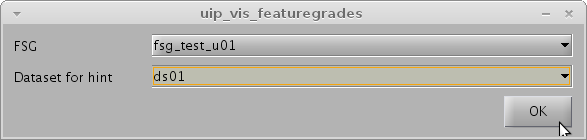
**16**

**17**

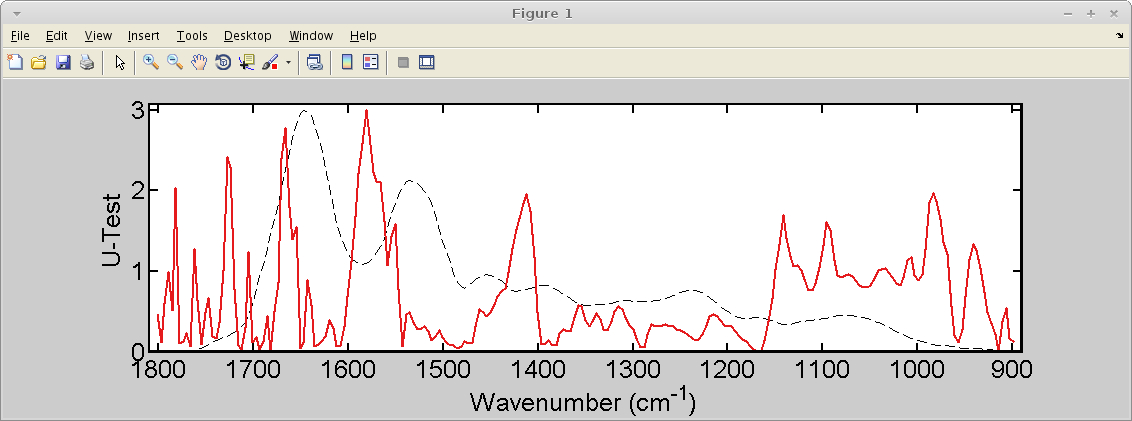
**18**

**19**

1. FSG: specify fsg\_test\_u01 created previously
2. Dataset for hint is optional. If specified, a dashed black spectrum will be drawn on the background of the figure. The objective is to help with the biochemical interpretation of the U-test per wavenumber curve.
3. Click on OK



The following figure should appear:

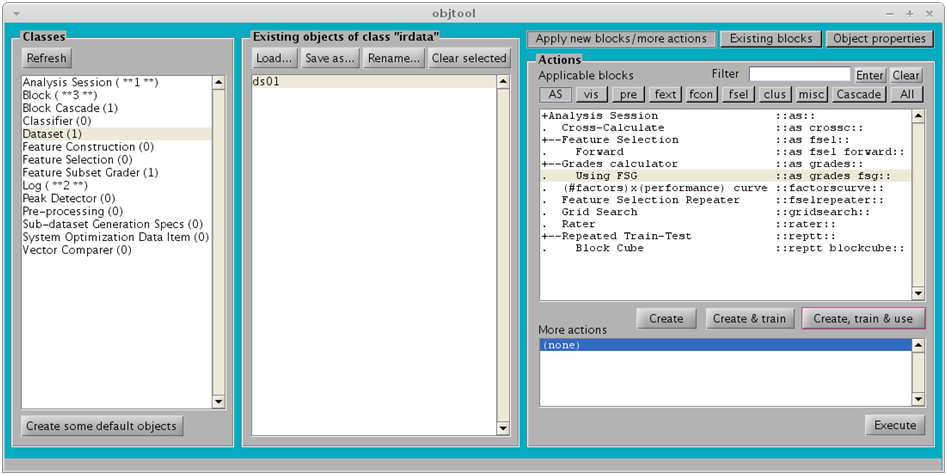


# Alternative way

This way has more steps, but the generated figure will have additional elements:

* Non-significant areas hachured in gray
* Significance threshold drawn as a dashed horizontal line

1. Click on Dataset
2. Click on Apply new blocks/more actions
3. Click on AS (Analysis session)
4. Click on Using FSG
5. Click on Create, train & use



**18**

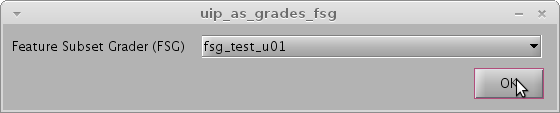
**16**

**17**

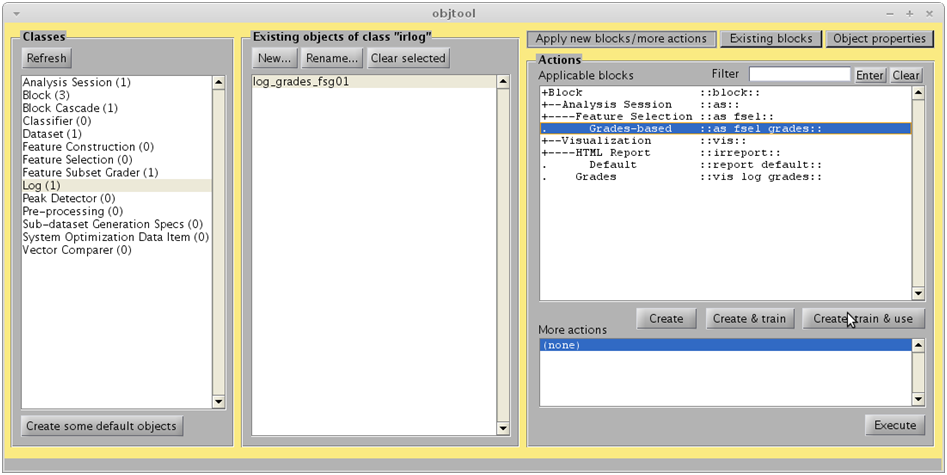
**19**

**20**

1. FSG: specify fsg\_test\_u01 created previously
2. Click on OK (the result will be a Log)



1. Click on Log
2. Click on Grades-based
3. Click on Create, train & use



**23**

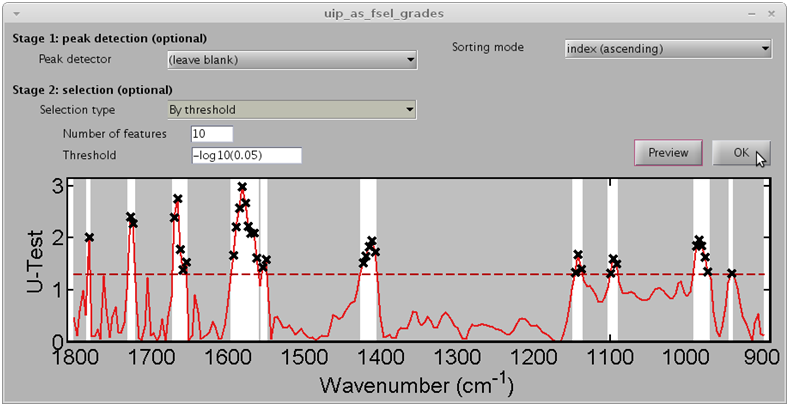
**24**

**25**

1. Selection type: choose By threshold
2. Click on OK (The result will be another log)

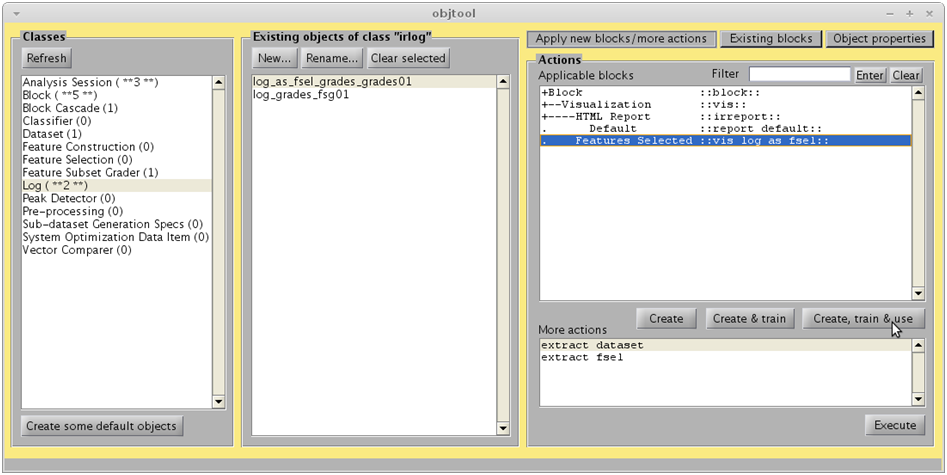
**Note** that the Threshold is specified as **-log10(0.05)**, where **0.05** is the significance level. The value –log10(0.05) is approximately equal to 1.3

You can Preview the figure (don’t worry about the “X” marks).



**26**

1. Click on log\_as\_fsel\_grades\_grades01
2. Click on Features Selected
3. Click on Create, train & use

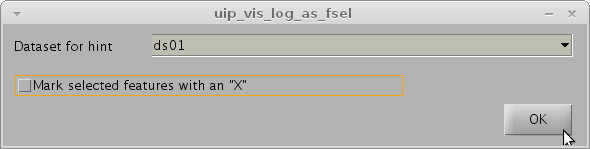


**28**

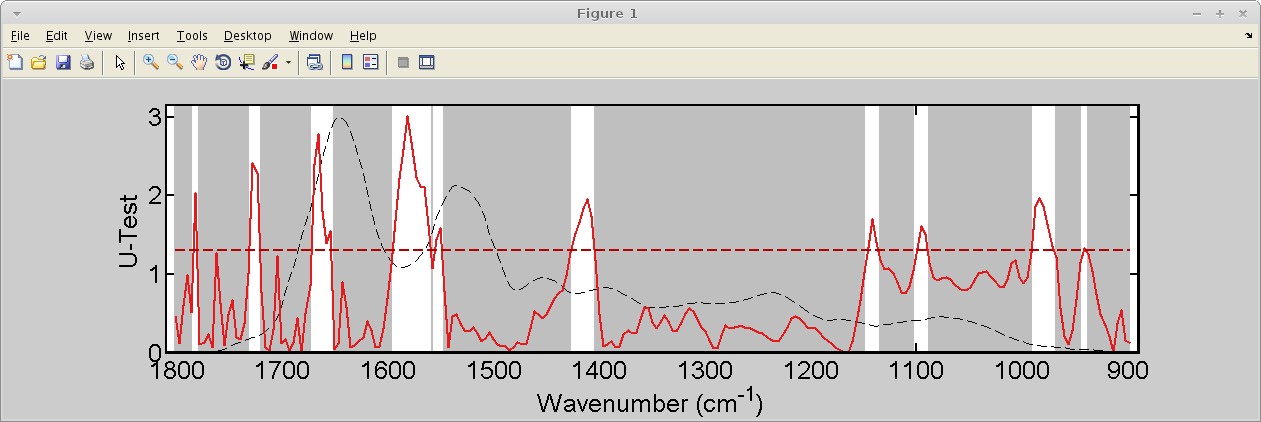
**29**

**30**

1. Dataset for hint is optional. If specified, a dashed black spectrum will be drawn on the background of the figure. The objective is to help with the biochemical interpretation of the U-test per wavenumber curve.
2. Uncheck Mark selected features with an “X”
3. Click on OK



The following figure should appear:



# **References**

[1] K. Gajjar, L. Heppenstall, W. Pang, K. M. Ashton, J. Trevisan, I. I. Patel, V. Llabjani, H. F. Stringfellow, P. L. Martin-Hirsch, T. Dawson, and F. L. Martin, “Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis,” *Analytical Methods*, vol. 44, no. 0, pp. 2–41, 2012.