Exercises to Lectures 7: Statistical Methods

This exercise is voluntary and meant for your fun. For the best solution that reaches us we offer a bottle of wine that will be handed over during the last lecture date. In case of more than one first prize candidate, dice have to decide, along the lines of this lecture...

Exercise 15 (Fitting Exercise):

For this exercise we have prepared two toy datasets with 20'000 events each in form of a binned histogram with 40 bins (of 5 GeV width) between fictive invariant mass values of 100 GeV and 300 GeV. The datasets are stored in two different *root* input files

- data-scenario-A.root
- data-scenario-B.root

You can find these two input files and two further files *fit_data.C* and *FitFunctions.h* in the tarball stored under this web link:

http://www-ekp.physik.uni-karlsruhe.de/~quast/vorlesung/TP2Higgs/Uebungen/fitting-exercise.tar.gz

we assume that you use the same setup as for **Exercise 14**. You can download and extract the linked file to some local space in your virtual machine.

a)
First investigate the content of the directory folder with name *fitting-exercise* that you have extracted. Open the root files that you find in there and inspect them in the *root Browser*:

```
> cd ./fitting-exercise/
> root -l data-scenario-A.root data-scenario-B.root
> TBrowser t
```

b)
Next make yourself familiar with the example macro *fit_data.C* that we provide. For this purpose open it with your favorite editor and go through the code and the comments that you find there. Finally you can execute the macro running the following commands:

```
> root -l
> .x fit_data.C++(120)
```

The argument 120 corresponds to the mass point that you want to test. You can find an explanation of the syntax used for the last command in the head of the file.

The macro loads one of the two *root* files (to be defined in line 81 of the macro), reads the data histogram from it and applies a maximum likelihood fit of a simple physics model to it. We have defined this physics model in the header file *FitFunctions.h*. It is made up from a falling exponential function with a potential signal in form of a Gaussian distribution on top. For simplicity reasons the width of the signal is restricted to be 20% of the tested mass value. The macro will prompt a plot of the data histogram and the fitted physics model and a significance value that has been estimated exactly as discussed in the lecture.

Here comes the exercise: in one file we have hidden a signal. Where is it and what significance does it have? For the first price we want three correct statements: (i) the file; (ii) the mass and; (iii) the significance of the signal. We open the secrete during the last lecture date. This will be the last exercise that we distribute for this course.

Note 1: the signal that you fit has a width. It does not make too much sense to choose a mass value at the very borders of the histogram for your fit. Make up you mind and decide from where on fitting makes sense.

Note 2: we do not care how you obtain the solution. You can modify the example macro how ever you like or use it just as is. The latter requires a bit more hand work that the computer can do for you in a reproducible way if you modify the macro correspondingly.

Note 3: if you want to learn more about the root classes that have been used throughout the example the easiest way to obtain this information is just to use *google*. Practical key words are the class name (e.g. *TF1*) plus *root*.