Statistical methods – an introduction (SS 2022)

Problem set 2

Decide carefully whether you want to perform the numerical implementations either in IDL or in `phython`. In case, discuss with your tutor.

Problem 1 [8 points] *Working with histograms*

To solve the following problems, calculate the variance with denominator 'N' (instead of 'N-1'). Please have your IDL-routines/`phython`-scripts handy during the tutorial. For specific functions and routines, check the internet (e.g., search for 'idl histogram' or 'python histogram').

Here is a data-set of 80 numbers

```
79  89  77  83  76  91  87  89  84  68
77  73  71  77  76  77  64  82  65  46
21  77  66  72  74  65  70  70  73  76
37  63  63  50  68  64  73  44  46  33
69  61  65  54  52  77  60  65  61  26
64  49  62  53  55  62  43  50  51  54
24  37  30  41  34  31  49  43  39  48
89  33  23  35  35  14  21  30  33  26
```

a) Draw (by hand) the corresponding histogram, with bin size 10 and bins starting at 0, i.e., bin 1 = [0,10), bin 2 = [10,20), ..., bin n = [90,100]. Note that all but the last bin of a histogram are half-open!

b) Write a small IDL-routine/`python`-script which calculates, from the *raw* data, mean $\bar{x}$, median, standard deviation σ and skewness γ_1 of the distribution. When calculating σ, use a denominator of $\sqrt{N}$ (instead of $\sqrt{N-1}$).

Hint1: Paste the above data directly into your routines, to avoid typos.
Hint2: In both languages, use `moment, median`; check internet for these commands/ routines!!!

Determine as well, by 'eye', the mode of the raw data. Here, you might use `sort`.

c) Calculate the same quantities (except for skewness) from your histogram, and compare the results.

d) **IDL**: Inspect the IDL procedure `my_histogram.pro` from the lecture's homepage (directory `sheet2`). This routine uses the system-supplied function `histogram` and adapts it for convenient use. Read the documentation for `histogram`, and try to understand the additional operations of `my_histogram.pro`.
Python: Inspect the `python` script `my_hist.py` from the lecture's homepage (directory `sheet2`). This script uses the `matplotlib.pyplot` function `hist` and adapts

it for convenient use. Read the documentation for `hist`, and try to understand the additional operations of `my_hist.py`.

e) Write a small IDL-routine/`python`-script (using `my_histogram.pro` or `my_hist.py`, respectively) to create and plot a histogram for the above data, in dependence of bin size and start of first bin (as input parameters). Complete the routine by calculating $\bar{x}$, σ and γ_1 *from the histogram* (hint for IDL: `n_elements(x), total(x)`).

f) Compare at first with your previous results from c) and convince yourself that the routine works reliably. Investigate the reaction to different bin sizes 5, 10, and 15 (with start of first bin at '0').

g) Compare then what happens (for bin size 10) when the start of the first bin is shifted from 0 to 1, 2, 3, ... 9. Write a small table for the corresponding results for $\bar{x}$, σ and γ_1 (calculated from the histogram), and compare with the results for the raw data. Plot the histograms with the smallest and largest skewness.

To be covered in the tutorial:

- non-uniform bins
- how many bins?
- start of the first bin?

Problem 2 [4 points] *Expectation value and variance of a convolution*

Prove the results for the expectation value and the variance of a convolution, as provided on page 41 of the script. Do NOT use the 'calculation rules' for expectation value and variance, but calculate these values directly from their definitions and the distribution $f(x')$ as given in the script.

Hint: Calculate the variance via $E(x'^2)$ and $E(x')$. Use $x' = x + u$, and remember that the individual pdf's are normalized.

Have fun, and much success!