Time	Group	Submission in Moodle; Mails with subject: [SMD2023]
Th. 12:00–13:00	A	lukas.beiske@udo.edu and tristan.gradetzke@udo.edu
Fr. 09:00–10:00	B	jonas.hackfeld@ruhr-uni-bochum.de and ludwig.neste@udo.edu
Fr. 10:00–11:00	C	stefan.froese@udo.edu and vincent.latko@udo.edu

Exercise 1 *Binning* **0 p.**

(a) Read in the distributions for height and weight from the file `height_weight.csv`. You can find this file in the Moodle. Create histograms for both distributions using matplotlib with 5, 10, 15, 20, 30, 50 bins respectively in a single figure, split into 3×2 subplots. What differences do you notice? Which binning seems reasonable to you? Why?

(b) What happens if you use data from much more than 250 people? To what extent might it make sense to use different numbers of bins for the two data sets? Specify a reasonable minimum bin width, as well as the position of the bin centers.

(c) Draw 10^5 uniformly distributed integers from the interval 1-100. Plot a histogram using bins that are spaced equidistantly in logarithmic space[1], using a logarithmic scale on the x-axis. Again, use different numbers of bins (analogous to **(a)**). Which effects are noticeable depending on the binning?

Exercise 2 *Chi-Squared* **0 p.**

(a) Generate 500 random numbers from a chi-squared distribution with 5 degrees of freedom using the `numpy.random.default_rng` known from the lecture.

(b) Using the random numbers generated earlier, create a one-dimensional histogram with error bars (the errors per bin should be $\sqrt{N_i}$ with N_i entries per bin i).

(c) Plot the histogram and the true density `scipy.stats.chi2.pdf` of the distribution appropriately (*tip:* normalization).

(d) Use the method `scipy.stats.chi2.fit` to perform a fit to the sample drawn in a) (*note:* Such a fit routine is called a *maximum likelihood fit*).

(e) Now plot the histogram along with both the fitted and true chi-squared distributions.

Exercise 3 *Birthdays* **0 p.**

(a) Estimate how many people are needed so that the probability for two birthdays to fall on the same date is greater than 0.5?

(b) Now calculate: in a group of n, what is the probability that at least two have their birthdays on the same day[2]? For which n does the probability become 0.5 or greater? Neglect leap years.

[1] See `https://numpy.org/doc/stable/reference/generated/numpy.geomspace.html`
[2] We assume that someone born on February 29 celebrates March 1 in years without that date