Applied Statistics

Exam in Applied Statistics 2024/25

This take-home exam was distributed Thursday the 16th of January 2025 at 08:00. A solution in PDF format must be submitted at **www.eksamen.ku.dk by 20:00 Friday the 17th**, along with all code used to work out your solutions (as appendix). Links to data files can also be found on the course webpage and github. Working in groups or discussing the problems with others is **NOT** allowed.

Thank you for all your hard work, Malthe, Beatrice, Rashmi, Marcela, Mathias, & Troels.

The knowledge of certain principles easily compensates the lack of knowledge of certain facts.
[Claude Adrien Helvétius, 1759]

I – Distributions and probabilities:

1.1 (6 points) Every day, you roll a normal die, and if you get a six, you do roll the die 100 times and do as many pushups as you get sixes. Otherwise you don't do any pushups.

- What is the distribution of days between doing pushups?
- What is the mean, median, and standard deviation of number of pushups in 30 days?

1.2 (4 points) The Djoser pyramide in Egypt is North-South aligned to 3 degrees.

- Estimate the probability that the pyramid is North-South aligned by coincidence.

II – Error propagation:

2.1 (5 points) Water on Earth ($\oplus$) has a Deuterium to Hydrogen ratio of $r_{\oplus} = (149 \pm 3) \times 10^{-6}$. The hydrogen of the proto-solar system ($\odot$) has a ratio of $r_{\odot} = (25 \pm 5) \times 10^{-6}$, while that of comets ($C$) have been measured to be $r_C = (309 \pm 20) \times 10^{-6}$.

- From these numbers, what fraction of water on Earth do you estimate come from the original proto-solar system, and what fraction do you attribute to comets?

2.2 (8 points) You run a detector for a time interval of $\Delta t = 98.4$s, during which the detector yields $N = 1971$ counts. The time interval uncertainty is $\sigma_{\Delta t} = 3.7$s, indepedent of Δt.

- What is the rate $r = N/t$ and its uncertainty?
- How long should you measure to get a relative uncertainty on the rate r below 2.5%?

2.3 (14 points) The file **www.nbi.dk/~petersen/data_PylonPositions.csv** contains position measurements (both with and without uncertainties) of four pylons for a bridge.

- Using measurements without uncertainty, determine the four pylon positions.
- Using measurements with uncertainty, determine the four pylon positions.
- Combine the two measurements. Do they match each other?
- Test if the four measured pylon positions are equidistant.
- The pylon distance should be 40.83m with a tolerance (i.e. maximally allowed deviation) of 1.05m. Do the pylon positions live up to this requirement?

III – Simulation / Monte Carlo:

3.1 (8 points) Circles A and B are centered at (0,0) and (3,7) and have radii of 6 and 4, respectively.

- What fraction of A overlaps with B? And conversely, what fraction of B overlaps with A?
- If the circles were 4D "hyperballs" centered at (0,0,0,0) and (3,7,-1,2), respectively, and with the same radii, what would the answers then be?

3.2 (15 points) You want to simulate the radial material distribution $m(r)$ from a uniform explosion.

- Generate 50000 x, y, and z values in the range $[-1, 1]$ and plot the spherical coordinate r.
- Selecting only points with $z > 0$ and $r < 1$, what distributions in θ and ϕ do you obtain?
- How would you produce random velocities v according to $f(v) = (v/v_0)^2 \exp(-v/v_0)$?
- Given $v_0 = 100$ m/s and that the radial distance of material r as a function of velocity is $r(v) = \sin(\theta)v^2/g$ ($g = 9.82\text{m/s}^2$), simulate 10000 values of θ and v. Combine these to obtain values of r, and plot the resulting distribution $m(r)$.

IV – Statistical tests:

4.1 (12 points) You get a permanently closed box with 20 normal (i.e. six-sided) dices in. One of the dice is potentially fake, with all the sides having the same (unknown) value. You can shake the box and see the resulting 20 dice roll as many times as you like.

- Simulate 100 box rolls and plot the die frequencies, both with and without a fake die in.
- For both of your simulated datasets, test if there is a fake die or not.
- How many rolls would you require before you would argue, that you could tell the difference?

V – Fitting data:

5.1 (14 points) The file **www.nbi.dk/~petersen/data_InconstantBackground.csv** contains molecular interspacing measurements d (in nm) from a scattering experiment.

- Plot the data and test to what extend the background in the range [8,10] is uniform.
- Fit the three Gaussian peaks at around $d = 0.9$, 3.4, and 5.9 nm, including local background.
- Test if the peaks have the same intensity, i.e. number of measurements in them.
- Try to fit the entire spectrum or parts of it best possible and comment on your results.

5.2 (14 points) The table below lists the North-South alignment of Egyptian pyramids (in arc minutes).

Pyramid	1.Meidum	2.Bent	3.Red	4.Khufu	5.Khafre	6.Menk.	7.Sahure	8.Nefer.
Align. year	2600 BC	2583 BC	2572 BC	2554 BC	2522 BC	2489 BC	2446 BC	2433 BC
East Align.	-20.6 ± 1.0	-17.3 ± 0.2	-8.7 ± 0.2	-3.4 ± 0.2	6.0 ± 0.2	12.4 ± 1.0	23 ± 10	30 ± 10
West Align.	-18.1 ± 1.0	-11.8 ± 0.2	–	-2.8 ± 0.2	6.0 ± 0.2	14.1 ± 1.8	–	–

- Test to what extend the East (E) and West (W) alignment values agree.
- Combine East and West values. Include systematic uncertainties to ensure agreement.
- If the alignments were done using circumpolar stars, these drift with Earth's precession. Test if the alignment of the pyramids shifts linearly as a function of time.
- The astronomically predicted shift as a function of time is 2467 BC + 0.274 arc min./year. Does the slope of the linear fit match the predicted value?
- What alignment date of Khufu (historically 2554 ± 100 BC) is the astronomical prediction?

The outcome of a repeated process follows not chance but statistics.