# Applied Statistics

Re-Exam in Applied Statistics 2025/26

This take-home exam was distributed Thursday the 16th of April 2026 at 08:00. A solution in PDF format must be submitted at **www.eksamen.ku.dk by 20:00 Friday the 17th of April**, 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, Nina, Janni, Gabriela, Preet, Clara, Marc, Mathias, & Troels.

---

*Science may be described as the art of systematic oversimplification.*

[Karl Popper, Austrian/British philosopher 1902-1994]

---

**I – Distributions and probabilities:**

**1.1** (8 points) Three professors arrive daily at 8:15, and meet if they arrive within 1 minute of each other. Assume their arrival time follows a Gaussian $G(\mu = 8{:}15, \sigma = 0.10)$ distribution.

- What is the probability that the three of them meet one morning?

One professor is away 40% of the time, another really has a spread $\sigma$ of 30 minutes, and the last follows a log-normal distribution with mean 8:15 and a width 60 minutes.

- How often should they expect to meet in 1000 days?

**II – Error propagation:**

**2.1** (7 points) Let $z = xy/((x-y)^2+1)$, with $x = 9.1 \pm 0.3$ and $y = 7.8 \pm 0.6$.

- If $x$ and $y$ are uncorrelated, what is the value and uncertainty of $z$? What if $\rho_{xy} = 0.87$?
- Produce a 99% Confidence Interval for $z$.

**2.2** (15 points) Iridium and Osmium were both discovered in 1803 in residues of Platinum. The file **www.nbi.dk/~petersen/data_OsmiumIridiumDensity.csv** contains density measurements of the two at 20°C, which are very similar and hard to measure.

- Determine the mean and median for each of the two samples and fit each with a Gaussian.
- Do you suspect any mismeasurements? If so, argue for their exclusion, and repeat the above question.
- Test the hypothesis that Osmium is the heaviest of the two elements (and all others).

**2.3** (8 points) Given three side lengths of a triangle $a = 1.04 \pm 0.03$, $b = 0.99 \pm 0.06$, and $c = 1.06 \pm 0.02$), Heron's formula gives the triangle area as $A = \sqrt{s(s-a)(s-b)(s-c)}$, where $s = (a+b+c)/2$.

- Is this triangle consistent with being equilateral triangle?
- Is this triangle consistent with being a unit side length triangle?
- What is the area of this triangle and its uncertainty?

**III – Simulation / Monte Carlo:**

**3.1** (16 points) A radar signal follows the PDF $A = A_0 \, G(t = 0, \sigma_t = 5) \cos^2(\omega t)$, with $\omega = 10$.

- Generate 20000 values of $t$ from the $A$ distribution. Discuss how this is done best.
- Fit the distribution and determine the precision with which $t_0$ can be determined.

Now apply a time resolution effect, shifting all $t$-values by a random Gaussian number $G(0, \sigma_{res})$.

- Using $\sigma_{res} = 0.1$, fit the oscillations in the spectrum with any function.
- Estimate the minimum resolution $\sigma_{res}$ needed in order to be able to fit the oscillatory spectrum.

**IV – Statistical tests:**

**4.1** (16 points) The file **www.nbi.dk/~petersen/data_GlacialFlow.csv** contains data on the position of an ice element in a tropical mountain glacier ($x$ in m) as a function of time ($t$ in days). The positional resolution of the glacier element is 10m.

- Fit the data with a linear function, and comment on the data and the result.
- Test to what extend there is a sudden change in the flow rate, and if so, when.
- Test if an oscillatory term in the fit function to include summer/winter variation is needed.

**4.2** (8 points) You shoot 80 penalties in football and score as follows (0: No goal, 1: Goal):

00110100111011010001011011011111101100110101110110000000110111010011010101110010

- Is the scoring order random, or are you occationally in "flow"?

**V – Fitting data:**

**5.1** (11 points) The strong coupling constant $\alpha_S$ changes with energy scale $Q$, as given in the file **www.nbi.dk/~petersen/data_RunningCoupling.csv**

- Fit the data with the function $\alpha_s(Q) = \alpha \log(Q/\Lambda)$, where $\alpha$ and $\Lambda$ are fit parameters.
- Try other functional forms, using maximally three fit parameters, and evaluate these.
- Evaluate $\alpha_s(Q)$ for $Q = 80.4$ GeV (the $W$ mass) including an uncertainty estimate.
- Given a measurement of $\alpha_s(91.2) = 0.127 \pm 0.005$, how does the above estimate change?

**5.2** (19 points) The file **www.nbi.dk/~petersen/data_BohriumVsEinsteinium.csv** contains lifetime (in $s$) and energy (in MeV) measurements from decays of a composit sample containing Bohrium-272 (Bh) and Einsteinium-242 (Es). The expectation is that the Bh decay times are exponential and the lifetime is the lowest, while the Bh energy desposits are Gaussian and the highest.

- Fit the energy distribution of the entire dataset with two Gaussian functions. Comment on the result, and possibly improve the fit.
- Consider if any decay time measurements are suspecious and should be omitted. Argue why.
- Select Bh and Es decays based on their decay energy, and fit each of their corresponding lifetime measurements with exponential functions.
- See if you can improve the lifetime goodness of fit by including possible background PDF in fit function.
- Quantify how certain are you that the lifetimes are different.

---

*Don't worry too much about statistics! Just tell us what you do, and do what you tell us.*

[Roger Barlow, ICHEP conference 2006, Moscow]