Applied Statistics

Problems in fundamental concepts of applied statistics

The following problem set is a review of the fundamental concepts of applied statistics. It covers the curriculum of the first two weeks of the course, and a written solution is to be handed in Tuesday the 20th of September. Problem solving in groups is allowed (please note this in the solutions), but individual solutions are required.

The use of computers and modifications of the programs we have used in class is both allowed and adviced, and will for certain problems be almost necessary.

Good luck, Troels

I – Distributions and probabilities:

1.1 Let x be distributed according to the PDF $f(x) = x^2$ in the interval $[0, C]$.

- For which value of C is the PDF $f(x)$ normalized?
- What is the mean and spread of x?

1.2 Little Peter goes to the casino and puts money on one number at a time ($p = 1/37$). If he is not cheating, what are the chances that he will win more than 3 times in 100 games?

1.3 Calculate the mean and spread of the following PDF: $f(x) = \ln(x)$, $x \in [1, e]$.

II – Error propagation:

2.1 Let $x = \cos(y - 2z) + z^3$. Given that $y = 1.0 \pm 0.2$ and $z = 0.2 \pm 0.1$, and $\rho_{yz} = -0.85$, what is x and the uncertainty in x?

2.2 Ten students have measured the speed of light and got the following results:

Measurement	1	2	3	4	5	6	7	8	9	10
Result (10^8 m/s)	3.61	2.08	3.80	2.53	2.82	2.48	2.43	3.56	4.43	3.08

- What is the average and the uncertainty on the average?
- If another group of students measured 3.04 ± 0.11, how would you combine the two?

2.3 A student measures the speed of sound 20 times and concludes that the spread in the results is 20 m/s. If the sources of uncertainties are many and random, which distribution should the measurements follow? And how many measurements would be needed to reach an uncertainty in the mean of 4 m/s?

2.4 The initial activity N_0 and lifetime τ of a radioactive source is known with a relative uncertainty of 1%. When estimating the activity $N = N_0 e^{-t/\tau}$ the uncertainty will initially be dominated by the uncertainty in N_0 and later by the uncertainty in τ. For what value of t/τ will the uncertainties contribute equally to the uncertainty on N?

III – Monte Carlo: (For this part the use of computers is adviced. Plots can be enclosed in the solution).

3.1 Let $f(x) = e^{-x^3+2x^2} - 1$ be proportional to a PDF for $x \in [0, 2]$.

- Which method should be used to generate numbers according to this distribution? Explain?
- Make an algorithm, which from a uniform distribtion in the interval $[0, 1]$ generates numbers following the PDF $f(x)$.
- Determine $\int_0^2 f(x)dx$ and its uncertainty by using this algorithm and use the result to normalize $f(x)$.

3.2 Make an algorithm, which simulates 1000 throws of three dices (or do this by hand!).

- Make a table and/or a histogram of the frequency (with errors) of each possible sum. Also, calculate the expected number of rolls for each sum. Do they "by eye"match?
- Calculate the χ^2 for the agreement between data and expectation, and determine the probability for obtaining such a χ^2 value or something more extreme.

IV – Estimators:

4.1 In the past years several groups of students have been measuring the lifetime of the muon in the basement at NBI. Their results and estimated uncertainties are listed below:

Group	1	2	3	4	5	6	7	8	9	10
Result (μs)	1.82	1.95	1.46	2.12	2.09	1.70	1.93	1.87	2.25	2.16
Uncertainty (μs)	0.06	0.09	0.12	0.13	0.24	0.11	0.07	0.10	0.21	0.14

- Calculate the average lifetime and its uncertainty both with an unweighted and a weighted calculated. Then use these averages to calculate a χ^2 and the probability of obtaining such χ^2 values or something more extreme.
- Is there a measurement, which does not fit very well in? Why?
- Repeat the previous calculation excluding the least probable measurement.
- How well do the results (unweighted and weighted) match the true value of τ_μ?

V – Fitting data:

5.1 An experiment has yielded the following results, where the uncertainty on y, σ_y has been estimated to be 0.06:

x	y	x	y	x	y	x	y
-2.0	-.29	1.0	0.06	4.0	0.63	7.0	0.81
-1.0	-.19	2.0	0.33	5.0	0.89	8.0	1.04
0.0	0.04	3.0	0.57	6.0	0.80	9.0	0.94

- Assume a linear relation between x and y and make a χ^2-fit to data.
- Calculate from this χ^2 and the number of degrees of freedom the probability of obtaining such a χ^2 value or something more extreme. Is it a good fit?
- Try other hypotheses for the relation between x and y, and discuss their validity.

Bonus problem:

6.1 Two years ago, the number of physics students starting at KU was 119. This year it is 171.

- How certain can you be, that this is not just a statistical fluctuation?

What we have to learn to do, we learn by doing. [Aristotle, 384-322 bc]