Applied Statistics

Problems in fundamental concepts of statistics

The following problem set is a review of the fundamental concepts of statistics, which we have covered till now. It will be handed out Friday the 17th of September 2010, and a written solution is to be handed in Tuesday the 21st of September 2010 by 13:15. Problem solving in groups is allowed, but separate 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 A student has measured the speed of light ten times and gotten the following results:

Måling	1	2	3	4	5	6	7	8	9	10
Result (10^8 m/s)	3.61	2.00	3.90	2.23	2.32	2.48	2.43	3.86	4.43	3.78

- What is the average, width, and the uncertainty on the average?
- If $m = 1.02$, $\sigma_m = 0.05$, and $\rho_{mc} = -0.2$, what is the error on $E = mc^2$?

2.2 Let $x = 2yz + z^2$. Given that $\sigma_y = 0.2$, $\sigma_z = 0.3$, and $V_{yz} = 0.6$, what is the uncertainty in x?

2.3 A student measures the speed of sound 20 times and concludes that the spread in the results are 12 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 2 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 to 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 a Monte Carlo algorithm, which simulates 200 throws of two dices (or do this by hand!).

- Make a histogram of the frequency of each possible sum.
- Plot the statistical uncertainties of this distribution along with the expected one.
- 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 and spread of the measurements along with the χ^2 and the probability of obtaining such a χ^2 value or something more extreme both in an unweighted and a weighted calculation.
- 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 hypothesis for the relation between x og y, and discuss their validity.

Bonus problem:

6.1 How many physics students are there in Denmark?

- Without any further knowledge, what uncertainty would you ascribe this number?
- If a journalist claimed that the number of physics students had rissen/fallen by 3% compared to the year before, what should he be told?

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