Applied Statistics

Problem set in applied statistics 2016/2017

The following problem set is for the course applied statistics. A solution in PDF format must be sent (by email) before noon Friday the 23rd of December 2016. Working in groups and/or discussing the problems with others is allowed, but you should note your collaboration in the problem set.

Good luck, Troels, Daniel & Niccolo.

Statistics may be defined as "a body of methods for making wise decisions in the face of uncertainty".
[W.A. Wallis, US statistician 1912-1998]

I – Distributions and probabilities:

1.1 - 2015 (7 points) You pick 20 M&M's from a (large) bag, where 5% are blue (you are told).

- What distribution does the number of blue M&M's you picked follow? Why?
- If you get six blue M&M's, do you trust the fraction of these to be 5%?

1.2 - 2015 (6 points) Two bags contain red and blue M&M's. Bag 1 has 50% red M&M's, while bag 2 has 75% red M&M's. You pick a random M&M from a random bag, and it is red.

- What is the probability that you picked an M&M from bag 1?

II – Error propagation:

2.1 (8 points) An experiment to measure resistance R using Ohm's Law $V = RI$ yields the measurements $V = 5.01 \pm 0.07V$ and $I = 1.178 \pm 0.013A$.

- Assuming no correlations, what is the resistance and its uncertainty?

The experiment was designed to have a linear correlation between V and I of -0.95.

- What is the uncertainty on R then?

2.2 (8 points) Given $x = 0.56 \pm 0.02$, what is the value and uncertainty of $\exp(x)$, $\sin(x)$ and $\tan(x)$? And what if $x = 1.56 \pm 0.02$? Also, comment on the degree of Gaussianity of the uncertainties.

2.3 (10 points) The age of old rocks can be estimated from the current amounts of ^{238}U and its decay product ^{206}Pb as follows: $N(^{238}U) = N(^{238}U + ^{206}Pb) \cdot \exp(-t/\tau_{U238})$. In a rock sample from Ilulissat in Greenland one finds **2689** ^{238}U and **3952** ^{206}Pb atoms. Assuming no initial Pb206 and given the half-life $\tau_{U238} = (4.47 \pm 0.03) \times 10^9 y$, what is the age of the rock and its uncertainty?

III – Monte Carlo:

3.1 (12 points) Let $f(x) = C \sin^2(\pi/x)/\sqrt{x}$ be a PDF for $x \in [0.1, 1.0]$.

- What method would you use to produce random numbers according to $f(x)$? Why?
- Produce 100000 random numbers according to $f(x)$ and plot these.
- In order for this PDF to be normalized, what value should C have?
- Perform a fit to the produced data points. Do you manage to get a good χ^2?

IV – Fitting data:

4.1 (12 points) The data **www.nbi.dk/∼petersen/data_MuonLifetime.txt** contains the results of an experiment that measures muon decay times. The exponential signal has a background at very short times resulting from random noise around $t = 0$, and one that is constant in time.

- Determine the constant background by fitting a suitable range at high times.
- Extract the lifetime of the decay along with its uncertainty.
- Fit the entire distribution with a suitable function. Do you obtain a good fit?

4.2 (12 points) You are measuring the gravitational acceleration g by letting a magnet drop a ball, which you then measure the position (d in meters) of at certain times (t in seconds). The data can be found in the file **www.nbi.dk/∼petersen/data_FreeFall.txt**. The Gaussian uncertainties are $\sigma_t = 0.001$s and $\sigma_d = 5$mm. To begin with, consider only the first 8 data points.

- Assuming that $d(t=0) = 0$ and $v(t=0) = 0$, determine g and its uncertainty.
- Is the ball released at $t = 0$? Repeat the fit, and measure the possible offset in time, Δt.
- Now consider all 20 points and test which of the following three hypothesis matches the data best, where τ (characteristic time) and v_∞ (terminal velocity) are parameters:
  ⋆ No air drag: $d(t) \sim \frac{1}{2}gt^2$
  ⋆ Linear drag: $d(t) \sim g\tau(t - \tau(1-e^{-t/\tau}))$
  ⋆ Quadratic drag: $d(t) \sim v_\infty^2 \ln(\cosh(gt/v_\infty))/g$

V – Statistical tests:

5.2 (13 points) Benford's ("first-digit") Law states that leading digits ($d \in \{1, \ldots, 9\}$) occur with probability $P(d) = \log_{10}(1 + 1/d)$. Below is a table showing the frequency of first digits of countries size measured in km^2 and miles^2 (km^2/miles^2).

Digit	1	2	3	4	5	6	7	8	9
Frequency	58/56	34/37	22/20	21/17	10/14	14/14	11/14	7/12	10/4

- Test if country sizes in km^2 and miles^2 follow Benford's Law.
- Are the two distributions consistent with being from the same underlying distribution?

5.2 (12 points) In 1929 Edwin Hubble investigated the relationship between distance (D) and radial velocity (v) of extra-galactic nebulae. His original data from 1929 is listed below.

D (Mpc)	0.032	0.034	0.214	0.263	0.275	0.275	0.45	0.5	0.5	0.63	0.8	0.9
v (km/s)	170	290	-130	-70	-185	-220	200	290	270	200	300	-30
D (Mpc)	0.9	0.9	0.9	1.0	1.1	1.1	1.4	1.7	2.0	2.0	2.0	2.0
v (km/s)	650	150	500	920	450	500	500	960	500	850	800	1090

- Assuming uncertainties of 12% on the distance and 170 km/s on the radial velocity, fit the data to extract the constant of linear proportionality H_0 (Hubble constant), $v = H_0 D$.
- Hubble's distance measurement was biased by a factor 5.3 ± 0.3 due to the existance of two types of Cepheids. Given this correction, how good is the agreement with the modern value of the Hubble constant $H_0 = 67.80 \pm 0.77$.

Numbers are like people; torture them enough and they'll tell you anything. [Unknown]