Homework 8 (Need not submit)

1. In the file `http://math.iisc.ernet.in/~manju/UGstatprob/simulatednormaldata.txt` you will see data *simulated* on a computer from a normal distribution with unknown mean and variance. The problem is to test the hypotheses $H_0 : \mu = 2$ versus $H_1 : \mu \neq 0$.

1. Use only the first n data points for $n = 20, 50, 100, 200, 400, 500, 800, 1000$, and carry out the test for each n at significance level 0.05. Report the p-values.

2. Repeat the same tests but now assume that the variance is given to be 9.

2. Are real coins fair? Formulate this as a hypothesis testing problem and perform the test at 0.01 level of significance using the following data. Report the p-value.

1. In an experiment reported in `http://www.stat.berkeley.edu/~aldous/Real-World/coin_tosses.html`, a real coin was tossed 20000 times. The number of heads observed was 10231.

2. In another experiment reported on the same page, 10014 heads appeared in 20000 tosses. Repeat the test with this data.

3. In the file `http://math.iisc.ernet.in/~manju/UGstatprob/twomidtermgrades.txt` you see the scores obtained in two exams by your batch in the first and second midterms, respectively. Test the hypothesis that the overall performance is worse in the second mid-term than in the first.

4. In `http://math.iisc.ernet.in/~manju/UGstatprob/heightweight2.txt` you will find data on heights (second column) and weights (third column) of 200 individuals.

1. Test the hypothesis that heights are normally distributed (this is the null hypothesis). Use χ^2-test with different choices of bins (i.e., do it with 10 bins and then with 15, etc. Each bin should have at least 5 observations).

2. Do the same for weights.

5. *Benford's law* is the probability distribution with mass function given by $f(k) = \log_{10}(k+1) - \log_{10}(k)$ for $k = 1, 2, \ldots, 9$. It is observed that for various quantities that vary over several orders of magnitude, the first digit follows Benford's law. Here we give a few. In each case, conduct a χ^2-test at level 0.10, with the null hypothesis being that the distribution is indeed Benford's law. Compute the p-value in each case.

1. Let $F_0, \ldots, F_{999}$ be the first 1000 Fibonacci numbers. This is a (non-random!) sequence of numbers defined by $F_0 = F_1 = 1$ and $F_k = F_{k-1} + F_{k-2}$ for $k \geq 2$. Although there is no randomness, extract the first digits, and compare against Benford's law by a χ^2-test.

2. Do the same for the sequence of factorials, $1, 2, 6, 24, \ldots$ (go up to 100 or wherever your computer stops to compute the first digit).

3. In `http://en.wikipedia.org/wiki/List_of_national_capitals_by_population` you will find the populations of the capitals of (almost) all countries in the world. For convenience, the list of populations is given in `http://math.iisc.ernet.in/~manju/UGstatprob/population.txt`. Again, compute the first digits and check the hypothesis that Benford's law applies. Report the p-value.