Course EPIB-621 - Data Analysis for the Health Sciences

Assignment 1

1. In a one-sided frequentist test of the null hypothesis $H_0 : \mu = 0$, versus $H_A : \mu > 0$, it is found that the p-value is p=0.0001. State whether each of the following statements are true or false, and explain why.

(a) After carrying out this experiment, there is only a one in ten thousand chance (0.0001) of being wrong if the conclusion is to reject the null hypothesis.

(b) Since the p-value is very small, the value of $\overline{x}$ observed in the experiment must have been far from zero.

2. The table below gives data on blood pressure before and after treatment for two groups of subjects participating in a clinical trial. One group took a daily calcium supplement, while the other group received a placebo.

Calcium Group		**Placebo Group**	
before	after	before	after
107	100	123	124
110	114	109	97
123	105	112	113
129	112	102	105
112	115	98	95
111	116	114	119
107	106	119	114
112	102	112	114
136	125	110	121
102	104	117	118
–	–	130	133

(a) Calculate a 95% confidence interval for the difference in blood pressure changes (before minus after) between the two groups. Give the interpretation of this confidence interval.

(b) Carry out a t-test of the null hypothesis that there is no difference in blood pressure changes between the two groups. State the null and alternative hypotheses, calculate the test statistic, and state your conclusion.

(c) You now must make a decision regarding whether or not to prescribe calcium supplementation to your patients with mild high blood pressure. In helping you to make this decision, would your answer to part (a) or (b) provide more useful information? Why?

3. The following data are observed in an experiment designed to compare a new treatment to a standard therapy:

	Therapy		
	New	Standard	
Success	60	30	90
Failure	10	40	50
	70	70	140

(a) Test the null hypothesis that there is no difference in success rates between the new and standard therapies. State the null and alternative hypotheses, and calculate a p-value using a χ^2 test. State your conclusion.

(b) Repeat part (a), but use a (two-sided) Fisher's Exact test instead.

(c) How do the p-values calculated by the two different procedures compare?

(d) Calculate a confidence interval for the difference in success rates between the new and standard therapies. Overall, what would you conclude? After knowing the confidence interval, do the p-values discussed in parts (a) or (b) add any clinically useful information? If so, what information do they add?

4. (a) Construct your normal prior distribution for the average age at which men get their first myocardial infarction (MI, i.e., a heart attack), among all men who do have MI's. There is no "correct" answer here, but you should justify your

choice; the mean and standard deviation from this distribution should correctly represent your prior knowledge about ages at which men have heart attacks.

(b) Update your prior to a posterior distribution using the data given below. Assume that the variance of the age at first MI is known to be 144 (so SD=12). First calculate this by hand, showing all of your calculations. Then, check your answer using the R program for Bayesian posterior means we discussed in class.

68 52 62 67 68 56 53 81 36 77

(c) Use R to plot your prior and posterior distributions on the same graph. Comment on the degree to which you think your prior distribution influences the posterior distribution.

(d) Find prior and posterior 95% credible intervals for the mean age. You can do this calculation either by hand, or using R.

5. Suppose that a person claims that they are a psychic, and you decide to test them by having them predict the outcome of a series of coin flips that you will carry out.

(a) Suppose you flip a coin five times, and they in fact correctly predict the outcome (heads or tails) all five times. What is your personal probability that they are in fact a psychic? [Note: there is no "correct" or "incorrect" answer here, I expect all students in the class to give different answers, but there is a correct way to derive your personal probability. Therefore, provide details about where your prior probability came from.]

(b) Same question as in part (a), but re-evaluate now if they get 20 flips in a row correctly.

(c) Now consider the following slightly different situation: A skeptics organization in the US offers a prize of $10 million to anyone who can prove they are a psychic, by correctly guessing the outcome of 20 coin flips in a row. Given the size of the prize, it is not surprising that the contest drew many entrants, all claiming to be psychics, and in fact, there were 1,000,000 such trials carried out on 1,000,000 different people, each claiming to be psychic. Out of these trials, one contestant actually indeed got all 20 flips correct. Do you believe their claim that they are psychic? Explain why or why not.