BST 140.652
Problem Set 5

Problem 1. A laboratory experiment found that in a random sample of 20 frog eggs having aquaporins, 17 exploded when put into water.

a. Plot and interpret the posteriors for p assuming a beta prior with parameters $(2, 2)$, $(1, 1)$ and $(.5, .5)$.

b. Calculate and interpret the credible interval for each of the beta prior parameter esttings. Note that the R package `binom` may be of use.

Problem 2. A study of blood alcohol levels (mg/100 ml) at post mortem examination from traffic accident victims involved taking one blood sample from the leg, A, and another from the heart, B. The results were:

Case	A	B	Case	A	B
1	44	44	11	265	277
2	265	269	12	27	39
3	250	256	13	68	84
4	153	154	14	230	228
5	88	83	15	180	187
6	180	185	16	149	155
7	35	36	17	286	290
8	494	502	18	72	80
9	249	249	19	39	50
10	204	208	20	272	271

Test whether or not the mean blood alcohol level differs between the heart and the leg. Give the appropriate null and alternative hypotheses. Give the relevant P-value. Interpret your results, state your assumptions.

Problem 3. Forced expiratory volume FEV is a standard measure of pulmonary function. We would expect that any reasonable measure of pulmonary function would reflect the fact that a person's pulmonary function declines with age after age 20. Suppose we test this hypothesis by looking at 10 nonsmoking males ages 35-39, heights 68-72 inches and measure their FEV initially and then once again 2 years later. We obtain this data.

Person	Year 0 FEV (L)	Year 2 FEV (L)	Person	Year 0 FEV (L)	Year 2 FEV (L)
1	3.22	2.95	6	3.25	3.20
2	4.06	3.75	7	4.20	3.90
3	3.85	4.00	8	3.05	2.76
4	3.50	3.42	9	2.86	2.75
5	2.80	2.77	10	3.50	3.32

a. Preform and interpret the relevant test. Give the appropriate null and alternative hypotheses. Intepret your results, state your assumptions and give a P-value.

b. A large test comparing the two-year decline in non-smokers of a different age. Perform a sample size calculation to detect a change in FEV over two years at least as large as that detected for males age 35-39. Use the data above a for any relevant constants that you might need.

Problem 4. Another aspect of the preceding study involves looking at the effect of smoking on baseline pulmonary function and on change in pulmonary function over time. We must be careful since FEV depends on many factors, particularly age and height. Suppose we have a comparable group of 15 men in the same age and height group who are smokers and we measure their FEV at year 0. The data are given (For purposes of this exercise assume equal variance where appropriate).

Person	FEV Year 0 (L)	FEV Year 2 (L)	Person	FEV Year 0 (L)	FEV Year 2 (L)
1	2.85	2.88	9	2.76	3.02
2	3.32	3.40	10	3.00	3.08
3	3.01	3.02	11	3.26	3.00
4	2.95	2.84	12	2.84	3.40
5	2.78	2.75	13	2.50	2.59
6	2.86	3.20	14	3.59	3.29
7	2.78	2.96	15	3.30	3.32
8	2.90	2.74			

Assume that A and B are normally distributed with equal variance, and let μ A and μ B denote the means of this distribution. We want to test,
H0 : μ A = μ B
HA : μ A 6= μ B
Using the t.test() function in R, we get a p-value of 0.0407. With $\alpha = 0.05$, we reject the null and conclude that the blood alcohol level may be different between samples from the leg and samples from the heart.

```
A <- c(44,265,250,153, 88,180, 35,494,249,204,
265, 27, 68,230,180,149,286, 72, 39,272)
B <- c(44,269,256,154, 83,185, 36,502,249,208,
277, 39, 84,228,187,155,290, 80, 50,271)
diff = A - B
# T-test
t.test(diff, var.equal = TRUE)
```

Test the hypothesis that the change in FEV is equivalent between non-smokers and smokers. State relevant assumptions and interpret your result. Give the relevant P-value.

Problem 5. Perform the following simulation. Randomly simulate $1,000$ sample means of size 16 from a normal distribution with means 5 and variances 1. Calculate $1,000$ test statistics for a test of $H_0 : \mu = 5$ versus $H_a : \mu < 5$. Using these test statistics calculate $1,000$ P-values for this test. Plot a histogram of the P-values. Note, this exercise demonstrates the interesting fact that the distribution of P-values is uniform.

Problem 6. Suppose that systolic blood pressures were taken on 16 oral contraceptive users and 16 controls at baseline and again then two years later. The average difference from follow-up SBP to the baseline (followup - baseline) was $11\ mmHg$ for oral contraceptive users and $4\ mmHg$ for controls. The corresponding standard deviations of the differences was $20\ mmHg$ for OC users and $28\ mmHg$ for controls.

a. Calculate and interpret a 95% confidence interval for the **relative** change in systolic blood pressure for oral contraceptive users; assume normality.

b. Does the change in SBP over the two year period appear to differ between oral contraceptive users and controls? Perform the relevant hypothesis test and interpret. Give a P-value. Assume normality and a common variance.

Problem 7. Will a Student's T or Z hypothesis test for a mean with the data recorded in pounds always agree with the same test conducted on the same data recorded in kilograms? (explain)

Problem 8. A researcher consulting you is very concerned about falsely rejecting her null hypothesis. As a result the researcher decides to increase the sample size of her study. Would you have anything to say? (explain)

Problem 9. Researchers studying brain volume found that in a random sample of 16 sixty five year old subjects with Alzheimer's disease, the average loss in grey matter volume as a person aged four years was .1 mm^3 with a standard deviation of .04 mm^3.

a. Calculate and interpret a P-value for the hypothesis that there is no loss in grey matter volumes as people age. Show your work.

b. The researchers would now like to plan a similar study in 100 healthy adults to detect a four year mean loss of .01 mm^3. Motivate a general formula for power calculations in this setting and calculate the power for a test with $\alpha = .05$? Assume that the variation in grey matter loss will be similar to that estimated in the Alzheimer's study.

Problem 10. A recent Daily Planet article reported on a study of a two week weight loss program. The study reported a 95% confidence interval for weight loss from baseline of [2 lbs, 6 lbs]. (There was no control group, all subjects were on the weight loss program.) The exact sample size was not given, though it was known to be over 200.

a. What can be said of a $\alpha = 5\%$ hypothesis test of whether or not there was any weight change from baseline? Can you determine the result of a $\alpha = 10\%$ test without any additional calculation or information? (explain your answer)

Problem 11. Suppose that 18 obese subjects were randomized, 9 each, to a new diet pill and a placebo. Subjects' body mass indices (BMIs) were measured at a baseline and again after having received the treatment or placebo for four weeks. The average difference from follow-up to the baseline (followup - baseline) was -3 kg/m^2 for the treated group and 1 kg/m^2 for the placebo group. The corresponding standard deviations of the differences was 1.5 kg/m^2 for the treatment group and 1.8 kg/m^2 for the placebo group. Does the change in BMI over the two year period appear to differ between the treated and placebo groups? Perform the relevant test and interpret. Give a P-value. Assume normality and a common variance.