

Biostatistics

Math 322 — Spring 2009

Final Exam

The exam is due Wednesday May 6th, at noon in Cupples I, room 100.

This exam contains twenty-five problems numbered 1, 2a, 2b, 3a, 3b, 4a, 4b, 5a, 5b, 5c, 6, 7a, 7b, 7c, 7d, 7e, 7f, 8a, 8b, 8c, 9a, 9b, 9c, 9d, and 9e, worth a total of 100 points. Each problem is worth 4 points.

You may use the textbook, your class notes, a calculator or computer, and any other previously written reference, but you may not receive assistance from any other person.

- Explicitly state any assumptions (e.g. on independence, distributions, etc.) you make.
- If you use R (or some other computer program) to solve a problem, supply the commands and the output.
- Whenever you are performing a hypothesis test, you should report the following: Statement of the null hypothesis and alternative hypothesis with an explanation of the parameters used. The significance level. The value and distribution of the test statistic. The critical value. The p -value. The conclusion of the test.

Problem 1

Given the 2×2 contingency table

	A	B
I	63	17
II	49	17

Analyze the table using both the χ^2 -test and the exact Fisher test.

Problem 2

A farmer wishes to compare two diets, to figure out which diet fattens his pigs up best. He has 10 pigs with which to compare the two diets, so 5 pigs are randomly assigned to each. He then weighs the pigs at the start of the experiment, and again 30 weeks later.

Pig	Diet	Weight at baseline	Weight after 30 weeks
1	A	92.4	241.3
2	A	102.7	208.0
3	A	92.7	222.2
4	A	84.2	191.8
5	A	107.3	230.4
6	B	91.1	205.1
7	B	79.8	176.4
8	B	83.6	173.8
9	B	71.1	158.5
10	B	98.4	226.2

Assume that the weights are normally distributed, and that the variance is unaffected by the diet.

- Independent of diet, is the average weight increase more than 100 pounds over the 30 weeks?
- Is there a difference between the two diets?

Problem 3

The file

http://www.math.wustl.edu/~hjelle/m322/r/m322_exam090506table3.txt

contains data from a study investigating whether the fore- and hind-legs of deer are the same lengths.

- Rank the data and set up a table of ranks similar to Table 9.1, page 367 in the book.
- Carry out a Wilcoxon test to see if there is a difference between the lengths of the fore- and hind-legs.

Problem 4

A survey of statisticians in the US finds that 19 out of 163 sampled are current cigarette smokers.

- a) Assuming that 25% of the general population are current smokers, test whether statisticians are representative of the general population regarding cigarette smoking.

The survey divided the statisticians into three groups based on their current employment.

	Academic	Finance	Other
Nonsmoker	42	60	42
Smoker	7	8	4

- b) Test whether the proportion of current smokers is the same across all three groups.

Problem 5

At Sacred Heart hospital, each patient with flu-like symptoms is diagnosed for flu by two different doctors. The results from last week were

	Dr. Dorian	
Dr. Reid	+	-
+	61	9
-	14	38

- a) Analyze the data using McNemar's test.
b) Assess the reproducibility of the diagnoses in terms of the Kappa statistic.
c) Find a 95% confidence interval for κ .

Problem 6

The hair colors of 200 Norwegian tourists visiting the Gateway Arch are recorded.

Blonde	Black	Brown	Red
96	47	40	17

Use a χ^2 -goodness-of-fit test to investigate

H_0 : the sample comes from a population having a 6: 3: 2: 1 ratio of blonde to black to brown to red hair

vs.

H_1 : the sample comes from a population *not* having that ratio.

Problem 7

Heights of biostatistics students are measured. Among 41 male students, the average height is $\bar{x}_M = 68.7$ inches, with sample standard deviation $s_M = 3.9$. Assume that the heights are normally distributed, and use significance level $\alpha = 0.05$.

- Based on this sample, are the heights of male biostatistics students different from the national average of $\mu_M = 69.2$ inches?
- If the true average height of male biostatistics students is 68.1 inches, what is the power of the test?
- How many students need to be measured to obtain a power of 80%?

The heights of 47 female students are also measured. Their average height is $\bar{x}_F = 65.3$ inches, with $s_F = 2.6$.

- Construct a 95% confidence interval for the true average height of female biostatistics students.
- Test the hypothesis

$$H_0: \sigma_M^2 = \sigma_F^2 \quad \text{vs.} \quad H_1: \sigma_M^2 \neq \sigma_F^2$$

at significance level $\alpha = 0.10$.

- Are the male students significantly taller than the female students?

Problem 8

Consider the data set

http://www.math.wustl.edu/~hjelle/m322/r/m322_exam090506table8.txt

We will try to predict the weight of people based on the lengths of their legs. Assume the multiple linear regression model where **Weight** is explained by **RightLeg** and **LeftLeg**.

- Do an F -test to find if the model is significant.
- Carry out t -tests to see which, if any, of the two variables are significant.
- Explain your findings in a) and b). Are they consistent? Why, or why not?

Problem 9

Consider the data set

http://www.math.wustl.edu/~hjelle/m322/r/m322_exam090506table9.txt

For 50 trees of the same species from a ten hectar area of mixed woodland, their diameter at breast height in millimeters, and the numbers of flowers on the tree at the time of measurement is recorded.

- a) Find the best fitting straight line predicting **Flowers** from **Diameter**.
- b) Is there significant correlation between **Flowers** and **Diameter**?
- c) Test the null hypothesis that the slope of the best fitting line equals 4.
- d) Predict the number of flowers on a tree with diameter 123 mm.
- e) Give a 95% prediction interval for your prediction in d).