# 2009 Honors Examination in Statistics

May 2009

Department of Mathematics and Statistics
Swarthmore College

Name: _____

**Instructions:** The examination consists of six questions. Number the questions clearly in your work and start each question on a **new** page. You must make it clear how you arrived at your answer. Answers without any work may lose credit even if they are correct.

This is a closed-book three-hour exam. You may not refer to notes or textbooks.

You may use a calculator that does not do algebra or calculus.

Normal, $t$ and F tables should be supplied with this exam.

1. Sixteen people volunteered to be part of an experiment. All sixteen people were Caucasian, between the ages of 25 and 35, and were supplied with nice clothes. Eight of them were male and eight were female. The question of interest in this experiment was whether females receive faster service at restaurants than males. Each of the eight male participants was randomly assigned a restaurant, and each of the eight females was randomly assigned to one of these same eight restaurants. One Friday night, all sixteen people went out to eat, each one alone. The male and female assigned to the same restaurant would arrive within 5 minutes of each other, with the order determined by flipping a coin (male first or female first). Each person then ordered a similar drink and a similar meal. The time (in minutes) until the food arrived at the table was recorded. They are shown below:

| *Restaurant* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| *Male* | 22 | 14 | 16 | 26 | 18 | 13 | 9 | 27 |
| *Female* | 25 | 12 | 13 | 21 | 21 | 14 | 9 | 16 |

Assume the time until food arrived followed a normal distribution.

(a) Construct a 95% confidence interval for the mean difference of the service time between females and males.

(b) Do you think the data provide strong evidence that females receive faster service at restaurants than males? State the null and alternative hypotheses, calculate test statistic and p-value, then make conclusion in both statistical language and plain English. Test at 5% level of significance.

(c) Can you use the confidence interval you constructed in (a) to answer the question in (b)? Why?

(d) Suppose it's known that the standard deviation of the difference of the service time is 5 minutes. Estimate the power of the test if on average, service to males is 4 minutes long than to females.

(e) Suppose it's known that the standard deviation of the difference of the service time is 5 minutes. If you want the 95% confidence intervals for the mean difference of the service time between females and males narrower than 4 minutes. Estimate how many volunteers you would need.

2. A box contains many coins that are either balanced or loaded. When tossed, a loaded coin will have probability $p \in (0.5, 1.0)$ landing on its head and a balanced coin will have probability 0.5 landing on its head. The proportion of balanced coin in the box is $\theta$, $\theta \in (0, 1)$. A coin is randomly selected from the box and independently tossed $n$ times. Let $X$ be the number of heads in these $n$ tosses. Given $p$ and $\theta$,

   (a) Find the probability mass function for $X$.

   (b) Find the expected value of $X$.

   (c) Find the variance of $X$.

   (d) Suppose it's known that $p = 0.9$. A student randomly chose a coin from the box and tossed it 20 times, getting 14 heads. Provide an estimate of $\theta$, the proportion of balanced coins in the box.

3. Three players simultaneous toss coins. The coin tossed by A (B) [C] turns up heads with probability $p_1$ ($p_2$) [$p_3$].

   (a) If all three players toss their coins just once. What's the probability that they all got the same outcome?

   (b) If all three players toss their coins just once. What's the probability that player A got an outcome different from those of the other two?

   (c) Now suppose the game is the following: If one person gets an outcome different from those of the other two, then he is the odd man out. If there is no odd man out, the players flip again and continue to do so until they get an odd man out. What is the probability that A will be the odd man out on the second round of flips?

   (d) In the same game as in (c), what is the probability that A will be the odd man out?

4. A complaint of sex discrimination has been filed against a division of a company that employs 31 (16 males) individuals in a particular job where all employees have a similar amount of education and prior work experience. Three variable are considered: The response variable is the monthly salary, in dollars; the Seniority is measured by months and Female is an indicator (dummy) variable with value =1 if the employee is a female.

   (a) The women submit the following regression analysis that they claim shows sex discrimination since they have adjusted for any difference in seniority at the company.

| | *Estimate* | *Std.Error* | *t value* | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| *Intercept* | 987.77650 | 40.69517 | 24.27 | $< .0001$ |
| *Seniority* | 5.04134 | 0.40725 | 12.38 | $< .0001$ |
| *Female* | −152.10662 | 34.59094 | ? | ? |

   Find the values of the two question marks. What are the estimated regression equations for male and females, respectively? Give the test the plaintiffs (women) used to support their claim of unequal pay. State the null and alternative hypotheses and describe the test statistic and p-value.

   (b) The company claims that the proper regression should allow for an interaction between gender and seniority and submits the following alternative regression

| | *Estimate* | *Std.Error* | *t value* | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| *Intercept* | 913.89714 | 52.12293 | 17.53 | $< .0001$ |
| *Seniority* | 5.95343 | 0.58043 | 10.26 | $< .0001$ |
| *Female* | −28.79427 | 67.24365 | −0.43 | 0.6719 |
| *Seniority* : *Female* | −1.62575 | 0.77492 | −2.10 | 0.0454 |

   The company argues that in their model the Female coefficient is not significant even at the .1 level. What part of the output of their regression are they referring to? Be sure to state the null and alternative hypotheses and carry out the statistical test the company did. Do you agree with the company? According to the company's analysis, what are the estimated regression equations for male and females, respectively?

   (c) The court appoints an expert who questions the analysis of the company and proposes a more appropriate procedure to test whether "sex" affects salary from the model submitted by the company. What procedure did the expert decide on and carry out the calculation? Again, formulate the correct null and alternative hypotheses and carry out the test. Does the proper test (at the level of .05) support the plaintiffs or the company?

5. Let $X_1, X_2, \ldots X_n$ be independent, identically distributed random variables with the density function:

$$f_X(x) = \frac{1}{\theta},\ 0 \leq x \leq \theta.$$

(a) Find the mean and standard deviation of $X_1$;

(b) Write out the Likelihood function $L(\theta)$ for the entire sample $X_1, X_2, \ldots X_n$;

(c) Find the Maximum Likelihood Estimate (MLE) $\hat{\theta}$ for the entire sample $X_1, X_2, \ldots X_n$;

(d) Find the distribution function and density function of the MLE $\hat{\theta}$;

(e) Is the MLE unbiased? Consistent?

(f) According to the distribution of the $\hat{\theta}$, construct a $100(1-\alpha)\%$ confidence interval for $\theta$.

6. Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with success probability $p$. Let $Y = X_1 + X_2 + \ldots + X_n$.

   (a) What kind of random variable is $Y$? Find the probability mass function for $Y$.

   (b) Is $Y$ a sufficient statistic for $p$?

   (c) Assume that $p$ has a noninformative prior distribution: Uniform on $[0, 1]$. Find the joint distribution of $Y$ and $p$.

   (d) Find the posterior distribution for $p|Y$.

   (e) Can you suggest a Bayes estimate for $p$?

   (f) Based on the posterior distribution, suggest a Bayes confidence interval for $p$.