

2007 Honors Examination in Statistics

**Department of Mathematics and Statistics
Swarthmore College**

Name: _____

Instructions: This 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 examination. You may not refer to notes or textbooks.

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

Normal, t , and χ^2 tables should be supplied with this exam.

1. Let Y have a Gamma(α, β) distribution with density given by

$$f_Y(y|\alpha, \beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-y/\beta} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

where α and β are positive. Let μ ($\mu > 0$) be the mean of Y and σ^2 be the variance of Y .

- (a) Derive the mean of Y .
- (b) Let Y_1, Y_2, \dots be independent and identically distributed as in part (a). Let $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$. The variance of Y is $\alpha\beta^2$. What are the mean and variance of \bar{Y}_n ? Explain why.
- (c) What is the probability density function (pdf) of $W = 1/Y$?

2. Let $Y_i, i = 1, \dots, n$ be independent and identically distributed Gamma($2, \beta$) random variables.

- (a) What is the likelihood function for β ?
- (b) Find the maximum likelihood estimator (MLE) of β .
- (c) What is the large sample distribution of the MLE from part (b)?

Suppose β has prior distribution specified as an Inverse-Gamma(μ, ν) distribution. The probability density function of β is

$$f(\beta) = \frac{1}{\Gamma(\mu)\mu^\nu} \frac{1}{\beta^{\mu+1}} e^{-1/(\beta\nu)}$$

for $\beta, \mu, \nu > 0$.

- (d) What is joint density of the data and β ?
- (e) What is the posterior distribution of β given observations $y_i, i = 1, \dots, n$?
- (f) Show that as n becomes large ($n \rightarrow \infty$) that the posterior mean of β is determined by the MLE.

3. Let the joint probability density function (pdf) for random variables X and Y be as follows:

$$f(x, y) = kx^{1/2}y^{3/2}$$

for $0 < x < y < 2$ and $f(x, y) = 0$ otherwise.

- (a) Find k .
- (b) Are X and Y independent? Why or why not?
- (c) What is the probability density function for X given that $y = 3/2$?
- (d) What is the probability that both X and Y are less than 1?

4. Two sets of test results from students in a Geological and Atmospheric Sciences course are reported. Three hundred twelve students are reported in the first set of tests. The average score (on a ten-point scale) was 7.11. The standard deviation was 1.90. Two hundred ninety-nine students are reported in the second set of tests. Their average and standard deviation were 7.36 and 1.89, respectively.
- (a) Given the information above, propose a test of whether or not the mean scores in the two sets of tests are the same. State hypotheses and any assumptions. Conduct the test and report a conclusion.
 - (b) It turns out (you learn later) that both tests were given in one class and they were pre and post tests. Of the 291 students who took both tests, the average of the difference between their pre- and post-test scores was 0.25. The standard deviation of the gains (post-pre) was 2.04. Given this new information, propose a test of whether or not the mean scores in the two sets of tests are the same. State hypotheses and any assumptions. Conduct the test and report a conclusion.
 - (c) Compare the results of the tests in the two parts. Comment on the effect of correlation between pre- and post-test scores.

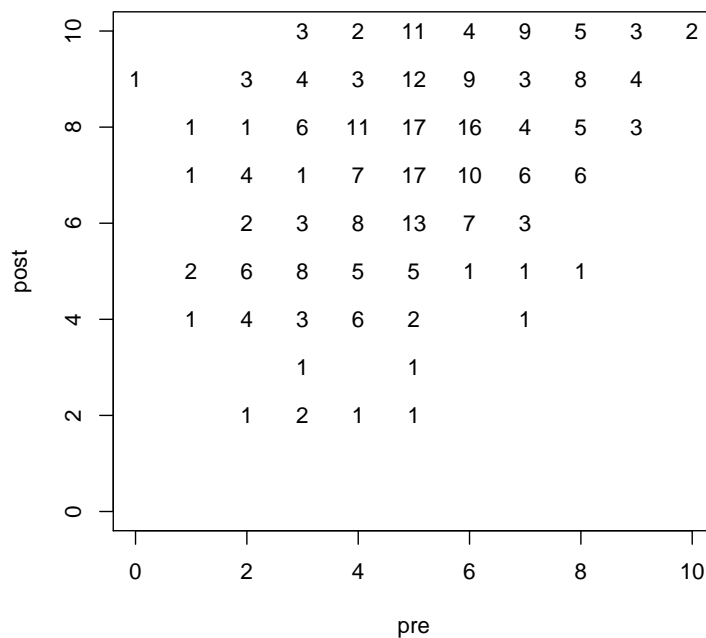
5. Two hundred ninety-one students in a Geological and Atmospheric Sciences course took a pre-test and a post-test. Their scores are plotted below in the figure. A simple linear regression model is fit to the data. Some computer output are given below.

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 2.00301 | 0.40679 | 4.924 | 1.43e-06 |
| x | 0.42414 | 0.05373 | 7.894 | 6.10e-14 |

| | Df | Sum Sq | Mean Sq | F value |
|-----------|-----|--------|---------|---------|
| x | ? | 185.78 | ? | ? |
| Residuals | 289 | 861.48 | ? | |

Figure 1: Plot of pre-test and post-test exam scores for 291 students: numbers indicate number of students with each pair of scores.



- Draw the estimated regression line on the graph. Interpret the meaning of the estimated slope. What do you predict on the post-test for a student with a pre-test score of 6?
- Compute the root mean squared error (RMSE) for the regression. Illustrate the RMSE on the graph.
- Compute the F statistic. Explain what a 'significant regression' means and determine whether or not this regression is significant.

6. Test data were collected from students in three schools. In school 1, teachers were given a short training course involving practice teaching sessions (training A) using computer technology in the classroom. Teachers in school 2 received training but without the practice sessions (training B). Teachers in school 3 are not scheduled to receive the training until next year. Students in all schools were given standardized grade-appropriate tests of basic mathematics skills. Below are the number of students in each school that scored at or above the standard and the numbers that scored below the standard on the exam.

| Classroom: | Exam results | |
|------------|----------------------|----------------|
| | At or above standard | Below standard |
| School 1 | 192 | 15 |
| School 2 | 290 | 21 |
| School 3 | 57 | 11 |

- (a) Produce a 95% confidence interval for the difference between the percent of students at or above grade level in the school with training A and the school with training B. Interpret the results.
- (b) Researchers want to test whether there is a difference in proportions of the children scoring at or above the standard for the three schools.
- Compute the proportion at or above the standard separately for the three schools. Does there appear to be any difference?
 - Assuming the chance that students score at or above the standard is really equivalent in the three schools, compute the proportion at or above the standard overall. Then, under the assumption of equivalence, compute the number of students in each school you would expect to be at or above standard and the number you would expect below the standard.
 - The statistic used to test the hypothesis is $X^2 = \sum_{\text{rows}} \sum_{\text{columns}} (O_{rc} - E_{rc})^2 / E_{rc}$, where O_{rc} is the observed count in row r and column c in the table above and E_{rc} is the expected value from the previous part. Compute this statistic.
 - State a conclusion about the researchers interest. If you know the appropriate distribution to use for a statistical test, then use it. If not, then compare the test statistic value to a distribution with mean 2 and variance 4.
- (c) Do you have any concern about the tests performed in the previous parts? Hint: classrooms have 25-30 students in them.