2010 Swarthmore Honors Examination in Statistics

Weiwen Miao
Haverford College

May 2010

Instructions: The examination consists of seven questions. Number the questions clearly in your work. 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. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Gamma density

$$f(x|\alpha,\theta) = \begin{cases} \left(\frac{1}{\Gamma(\alpha)\theta^\alpha}\right) x^{\alpha-1} e^{-x/\theta}, & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

   where $\alpha > 0$ is **known** and $\Gamma(\alpha)$ is the Gamma function with $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$. Also note that $\theta > 0$.

   (a) Find the MLE $\hat{\theta}$ of θ.

   (b) Is $\hat{\theta}$ unbiased?

   (c) Suggest a sufficient statistic for θ.

   (d) Find the Fisher information $I(\theta)$ about the parameter θ.

2. In New Haven's fire department, promotions to Lieutenant and Captain positions are based on scores of the promotion exams. Under the contract between the City of New Haven and the firefighter's union, applicants with a total score of 70% or above pass the exam and become eligible for promotion. In 2003, the New Haven Fire Department administered the promotion exams, but decided not to validate the test results, because an insufficient number of minorities would be promoted to an existing position. Ricci and other 17 firefighters who achieved high scores on promotion examinations sued the

Table 1: Summary Statistics for *Ricci v. DeStefano*, Captain exam

Group	Sample size	Mean	S.D.	# of Pass
Majority	25	74.11	8.25	16
Minority	16	66.16	8.66	6

city for "reverse discrimination". The following table provides summary statistics for the test scores of the Captain exam.

Suppose the distributions of test scores for majority and minority are normal.

(a) For the Captain exam, do you think the average score for majority is significantly different than that of the minority? ($\alpha = 5\%$)

(b) Construct a 95% confidence interval for the true average difference between the majority and the minority test scores.

(c) For the Captain exam, do you think the pass rate for majority and minority are significantly different? ($\alpha = 5\%$)

(d) Construct a 95% confidence interval between the two pass rates.

(e) Suppose it's known that the standard deviations of the test scores for majority and minority are equal. The common standard deviation is 8.5 points. Estimate the power of the test if the average test score for majority is 10 points higher than that of the minority.

(f) Only for the sake of the exam, suppose the courts wanted the power to be 99% when the average test score for majority is 10 points higher than that of the minority. Given that 25 whites and 16 minorities took the exam, estimate the level of significance of the test.

3. (Continue from Problem 2) In the actual law case, the courts considered three races: black, Hispanic and white. For the Captain exam, 3 out of 8 blacks, 3 out of 8 Hispanics, and 16 out of 25 whites passed the exam.

   (a) Do you think the pass rate for the three races are significantly different? ($\alpha = 5\%$).

   (b) The following is the one-way ANOVA table for the Captain exam:
   Using the available information, find the values of the question marks "?".

   (c) Do you think the average test scores are significantly different for the three races?

Table 2: One-way ANOVA output for Captain test scores

	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Race	?	707.19	?	?	?
Residuals	?	2669.15			

4. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, 1)$. Assume that the prior distribution for μ is also normal with mean 0 and variance 1. i.e. $\mu \sim N(0,1)$. To get the Bayesian estimator for μ, you start with the sample mean $\overline{X}$.

   (a) What's the distribution of $\overline{X}$, given the mean μ? Write out the density function of $\overline{X}|\mu$, i.e. $f_{\overline{X}|\mu}(\overline{x}|\mu)$.

   (b) Find the joint density of $f_{(\overline{X},\mu)}(\overline{x}, \mu)$.

   (c) Find $f_{\overline{X}}(\overline{x})$, the marginal density of $\overline{X}$.

   (d) Find the posterior distribution of μ, the $f_{\mu|\overline{X}}(\mu|\overline{X})$. What kind of distribution is this $\mu|\overline{X}$?

   (e) Find the Bayesian estimator of the μ if the square error loss is used.

   (f) Find the Bayesian estimator of the μ if the absolute error loss is used.

5. In a study of the relation of amount of body fat to several possible predictor variables, a sample of 20 healthy females 25-34 years of age were selected and their body fat (Y), triceps skinfold thickness (X_1) and midarm circumference (X_2) were measured. The amount of body fat was obtained by cumbersome and expensive procedure requiring the immersion of the person in water. It would therefore be very helpful if a regression model with some or all of these predictor variables would provide reliable estimate of the amount of body fat since the measurements needed for the predictor variables are easy to obtain.

   Assume that the usual linear model assumptions are met.

   (a) The table below is the regression output for body fat (Y) to triceps skinfold thickness X_1. What's the least squares regression line? Do you think triceps skinfold thickness (X_1) is useful to predict body fat? Is it a good predictor to the body fat?

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	-1.4961	3.3192	-0.451	0.658
X_1	0.8572	0.1288	6.656	3.02e-06

Residual standard error: 2.82 on 18 degrees of freedom

Multiple R-squared: 0.7111, Adjusted R-squared: 0.695

F-statistic: 44.3 on 1 and 18 DF, p-value: 3.024e-06

(b) The table below is the regression output for body fat (Y) to midarm circumference X_2. What's the least squares regression line? Do you think midarm circumference (X_2) is useful to predict body fat? It is a good predictor to the body fat?

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	-14.6868	9.0959	1.615	0.124
X_2	0.1994	0.3266	0.611	0.549

Residual standard error: 5.193 on 18 degrees of freedom

Multiple R-squared: 0.02029, Adjusted R-squared: -0.03414

F-statistic: 0.3728 on 1 and 18 DF, p-value: 0.5491

(c) Finally the table below gives the regression output with both X_1 and X_2 in the model. What's the regression equation? Do you think midarm circumference (X_2) is useful to predict body fat? Interpret the slope (coefficient) of X_2.

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	6.7916	4.4883	1.513	0.1486
X_1	1.0006	0.1282	7.8003	5.12e-17
X_2	-0.4314	0.1766	-2.443	0.0258

Residual standard error: 2.496 on 17 degrees of freedom

Multiple R-squared: 0.7862, Adjusted R-squared: 0.761

F-statistic: 31.25 on 2 and 17 DF, p-value: 2.022e-06

(d) Without further information, which one is the best model?

(e) Notice that the signs for the coefficient of X_2 change from positive in (b) to negative in (c). Is that possible? How do you interpret this change?

6. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. Prove that if the $100(1-\alpha)\%$ confidence interval for σ^2 contains the constant σ_0^2, then the conclusion for the hypothesis testing: $H_0 : \sigma^2 = \sigma_0^2$ v. $H_a : \sigma^2 \neq \sigma_0^2$ at α level of significance must be "Do not reject null hypothesis".

7. Let X and Y be independent standard normal random variables. Define $Z = min(X, Y)$. Prove that $Z^2 \sim \chi_1^2$. (Hint: The density for χ_1^2 is: $f(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} e^{-\frac{1}{2}x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$.)