McGill University December 2011
Faculty of Science Final examination

Statistics

Math 324

Thursday, December 22nd, 2011

Time: 9am-12pm

Examiner: Prof. Russell Steele Associate Examiner: Prof. J. Nešlehová

INSTRUCTIONS

1. The questions have to be answered in the exam booklets provided
2. The total possible number of points for the exam is 160.
3. This is a closed book exam. One 8 1/2" $\times$ 11" double sided crib sheet is allowed.
4. Calculators (both programmable and non-programmable) are permitted.
5. Use of a regular dictionary is permitted.
6. Use of a translation dictionary is permitted.

This exam comprises the cover page, three pages of questions, numbered 1 to 10, and four pages of statistical tables.

1. (10 pts) A study of male navy enlisted personnel was conducted. It was found that 90 of 231 left-handers had been hospitalized for injuries whereas 623 of 2148 right-handers had been hospitalized. Assume that the two groups were randomly sampled from the population of enlisted personnel.

   (a) (6 pts) Is there evidence that the probability of injury for right-handers is less than that of left-handers? Test your hypothesis at a significance level of 0.01.

   (b) (4 pts) Construct a 95% confidence interval for the difference in population proportions.

2. (10 points) Let $Y_1, .., Y_n$ be independent and identically distributed Gamma random variables with parameters α and β. Let $\bar{Y}$ be the sample mean of the Gamma random variables.

   (a) (6 points) Derive the exact distribution of $\bar{Y}$.

   (b) (4 points) What is the approximate distribution of $\bar{Y}$ for large n? Be sure to clearly state the values of the parameters for the approximate distribution.

3. (10 pts) An experimenter wanted to compare the compression strength for a sample of 12-oz aluminum cans filled with strawberry drink and another sample of cans filled with cola. The data are shown in the table below:

Beverage	n	Sample mean	Sample standard deviation
Strawberry drink	15	540	21
Cola	15	554	15

   (a) (4 pts) Do the data suggest that cola cans have a higher average compression strength than strawberry soda cans? Test your hypothesis with a Type I error rate of $\alpha = 0.01$.

   (b) (3 pts) Calculate and **interpret** the p-value for the test in part (a).

   (c) (3 pts) What assumptions are necessary for the validity of your test and p-value?

4. (20 pts) A transportation engineer conducted a study which examined the relationship between available street travel space (x) (in feet) and the separation distance (y), in feet, between a bike and a passing car as determined by photography. The data for ten randomly sampled streets with bike lanes can be summarized as follows:

$$\sum_{i=1}^{10} x_i = 154.20; \quad \sum_{i=1}^{10} y_i = 80;$$
$$\sum_{i=1}^{10} x_i^2 = 2452.17; \quad \sum_{i=1}^{10} x_i y_i = 1282.74;$$
$$\sum_{i=1}^{10} y_i^2 = 675.16;$$

   (a) (5 pts) Derive the equation of the estimated regression line.

   (b) (5 pts) Give a 90% confidence interval for the slope parameter and interpret your interval.

   (c) (5 pts) Find a 95% prediction interval for the separation distance on a street that has 15.0 feet as its available travel space value.

   (d) (5 pts) What would be the estimate of the *expected value* for all streets having available travel space of 15.0 feet? Give a 95% confidence interval for the expected value.

5. (15 pts) Let N_{ij} be the count in the (i,j)-th cell and $\hat{E}_{ij}$ be the estimator of the expectation of the (i,j)-th cell under a hypothesis of independence of the rows and columns in an $I \times J$ contingency table. Let n be the total number of subjects, i.e. $n = \sum_{i=1}^{I}\sum_{j=1}^{J} N_{ij}$.

   (a) (8 pts) Show that the χ^2 statistic for the test of independence (denoted by X^2) can be written in the form:

$$X^2 = \left(\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{N_{ij}^2}{\hat{E}_{ij}}\right) - n.$$

   (b) (7 pts) Perform a test of independence (at significance level $\alpha = 0.01$) for the following 2×3 table, using your result in part (a):

	X_2			
X_1	A	B	C	Total
Y	13	19	28	60
N	7	11	22	40
Total	20	30	50	100

6. (15 pts) Three types (labelled A, B, and C) of soil preparation were each randomly installed in plots at four different locations (labelled 1, 2, 3 and 4), i.e. each type of preparation was installed at each of the four locations. The researcher measured the growth of seedlings planted in each of the 12 plots and constructed the ANOVA table (although some of the cells are missing):

	Df	Sum Sq	Mean Sq	F value	p-value
Soil Prep		48.667			
Location		51.333			
Residuals					
Total	11	108.667			

   (a) (5 pts) Write down the ANOVA table above in your exam booklet, correctly filling in the missing cells.

   (b) (5 pts) Is there evidence to indicate mean differences in growth between the cell preparations at a significance level of $\alpha = 0.01$?

   (c) (5 pts) Is there evidence to indicate mean differences in growth between the locations at a significance level of $\alpha = 0.05$?

7. (10 points) Let X_1 and X_2 be random variables that have the following joint pdf:

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} \frac{4}{5}x_1x_2, & \frac{x_1}{3} < x_2 < \frac{x_1}{2}, 0 < x_1 < 1 \\ 0, & \text{otherwise} \end{cases}$$

   (a) (7 pts) Find the joint distribution of $Y_1 = X_2/X_1$ and $Y_2 = X_2$.

   (b) (3 pts) Find the marginal distribution of Y_1.

8. (15 pts) Let $Y_1, .., Y_n$ be independently and identically distributed according to a Normal distribution with unknown mean μ and unknown variance σ^2.

   (a) (5 pts) Show that T is a pivotal quantity for μ, where
   $$T = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$$
   where $\bar{Y}$ is the sample mean and S^2 is the sample variance.

   (b) (5 pts) Assume that in a sample of 100 observations from a Normal distribution, we observe a sample mean $\bar{y} = 11$ and a sample standard deviation $s = 3$. Construct a 90% confidence interval for μ.

   (c) (5 pts) Assume that we want to test a null hypothesis $H_0 : \mu = 10$ vs. an alternative hypothesis $H_a : \mu \neq 10$, again with a sample size of 100 observations, but now assuming a known population standard deviation of 3.0. What is the power of a hypothesis test of size $\alpha = 0.1$ based on T when the true $\mu = 10.5$?

9. (30 pts) Let $X_1, X_2, .., X_n$ be a random sample from the Rayleigh distribution which has density function:
   $$f(x; \theta) = \frac{x}{\theta} \exp(-\frac{x^2}{2\theta})$$
   for $x > 0$ and is otherwise 0. The $E(X_i) = \sqrt{\frac{\theta\pi}{2}}$ and $Var(X_i) = \left(\frac{4-\pi}{2}\right)\theta$.

   (a) (8 pts) Construct a method of moments estimator for θ using $\bar{X}$ and calculate the bias of your estimator. *Hint:* Use the fact that $Var(Y) = E(Y^2) - E(Y)^2$.

   (b) (4 pts) Is your estimator in part (a) a consistent estimator of θ? Why or why not?

   (c) (6 pts) Suggest a different and unbiased method of moments estimator based on $E(X_i^2)$. *Hint:* Again, use the fact that $Var(Y) = E(Y^2) - E(Y)^2$.

   (d) (6 pts) Show that $\sum_{i=1}^{n} X_i^2$ is a sufficient statistic for θ.

   (e) (6 pts) Find the maximum likelihood estimator for θ.

10. (25 pts) Let $X_1, X_2, ...X_n$ be a random sample from a distribution that has pdf:
    $$f(x; \theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2} \exp(-\frac{x - \theta_1}{\theta_2}) & \text{for } \theta_1 \leq x < \infty; \\ 0 & \text{otherwise .} \end{cases}$$

    (a) (7 pts) Find the maximum likelihood estimators for θ_1 and θ_2.

    (b) (9 pts) Find a most powerful test of the simple null hypothesis: $H_0 : \theta_1 = \theta_1^{(0)}$ and $\theta_2 = \theta_2^{(0)}$ vs. the simple alternative $H_a : \theta_1 = \theta_1'$ and $\theta_2 = \theta_2'$.

    (c) (9 pts) Derive a likelihood ratio test of the simple null hypothesis $H_0 : \theta_2 = \theta_2^{(0)}$ vs. the composite alternative $H_a : \theta_2 > \theta_2^{(0)}$ for unknown θ_1.