Statistics Qualifying Exam

TUESDAY, JUNE 14, 2011

NOON - 4 PM

Name :

1. Let U_1 and U_2 be independent uniform $(0, 1)$ random variables with pdf, $f(u) = 1$ for $0 < u < 1$; 0 elsewhere. Find the pdf of the ratio $Y = U_1/U_2$.

2. Let $Y_1 \leq Y_2 \leq \cdots \leq Y_6$ be the order statistics from a random sample of size 6 from a population with pdf $f(x)$ and cdf $F(x)$. Let the interval defined by $[Y_5, Y_6]$ be a confidence interval for the upper quartile of the population, $F^{-1}(0.75)$. What is the confidence level of this interval?

3. Let X_i be independent Poisson random variables with mean$=i\mu$ for $i = 1, \ldots, n$. Find $\hat{\mu}$, the maximum likelihood estimator of μ. (Recall that, for a Poisson random variable, X, with mean$=\mu$, $P(X = x) = \mu^x e^{-\mu}/x!$.)

4. A local radio station plays 40 songs during each 4-hour show. The program director at the station needs to know the total amount of airtime for the 40 songs so that time can also be programmed during the show for news and advertisements. The distribution of the lengths of songs, in minutes, is roughly symmetric with a mean length of 3.9 minutes and a standard deviation of 1.1 minutes.

   (a) Describe the sampling distribution of the sample mean song lengths for the random samples of 40 songs.

   (b) If the program manager schedules 80 minutes of news and advertisements for the 4-hour (240-minute) show, only 160 minutes are available for music. Approximately what is the probability that the total amount of time needed to play 40 randomly selected songs exceeds tha available airtime?

5. Let $X_1, \ldots, X_n$ be independent exponential random variables with pdf, $f(x) = e^{-x}$ for $x > 0$; 0 elsewhere. Let Y_1 be the minimum of the X's. Find the limiting distribution of $W_n = nY_1$ and its pdf.

6. Consider the linear model

$$Y_1 = \beta_1 + \beta_2 + \beta_3 + \epsilon_1; \quad Y_2 = \beta_1 + \beta_3 + \epsilon_2; \quad Y_3 = \beta_2 + \beta_3 + \epsilon_3$$

where $\epsilon_1, \epsilon_2, \epsilon_3$ i.i.d. $N(0, \sigma^2)$.

(a) Express in the form $\mathbf{y} = \mathbf{X}\beta + \epsilon$ where $\beta = (\beta_1, \beta_2, \beta_3)'$.

(b) Estimate $\beta_1 - 2\beta_2 + \beta_3$ and obtain its variance.

7. For ANOVA models with unequal sample size, we know that their type I and type III sum of squares are not the same. This is due to the fact that Type I SS weights each observation equally, while Type III SS weights each treatment equally.

Consider the bone data set where factor A is gender, $a = 2$ levels: male, female; and factor B is bone development, $b = 3$ levels: severely, moderately, or mildly depressed. The sample sizes are $3, 2, 2$ for male and $1, 3, 3$ for female. We use contrast to see what is being calculated for type I and type III SS.

(a) (5+5 pt) What is the type I and type III contrast statements (in SAS) for gender effect, the hypothesis is $H_0 : \mu_{1\cdot} = \mu_{2\cdot}$.

(b) (5+5 pt) Find the type I and type III contrasts for bone effect, $H_0 : (\mu_{\cdot 1} + \mu_{\cdot 2})/2 = \mu_{\cdot 3}$.

8. A survey of 1000 students concluded that 274 students chose a professional baseball team, A, as his or her favorite team. In 1991, the same survey was conducted involving 760 students. It concluded that 240 of them also chose team A as their favorite.

(a) Estimate the proportions of students favoring team A between the two surveys.

(b) Compute a 95% confidence interval for the difference between the proportion of students favoring team A between the two surveys.

(c) Is there a significant difference between two proportions? Make use of a P-value.

9. A production plant cost-control engineer is responsible for cost reduction. One of the costly items in his plant is the amount of water used by the production facilities each month. He decided to investigate water usage by collecting 17 observations on his plant's water usage and other variables. He had heard about multiple regression, but since he was quite skeptical he added a column of random numbers to his original observations. Use the attached information.

Data code
X_1 = average monthly temperature(F)
X_2 = average of production (M pounds)
X_3 = number of plant operating days in the month
X_4 = number of persons on the monthly plant payroll
X_5 = two-digit random number
Y = is the monthly water usage (gallons)

(a) Find the fitted regression model of $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \epsilon_i$ and complete the ANOVA table on your answer sheet (not on the exam sheet).

(b) Test the hypothesis $H_0 : \beta_1 = \beta_3 = \beta_5 = 0$ v.s. $H_1 : \beta_3 = \beta_5 = 0$. Use $\alpha = .05$.

(c) Perform a stepwise regression using a $\alpha = .05$ level of significance for entering and staying.

(d) Comment on the role of the variable X_5.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model				
Error		722691		
Corrected Total		3192632		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	$Pr > \|t\|$
Intercept	1	6487.35183	1371.64753	4.73	0.0006
x1	1	14.11699	5.33024	2.65	0.0226
x2	1	0.21378	0.04703	4.55	0.0008
x3	1	-126.99857	49.44390	-2.57	0.0261
x4	1	-22.37849	7.56462	-2.96	0.0130
x5	1	-1.34918	2.38034	-0.57	0.5822

Number in Model	R-Square	C(p)	AIC	MSE	SSE	Variables in Model
1	0.3978	16.2616	201.8103	128164	1922459	x2
1	0.1708	27.2961	207.2500	176495	2647418	x4
1	0.0817	31.6266	208.9853	195462	2931930	x1
1	0.0079	35.2113	210.2988	211163	3167442	x3
1	0.0043	35.3854	210.3601	211926	3178883	x5
2	0.5742	9.6907	197.9183	97097	1359361	x2 x4
2	0.4885	13.8573	201.0373	116650	1633106	x1 x2
2	0.4223	17.0730	203.1054	131741	1844373	x2 x3
2	0.3979	18.2605	203.8097	137313	1922389	x2 x5
2	0.2737	24.2966	206.9981	165640	2318960	x1 x4
2	0.1812	28.7900	209.0351	186726	2614168	x3 x4
2	0.1708	29.2961	209.2500	189101	2647418	x4 x5
2	0.1382	30.8770	209.9042	196520	2751284	x1 x3
2	0.0911	33.1699	210.8104	207280	2901925	x1 x5
2	0.0118	37.0226	212.2321	225360	3155042	x3 x5
3	0.6319	8.8875	197.4434	90399	1175191	x1 x2 x4
3	0.6268	9.1372	197.6791	91662	1191601	x2 x3 x4
3	0.5929	10.7846	199.1570	99987	1299830	x1 x2 x3
3	0.5774	11.5361	199.7908	103785	1349205	x2 x4 x5
3	0.4892	15.8213	203.0126	125441	1630736	x1 x2 x5
3	0.4225	19.0653	205.1008	141836	1843869	x2 x3 x5
3	0.3475	22.7068	207.1747	160239	2083110	x1 x3 x4
3	0.2746	26.2503	208.9757	178147	2315914	x1 x4 x5
3	0.1812	30.7890	211.0347	201085	2614103	x3 x4 x5
3	0.1481	32.3963	211.7080	209208	2719705	x1 x3 x5
4	0.7670	4.3213	191.6673	61983	743798	x1 x2 x3 x4
4	0.6379	10.5974	199.1654	96344	1156134	x1 x2 x4 x5
4	0.6293	11.0144	199.5636	98628	1183531	x2 x3 x4 x5
4	0.5935	12.7516	201.1287	108139	1297663	x1 x2 x3 x5
4	0.3485	24.6612	209.1502	173343	2080114	x1 x3 x4 x5
5	0.7736	6.0000	193.1779	65699	722691	x1 x2 x3 x4 x5