Statistics Qualifying Exam

9am-1pm, Monday, May 4, 2015

Name :

1. Let $X_1 \sim Gamma(2,2)$, $X_2 \sim Gamma(2,4)$ and $X_3 \sim Gamma(1,2)$ be independent random variables. The pdf of a gamma distribution is $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}\exp(-\frac{x}{\beta})$, $x > 0$.

   (a) Find the mean of $Y = \frac{4X_2}{X_1+X_3} + \frac{2}{X_2} + 3X_2X_3^4 + 1$, that is $E(Y)$.

   (b) Prove that $E(\frac{1}{X_1} - \log(X_1)X_2 + X_2^2 + 1) \geq \frac{1}{4} - 8log(4) + 9$. (Explain)

   (c) If $Y_1 = 3X_1 + 2X_2$ and $Y_2 = 5X_1 + 2$ find the joint pdf of Y_1 and Y_2. Are Y_1 and Y_2 independent?

   (d) Find the conditional mean of $Y_1|Y_2 = y_2$, $E(Y_1|Y_2 = y_2)$.

2. An insect lays a large number of eggs, each surviving with probability $p = \frac{1}{2}$. Suppose that the number of eggs laid, Y, follow a Poisson($m = 8$) distribution. Further, if we assume that each egg's survival is independent Bernoulli trial. On average, how many eggs will survive? (Explain)

3. (a) Let $\overline{X}$ denote the mean of a random sample of size 25 from a gamma-type distribution with $\alpha = 4$ and $\beta(> 0)$. Use the Central Limit Theorem (CLT) to find an approximate 95.4% confidence interval for β.

   (b) The time to process orders at the service counter of a pharmacy are exponentially distributed with mean 10 minutes. If 100 customers visit the counter in a 2-day period, what is the probability that at least half of them need to wait more than 10 minutes?

4. Suppose that $X_1, X_2, \ldots$ are $i.i.d$ $Uniform(0,4)$ random variables.
   Denote $U_n = \sum_{i=1}^n X_i/n$ and $V_n = \frac{2}{n(n+1)}\sum_{i=1}^n i \cdot X_i$. (Use $\sum_{i=1}^n i^2 = n(n+1)(2n+1)/6$.)

   (a) Show that U_n converges to 2 in Probability, i.e., $U_n \xrightarrow{P} 2$.

   (b) Show that V_n converges to 2 in Probability, i.e., $V_n \xrightarrow{P} 2$.

5. Let $X_1, \ldots, X_m$ be iid Uniform$(0,\theta)$, $Y_1, \ldots, Y_n$ be iid Uniform$(-\theta,\theta)$, and also let X's be independent of the Y's where $\theta(> 0)$ is the unknown parameter.

   (a) Show that the distribution of $|Y_1|$ is $Uniform(0,\theta)$.

   (b) Find the MLE of θ based on $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$.

6. Consider the following ANOVA table for the multiple linear model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon,$$

where the error term ϵ are independent and identical normal random variables with mean 0 and variance σ^2.

Source	SS	df	MS
Model	2176606	3	725535
X_1	136366	1	136366
$X_2 \mid X_1$	2033565	1	2033565
$X_3 \mid X_1, X_2$	6675	1	6675
Error	985530	48	20532
Total	3162136	51	

(a) Find $SSR(X_2, X_3 \mid X_1)$

(b) For the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$, using the answer to part a) test the following hypothesis: $H_0 : \beta_2 = \beta_3 = 0$ vs. H_1: Not both β_2 and $\beta_3 = 0$. Use $\alpha = 0.05$.

(c) Compute the coefficient of partial determination between Y and X_2 given that X_1 is in the model, i.e., find $R^2_{Y2|1}$. Interpret the result.

(d) With the given information, can a "best" multiple linear regression model be selected using a forward selection procedure for this application? If yes, show your detailed procedure; if no, explain.

7. A study is conducted to investigate the effect of temperature (factor A) and humidity (factor B) on the force required to separate an adhesive product from a certain material. Both factors are treated as fixed effects. The following output was obtained from SAS program that performed analysis on a two-factor study on this experiment. PROC GLM was used.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
A	___	488.83	___________	_______	< 0.0001
B	___	352.67	___________	_______	< 0.0001
Interaction	___	3.00	___________	_______	0.9579
Error	___	______	___________		
Corrected Total	23	1001.83			

The GLM Procedure

Level of temperature	N	force Mean	force Std Dev
1	6	36.3333333	4.84424057
2	6	31.8333333	5.49241902
3	6	28.0000000	4.42718872
4	6	24.1666667	5.41910202

Level of humidity	N	force Mean	force Std Dev
1	12	33.9166667	5.31649805
2	12	26.2500000	5.54526825

Level of temperature	Level of humidity	N	force Mean	force Std Dev
1	1	3	39.6666667	3.51188458
1	2	3	33.0000000	3.60555128
2	1	3	36.0000000	3.00000000
2	2	3	27.6666667	3.78593890
3	1	3	31.6666667	2.51661148
3	2	3	24.3333333	1.52752523
4	1	3	28.3333333	4.16333200
4	2	3	20.0000000	2.0000000

(a) What conclusions would you draw about the factor effects based on the information given above? Use a nominal significance level $\alpha = 0.05$. (Note: You are not required to complete the ANOVA table above.)

(b) Examine the factor effects of temperature by pairwise comparison. Use the Tukey multiple comparison procedure with a familywise error rate of $\alpha = 0.05$.

(c) Denote $L = (\mu_1+\mu_2)/2-(\mu_3+\mu_4)/2$, where μ_i represents the factor level mean at level i of temperature (factor A). Test the hypothesis $H_0 : L = 0$ versus $H_1 : L \neq 0$ using a t-test ($\alpha = 0.05$).

8. A consumer organization studied the effect of age of automobile owner on size of cash offer for a used car by utilizing 12 persons in each of the two age groups (young, middle) who acted as the owner of a used car. A medium price, six-year-old car was selected for the experiment, and the "owners" solicited cash offers for this car from 24 dealers selected at random from the dealers in the region. Randomization was used in assigning the dealers to the "owners." The offers(in hundred dollars) follow.

	j													
i	1	2	3	4	5	6	7	8	9	10	11	12		
Young	23	25	21	22	21	22	20	23	19	22	19	21	$\bar{Y}_{1.} = 21.5$	$S_1^2 = 1.73^2$
Middle	28	27	27	29	26	29	27	30	28	27	26	29	$\bar{Y}_{2.} = 27.75$	$S_2^2 = 1.29^2$
													$\bar{Y}_{..} = 24.625$	$S^2 = 3.52^2$

Note: $\bar{Y}_{1.}, \bar{Y}_{2.}, S_1^2$, and S_2^2 represent the sample means and the unbiased sample variances for the age group "young" and "middle", respectively. $\bar{Y}_{..}$ and S^2 represent the grand sample mean and unbiased sample variance.

(a) Use a two-sample t-test to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$, where μ_1 and μ_2 represent the mean size of cash offers for age group "young" and "middle", respectively. Please clearly specify the assumptions made in the procedure. Use $\alpha = 0.05$.

(b) Assume the ANOVA model $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ applicable to the above data, where ϵ_{ij} are independent and identical random variables with $\epsilon_{ij} \sim N(0, \sigma^2)$, $i = 1, 2$, and $j = 1, \ldots, 12$. Complete the ANOVA table below and conduct the F test for the equality of factor level means. Clearly state the value of the test statistic, the sampling distribution under the null hypothesis, and your conclusion. Use $\alpha = 0.05$.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Age Group	___	___________	___________	______	<.0001
Error	___	___________	___________		
Corrected Total	___	285.6250000			