Statistics Qualifying Examination

Answer all questions and show all work.
This exam is closed-note/book. You need to use a calculator.

1. Let X_1, X_2, X_3 be iid with common probability density function (pdf) $f_{X_j}(x) = e^{-x}$ if $x > 0$ and zero elsewhere, for $j = 1, 2, 3$.

   (a) Find the joint pdf of $Y_1 = X_1$, $Y_2 = X_1 + X_2$, and $Y_3 = X_1 + X_2 + X_3$.

   (b) Find the conditional joint pdf of Y_1 and Y_2 given $Y_3 = y_3$.

2. Answer the following questions:

   (a) Let X follow the chi-square distribution with k degrees of freedom. Derive the moment generating function of X, and use it to find the mean and variance of X.

   (b) Let X_1 follow the chi-square distribution with r_1 degrees of freedom. Let X_2 be a random variable independent of X_1, and its pdf is unknown. Let $Y = X_1 + X_2$ and we find that Y follows the chi-square distribution with r degrees of freedom where $r_1 < r$. Show that X_2 has a chi-square distribution with $r - r_1$ degrees of freedom.

3. Let $X_1, \ldots, X_n$ be a random sample from a binomial distribution with the probability mass function as $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, 2, \ldots, n$ and $0 < p < 1$.

   (a) Find the maximum likelihood estimator (MLE) $\hat{p}$ of p.

   (b) Is the MLE $\hat{p}$ an efficient estimator of p? Clearly justify your answer.

   (c) Find the MLE $\hat{\tau}$ of $\tau = P(X \leq 1)$.

4. Let $X_1, \ldots, X_n$ be a random sample from a $Gamma(3, \theta)$ distribution, where $0 < \theta < \infty$.

   (a) Find the exact likelihood ratio test of size α for testing the hypotheses $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Clearly state the likelihood ratio, the rejection region and the decision rule.

(b) For $\theta_0 = 3$ and $n = 5$, specify the rejection region so that the test that rejects the null hypothesis has a significant level 0.05.

(c) Obtain the power function of the test in (a) with $\theta_0 = 3$, $n = 5$ and $\alpha = 0.05$.

5. A simple linear regression model is specified and fit: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $i = 1, \ldots, n$. We assume that ϵ_i's follow $N(0, \sigma^2)$ independently. Use the partially complete output below to answer questions.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	***	***	***	***	***
Error	14	***	0.399		
Corrected Total	***	89.696			

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	$Pr > \|t\|$
Intercept	1	0.9798	***	***	0.011
X	1	-8.3088	***	***	***

df	13	14	15	16
$t_{0.025;df}$	2.1604	2.1448	2.1315	2.1199
$t_{0.05;df}$	1.7709	1.7613	1.7531	1.7459

(a) Calculate R^2.

(b) Conduct an F test to decide whether or not there is a significant linear association between X and Y. Use $\alpha = 0.05$. To get full credit, you need to give the hypotheses, test statistic, degrees of freedom, (range of) p-value, and your conclusion.

(c) Construct a 95% confidence interval for β_1.

6. The electric power consumed each month by a chemical plant is thought to be related to the average ambient temperature X_1, the number of days in the month X_2, the average product purity X_3, and the tons of product produced X_4. The past year's historical data are available. Consider all 4 explanatory variables as covariates in the multiple regression model with normal errors. The sample size is $n = 12$. Use the following output to answer the questions.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model		4957.24074			
Error					
Corrected Total		6656.25000			

Parameter Estimates

Variable	Parameter DF	Estimate	Standard Error	t Value	$Pr > \|t\|$
Intercept	1	-102.71324	207.85885	-0.49	0.6363
x1	1	0.60537	0.36890	1.64	0.1448
x2	1	8.92364	5.30052	1.68	0.1361
x3	1	1.43746	2.39162	0.60	0.5668
x4	1	0.01361	0.73382	0.02	0.9857

Number in Model	R-Square	C(p)	AIC	MSE	SSE	Variables in Model
1	0.6446	1.7471	67.4074	236.57679	2365.76786	x2
1	0.5647	3.9371	69.8397	289.73308	2897.33079	x1
1	0.0024	19.3586	79.7922	664.03676	6640.36765	x3
1	0.0001	19.4218	79.8198	665.56870	6655.68695	x4
2	0.7314	1.3665	66.0471	198.66239	1787.96148	x1 x2
2	0.6463	3.6989	69.3479	261.56262	2354.06360	x2 x3
2	0.6447	3.7437	69.4032	262.77130	2364.94169	x2 x4
2	0.6412	3.8385	69.5194	265.32872	2387.95845	x1 x3
3	0.7447	3.0003	67.4353	212.38659	1699.09274	x1 x2 x3
3	0.7316	3.3612	68.0386	223.33621	1786.68970	x1 x2 x4
3	0.6466	5.6930	71.3406	294.07956	2352.63647	x2 x3 x4
3	0.6414	5.8343	71.5143	298.36753	2386.94025	x1 x3 x4
4	0.7447	5.0000	69.4347	242.71561	1699.00926	x1 x2 x3 x4

(df_1, df_2)	(1,7)	(1,8)	(1,9)	(1,10)	(1,11)	(2,7)	(2,8)	(2,9)	(2,10)	(2,11)
$F_{0.1;df_1,df_2}$	3.5894	3.4579	3.3603	3.2850	3.2252	3.2574	3.1131	3.0065	2.9245	2.8595
$F_{0.05;df_1,df_2}$	5.5914	5.3177	5.1174	4.9646	4.8443	4.7374	4.4590	4.2565	4.1028	3.9823
$F_{0.25;df_1,df_2}$	8.0727	7.5709	7.2093	6.9367	6.7241	6.5415	6.0595	5.7147	5.4564	5.2559

(a) In the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$, test $H_0 : \beta_2 = 0$ vs. $H_a : \beta_2 \neq 0$. To get full credit, give the test statistic, degrees of freedom, (range of) p-value, and your conclusion. Use $\alpha = 0.05$.

(b) In the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$, test $H_0 : \beta_1 = \beta_3 = 0$ vs. H_a : Not H_0. To get full credit, give the test statistic, degrees of freedom, (range of) p-value, and your conclusion. Use $\alpha = 0.05$.

(c) Pick the best model and explain your choice.

(d) Build a regression model with the forward procedure using $\alpha = 0.05$.

7. The following data are the treatment means and variance from an experiment where each treatment was randomly and equally allocated to a total of 32 experimental units.

Trt	1	2	3	4
Mean	3	0	1	2
Variance	5	7	6	7

Suppose the experimenter planned to test the following three hypotheses

(i) $H_o : 3\mu_2 = \mu_1 + 2\mu_4$

(ii) $H_o : 2\mu_2 = \mu_1 + \mu_3$

(iii) $H_o : \mu_3 = \mu_1$

(iv) $H_o : \mu_2 = -2\mu_1 + \mu_3$

Answer the following questions:

(a) Calculate the MSE (Mean Square Error).

(b) Using $\alpha = 0.05$, test each of the four hypotheses (two sided).

(c) Which of the four linear combinations of means are contrasts? Why?

(d) Are any pairs of contrasts orthogonal? Which ones and why?

(e) Find one contrast that is orthogonal to the first contrast (i).

df	27	28	29	30	31
$t_{0.025;df}$	2.0518	2.0484	2.0452	2.0423	2.0395
$t_{0.05;df}$	1.7033	1.7011.	1.6991	1.6973	1.6955

Table 1: Weight Gain of Rabbits Under Six Diets

	Diet						
Breed-Litter	1	2	3	4	5	6	Litter Mean
1		32.6	35.2			42.2	36.67
2	40.1	38.16	40.9				39.70
3			34.6	37.5		34.3	35.47
4	44.9		43.9		40.8		43.2
5			40.9	37.3	32.0		36.73
6		37.3			40.5	42.8	40.20
7	45.2	40.6		37.9			41.23
8	44.0				38.5	51.9	44.80
9		30.6		27.5	20.6		26.23
10	37.3			42.3		41.7	40.43
Diet Total	211.5	179.2	195.5	182.5	172.4	212.9	
Diet Mean	42.3	35.84	39.1	36.5	34.48	42.58	38.47

8. A study of the difference of 6 proposed Diets on the weight gain of young rabbits is proposed. Because weight varies considerably amongest young rabbits, it is proposed to block the experiment based on the ten available breeds. For the 10 breeds, there is 1 litters of rabbits available of varying sizes. The minimum litter size for the 10 breeds is 3. Therefore, only 3 of the 6 diets can be observed in any particular breed-litter. The design was applied as shown in Table 1 with data from this experiment. Answer the following questions.

(a) Verify that this is a balanced incomplete block design (BIBD) also find the parameters of BIBD.

(b) Write a mathematical model to describe the above experiment. Completely identify all terms in your model and include all conditions (constraints or distributional) placed on the terms in the model.

(c) Test if there is a difference between the six Diets. (Based on the partial SAS output (D=Diet, L=Litter) on the next page or calculating by hand).

(d) Obtain the estimate of treatment mean for Diet 4 (i.e., the least square mean). (Based on the partial SAS output (D=Diet, L=Litter) or calculating by hand)

The GLM Procedure

Class	Levels	Values
L	10	L1 L10 L2 L3 L4 L5 L6 L7 L8 L9
D	6	D1 D2 D3 D4 D5 D6

Number of Observations Read 60
Number of Observations Used 30

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	*	889.113889	*	*	0.0005
Error	*	*	*		
Corrected Total	*	1039.886667			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
D	*	293.3786667	58.6757333	5.84	0.0035
L	*	*	*	*	*

Source	DF	Type III SS	Mean Square	F Value	Pr > F
D	*	158.7272222	31.7454444	3.16	0.0382
L	*	*	*	*	*

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		31.83333333	B	2.58863137	12.30	<.0001
D	D1	-3.30000000	B	2.24182052	-1.47	0.1617
D	D2	-5.04166667	B	2.24182052	-2.25	0.0400
D	D3	-2.90000000	B	2.24182052	-1.29	0.2154
D	D4	-3.23333333	B	2.24182052	-1.44	0.1698
D	D5	-8.52500000	B	2.24182052	-3.80	0.0017
D	D6	0.00000000	B	.	.	.
L	L1	7.48055556	B	2.79604144	2.68	0.0173
L	L10	10.77777778	B	2.79604144	3.85	0.0016
L	L2	11.61388889	B	2.79604144	4.15	0.0008
L	L3	5.67777778	B	2.79604144	2.03	0.0604
L	L4	16.27500000	B	2.79604144	5.82	<.0001
L	L5	9.78611111	B	2.69433295	3.63	0.0025
L	L6	12.88888889	B	2.69433295	4.78	0.0002
L	L7	13.25833333	B	2.69433295	4.92	0.0002
L	L8	16.90833333	B	2.79604144	6.05	<.0001
L	L9	0.00000000	B	.	.	.