Statistics Qualifying Exam

12:00 pm - 4:00 pm, Thursday, May 3^{rd}, 2018

1. Suppose that X_1 and X_2 are independent and identically distributed with the exponential distribution whose marginal probability density function (pdf) is given by

$$f_{X_j}(x_j) = \begin{cases} \frac{1}{2}e^{-\frac{x_j}{2}} & x_j > 0, \\ 0 & \text{elsewhere,} \end{cases}$$

   for $j = 1, 2$.

   Define $Y_1 = \frac{1}{2}(X_1 - X_2)$ and $Y_2 = X_2$.

   (a) Find the joint pdf of X_1 and X_2.

   (b) Find the joint pdf of Y_1 and Y_2.

   (c) Find the marginal pdf of Y_1 (whose distribution is called the *Laplace* or *double exponential* distribution).

2. Suppose that $X_1, \ldots, X_n$ are independent and identically distributed with the Bernoulli distribution, i.e.,

$$f_{X_i}(x_i) = p^{x_i}(1-p)^{1-x_i}, \quad x_i = 0,\ 1, \quad i = 1, \ldots, n.$$

   (a) Find the maximum likelihood estimator (mle) of p when $x_1, \ldots, x_n$ are provided.

   (b) Find the moment generating function (mgf) of X_i.

   (c) Prove that $Y = \sum_{i=1}^n X_i$ follows the binomial distribution whose mgf is written by $M(t) = (1 - p + pe^t)^n$.

   (d) Find the mean and variance of Y.

3. Let $X_1, \ldots, X_n$ be a random sample from a distribution with the pdf as

$$f_{X_i}(x_i|\theta) = \theta^{-1}\exp(-x_i/\theta),$$

   where $\theta > 0, x_i \geq 0$, for $i = 1, \ldots, n$.

   (a) Find mle $\hat{\tau}$ of $\tau(\theta)$, where $\tau(\theta) = 1/\theta$.

   (b) Find the asymptotic distribution of $\sqrt{n}(\hat{\tau} - \tau(\theta))$ as $n \to \infty$.

   (c) Show $\hat{\tau}$ obtained in Part (a) is a biased estimator for $\tau(\theta)$. Derive an unbiased estimator $\tilde{\tau}$ using $\hat{\tau}$.

   (d) Let $W = \sum_{i=1}^n X_i$ and $Y_1 < Y_2 < \cdots < Y_n$ be the ordered sample of $X_1, \ldots, X_n$. Show that W and Y_n/W are independent.

4. Consider a random sample $X_1, \ldots, X_n$ from a $N(\mu, \sigma^2)$ distribution with the pdf given as

$$f_{X_i}(x_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right],$$

where $-\infty < \mu < \infty$, $\sigma^2 > 0$, and $-\infty < x_i < \infty, \quad i = 1, \ldots, n$. Assume μ is unknown and σ^2 is known.

(a) Find the minimum variance unbiased estimator (MVUE) $\hat{\theta}$ of $\theta = \mu^2$.

(b) Derive the **exact** likelihood ratio test for the following hypothesis

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu \neq \mu_0.$$

Please carefully specify the likelihood ratio Λ, the test statistic, the sampling distribution of the test statistic under the null hypothesis, and the decision rules of a size α test.

(c) Obtain the power function of the test using the decision rule derived in Part (b).

5. The following data show the effect of two soporific drugs (change in hours of sleep) on two groups consisting of 10 patients each:

group	change in hours of sleep	sample mean	sample s.d.
1	0.7, -1.6, -0.2, -1.2, -0.1, 3.4, 3.7, 0.8, 0.0, 2.0	0.75	1.79
2	1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4	2.33	2.00

Perform a two-sample t-test for the effect of two soporific drugs:
H_0 : the effect of drug 1 = the effect of drug 2 vs H_1 : the effect of drug 1< the effect of drug 2.

6. A simple linear regression model is specified as below

$$Y_i = \text{Constant} + \beta_1 X_i + \epsilon_i, \quad i = 1, \ldots, n,$$

where Y_i is the value of the response variable in the i^{th} observation, Constant and β_1 are parameters, and X_i is the value of the predictor variable in the i^{th} observation. The random errors ϵ_i's are independently and identically distributed as $N(0, \sigma^2)$, σ^2 is unknown.

Use the partially completed output from the simple linear regression to answer the following questions.

Predictor	Coeff	SE	T	P
Constant	.9798	.3367	2.91	.011
X	-8.3088	.5725	??	??

s=?? R-sq = 93.8%

Analysis of Variance

Source	DF	SS	MS	F
Regr	??	84.106	??	??
Error	14	5.590	??	
Total	15	??		

(a) Find all of the missing values labeled by ??.

(b) Find a point estimate of σ^2.

(c) Test for significance of regression. Test for significance of β_1. Comment on the two results. Use $\alpha = .05$.

(d) Construct a 95% confidence interval (CI) on β_1.

(e) Write down the fitted regression model and use it to compute the residual when $x = .58$ and $y = -3.30$.

(f) Use the fitted regression model to compute the mean response and predicted future response when $x = .6$. Given that $\overline{x} = \sum_{i=1}^{n} x_i/n = .52$ and $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = 1.218294$, construct a 95% CI on the mean response and a 95% prediction interval (PI) on the future response.

7. 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 litter 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 following design was applied:

Table 1: Weight Gain of Rabbits Under Six Diets

	Diet						
Breed	1	2	3	4	5	6	Breed 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

(a) Verify that this is a balanced incomplete block design (BIBD) and 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 assumptions) placed on the terms in the model.

(c) Test if there is a difference between the six diets. (Based on the partial SAS output or calculating by hand)

(d) Obtain the estimates of treatment means for Diet 4 (i.e., the adjusted means or the least square means). (Based on the partial SAS output or calculating by hand)

The GLM Procedure

Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	**	889.108236	63.507731	6.32	0.0005
Error	**	********	******		
Corrected Total	29	1039.846147			

R-Square	Coeff Var	Root MSE	resp Mean
0.855038	8.240593	3.170046	38.46867

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Breed	**	730.5357467	81.1706385	8.08	0.0002
Diet	**	********	********	***	****

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Breed	**	596.0441689	66.2271299	6.59	0.0008
Diet	**	********	********	***	****

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	42.61277778	B	2.24156129	19.01	<.0001
Breed 1	-3.30055556	B	2.79571813	-1.18	0.2562
Breed 2	0.85444444	B	2.79571813	0.31	0.7641
Breed 3	-5.10000000	B	2.69402139	-1.89	0.0778
Breed 4	5.49888889	B	2.79571813	1.97	0.0680
Breed 5	-0.99166667	B	2.79571813	-0.35	0.7277
Breed 6	2.10611111	B	2.79571813	0.75	0.4629
Breed 7	2.47722222	B	2.69402139	0.92	0.3724
Breed 8	6.13055556	B	2.69402139	2.28	0.0380
Breed 9	-10.78277778	B	2.79571813	-3.86	0.0016
Breed 10	0.00000000	B	.	.	.
Diet 1	-3.30500000	B	2.24156129	-1.47	0.1610
Diet 2	-5.03166667	B	2.24156129	-2.24	0.0403
Diet 3	-2.90500000	B	2.24156129	-1.30	0.2146
Diet 4	-3.23333333	B	2.24156129	-1.44	0.1697
Diet 5	-8.52500000	B	2.24156129	-3.80	0.0017
Diet 6	0.00000000	B	.	.	.

8. A mechanical engineer is studying the thrust force developed by a drill press. She suspects that the drilling speed and feed rate of the material are the most important factors. She selects three feed rates, and three drill speeds. Initially, three reps of each of the 9 treatments were to be run. However, it was found that certain feed rates were incompatible with certain drill speeds. Thus, the experiment was unbalanced. The following thrust force values for the 9 treatments are given below:

	FEED RATES		
DRILL SPEED	15	30	45
1	26, 27.7	24.7, 23.4, 25.3	26.6, 24.5
2	27.1	23.4, 21.2	27.5,28.5, 26.8
3	28.4, 29.1, 27.7	23.4,22.4	28.6, 27, 29.1

Suppose the following statistical model is used to fit the data.

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk},$$

$$i = 1, 2, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n_{ij},$$

where τ_i: ith level effect of Factor A (DRILL SPEED), β_j: jth level effect of Factor B (FEED RATES), $(\tau\beta)_{ij}$: interaction effect of combination ij, and $\epsilon_{ijk} \sim N(0, \sigma^2)$ i.i.d.. For parameter estimation, we impose the following constraints as in the lecture notes: $\sum_i \tau_i = 0$, $\sum_j \beta_j = 0$ and $\sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0$. The ANOVA of the data was done in SAS and the output are given in the next page.

(a) What are the least square estimates of τ_3 and $(\tau\beta)_{33}$?

(b) Test if the interaction between feed rate and drilling speed is significant. To get full credits, include all the steps and give a test statistic, the corresponding p-value, and your conclusion. (use $\alpha = 0.05$)

(c) Assume an additive model (two factors without interaction):

$$Y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}.$$

Based on this model calculate the type I sum of squares due to the main effects of feed rate of the material (FEED RATES) and the sum of squares of errors (SSE). (Explain all the steps.)

The GLM Procedure

Dependent Variable: force

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	96.3210476	12.0401310	10.83	0.0002
Error	12	*****	******		
Corrected Total	20	109.6657143			

R-Square	Coeff Var	Root MSE	force Mean
0.878315	4.038175	1.054540	26.11429

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DRILL	2	4.79357143	2.39678571	**	0.1586
FEED	**	*******	*******	33.58	***
DRILL*FEED	**	*******	*******	***	***

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DRILL	2	3.28144405	1.64072203	***	0.2674
FEED	**	*******	35.09443033	31.56	*****
DRILL*FEED	**	*******	*******	***	****

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	28.60000000	B	1.05454045	27.12	<.0001
DRILL 1	-3.05000000	B	1.29154300	-2.36	0.0360
DRILL 2	-0.82000000	B	1.15519118	-0.71	0.4914
DRILL 3	0.00000000	B	.	.	.
FEED 15	-0.20000000	B	1.21767842	-0.16	0.8723
FEED 30	-5.70000000	B	1.29154300	-4.41	0.0008
FEED 45	0.00000000	B	.	.	.
DRILL*FEED 1 15	1.50000000	B	1.61083714	0.93	0.3701
DRILL*FEED 1 30	4.61666667	B	1.61083714	2.87	0.0142
DRILL*FEED 1 45	0.00000000	B	.	.	.
DRILL*FEED 2 15	-0.48000000	B	1.67845387	-0.29	0.7798
DRILL*FEED 2 30	0.22000000	B	1.56413625	0.14	0.8905
DRILL*FEED 2 45	0.00000000	B	.	.	.
DRILL*FEED 3 15	0.00000000	B	.	.	.
DRILL*FEED 3 30	0.00000000	B	.	.	.
DRILL*FEED 3 45	0.00000000	B	.	.	.