Statistics Qualifying Examination

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

1. Suppose that Y is a random variable with moment-generating function $m(t)$ and U is given by $U = aY + b$, where a and b are known constants.

(a) Show that the moment-generating function of U is $e^{bt}m(at)$.

(b) If Y has mean μ and variance σ^2, use the moment-generating function of U to derive the mean and variance of U.

(c) Assume that Y is a normally distributed random variable with mean μ and variance σ^2. Use the result in (a) to derive the moment-generating function of $U = -3Y + 4$ and identify the distribution of U.

2. Suppose that $X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, \ldots, Y_{n_2}$, and $W_1, W_2, \ldots, W_{n_3}$ are independent random samples from normal distributions with respective unknown means μ_1, μ_2, and μ_3 and common variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$. Suppose that we want to estimate a linear function of the means: $\theta = a_1\mu_1 + a_2\mu_2 + a_3\mu_3$. Because the maximum-likelihood estimator (MLE) of a function of parameters is the function of the MLEs of the parameters, the MLE of θ is $\hat{\theta} = a_1\bar{X} + a_2\bar{Y} + a_3\bar{W}$.

(a) Find the distribution of the estimator $\hat{\theta}$. Clearly specify the mean and variance of $\hat{\theta}$.

(b) If the sample variances are given by S_1^2, S_2^2, and S_3^2, respectively, derive the distribution of the following statistic T, where

$$T = \frac{\hat{\theta} - \theta}{\sqrt{\left[\frac{(n_1-1)S_1^2+(n_2-1)S_2^2+(n_3-1)S_3^2}{n_1+n_2+n_3-3}\right]\left(\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}\right)}}$$

Show all the steps.

(c) Give a confidence interval for θ with confidence level $1 - \alpha$.

3. Let $X_1, \ldots, X_n$ be a random sample from a Poisson(λ) distribution with the probability mass function (pmf),

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots.$$

(a) Find the maximum likelihood estimator (mle) $\hat{\lambda}$ of λ.

(b) Prove that the mle $\hat{\lambda}$ is the minimum variance unbiased estimator of λ.

(c) Show that the mle $\hat{\lambda}$ achieves the Rao-Cramér lower bound.

4. Let $X_1, \ldots, X_n$ be a random sample from a distribution with the probability density function,

$$f(x|\beta) = \frac{1}{\Gamma(2)\beta^2}\, x\, \exp\left(-\frac{x}{\beta}\right)$$

for $0 < x < \infty$ for $\beta > 0$.

(a) Derive the exact likelihood ratio test (LRT) for the hypotheses

$$H_0 : \beta = \beta_0 \quad \text{vs.} \quad H_1 : \beta \neq \beta_0.$$

You need to specify the test statistic, its sampling distribution under H_0, and the decision rule with the rejection region for a size α-test.

(b) Obtain the power function for the exact LRT.

5. The following data pertain to the demand for a product (in 1000's of units) and its price (in dollars) charged in seven different markets:

Price (x)	11	9	12	10	15	12	6
Demand (y)	145	177	109	135	81	118	218

Note that : $\sum x = 75$, $\sum x^2 = 851$, $\sum y = 983$, $\sum y^2 = 150469$ and $\sum xy = 9785$.

(a) Fill in the ANOVA Table below for the simple linear regression model, i.e., $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ for $i = 1, \ldots, 7$.

Source	df	Sum of Squares (SS)	Mean Squares (MS)	F-ratio
Regression	?	?	?	?
Error	?	?	?	?
Total (Corrected)	?	?		

(b) Test $H_0 : \beta_1 = 0$ v.s. $H_1 : \beta_1 \neq 0$ using $\alpha = .05$ and obtain a 95% confidence interval for the slope, β_1.

(c) Obtain a 95% confidence interval for the mean response at $x = 13$.

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 variables as covariates in the multiple regression model with normal errors. Use the following information,with $n = 12$.

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	Standard Estimate	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
2	0.5648	5.9352	71.8377	321.87210	2896.84886	x1 x4
2	0.0026	21.3539	81.7901	737.69185	6639.22662	x3 x4
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

(a) Complete the ANOVA table with the regression model, $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \epsilon_i$.

(b) Test the hypothesis $H_0 : \beta_3 = 1.0$ v.s. $H_1 : \beta_3 \neq 1.0$ using $\alpha = .05$.

(c) Test the hypothesis $H_0 : \beta_1 = \beta_4 = 0$ v.s. $H_1 : \beta_4 = 0$ using $\alpha = .05$.

(d) Perform a stepwise variable selection method using a 5% level of significance for entering and staying.

7. An engineer suspects that the surface finish of a metal part is influenced by the feed rate and the depth of cut. He selects three feed rates and four depths of cut. He then conducts a factorial experiment and obtains the following data.

		Depth of	Cut (in)	
Feed rate (in/min)	1	2	3	4
1	74, 64, 60	79, 68, 73	82, 88, 92	99, 104, 96
2	92, 86, 88	98, 104, 88	99, 108, 95	104, 110, 99
3	99, 98, 102	104, 99, 95	108, 110, 99	114, 111, 107

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

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}; k = 1, 2, 3$$

where τ_i ($i = 1, 2, 3$) and β_j ($j = 1, 2, 3, 4$) are the effects of feed rate and cut depth, and $(\tau\beta)_{ij}$ are their interactions. For parameter estimation, we impose the following constraints as in the lecture notes: $\sum_i \tau_i = \sum_j \beta_j = \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0$.

Some summary statistics and ANOVA results from SAS are give on the next page.

(a) What are the estimates of τ_1 and $(\tau\beta)_{22}$?

(b) Test if the interaction between feed rate and cut depth is significant. To get full credits, give a test statistic, the corresponding p-value and your conclusion.

(c) If we plan to perform pairwise comparison for *all* treatment combinations, which procedure should we use? What is the corresponding critical difference (using $\alpha = 5\%$)? *[You don't need to find the results of comparison. Calculating the critical difference alone is sufficient.]*

(d) Use the Bonferroni method to compare the following treatments (i.e. the level combinations of speed and depth): (2,3), (2,4), (3,3) and (3,4), *pairwisely* (Use $\alpha = 6\%$). Calculate the critical difference and report your results of comparison. You can report the result as we have seen in SAS output by labeling significantly different combinations with different Latin letter.

Grand mean: 94.333

feed	mean	depth	mean
1	81.583	1	84.778
2	97.583	2	89.778
3	103.833	3	97.889
		4	104.889

feed	depth	mean
1	1	66.000
1	2	73.333
1	3	87.333
1	4	99.667
2	1	88.667
2	2	96.667
2	3	100.667
2	4	104.333
3	1	99.667
3	2	99.333
3	3	105.667
3	4	110.667

Dependent Variable: finish

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	5842.667	531.151515	18.49	<.0001
Error	24	689.333	28.722222		
Co Total	35	6532.000			

ource	DF	Type I SS	Mean Square	F Value	Pr > F
feed	2	3160.500	1580.250	55.02	<.0001
depth	3	2125.111	708.370	24.66	<.0001
feed*depth	6	557.056	92.843	3.23	0.0180

8. A agriculture research facility wants to investigate the relationship between corn yield and 4 important factors. An unreplicated two-level 2^4 factorial experiment is designed for the sake of saving time and money. The design and generated data are given below:

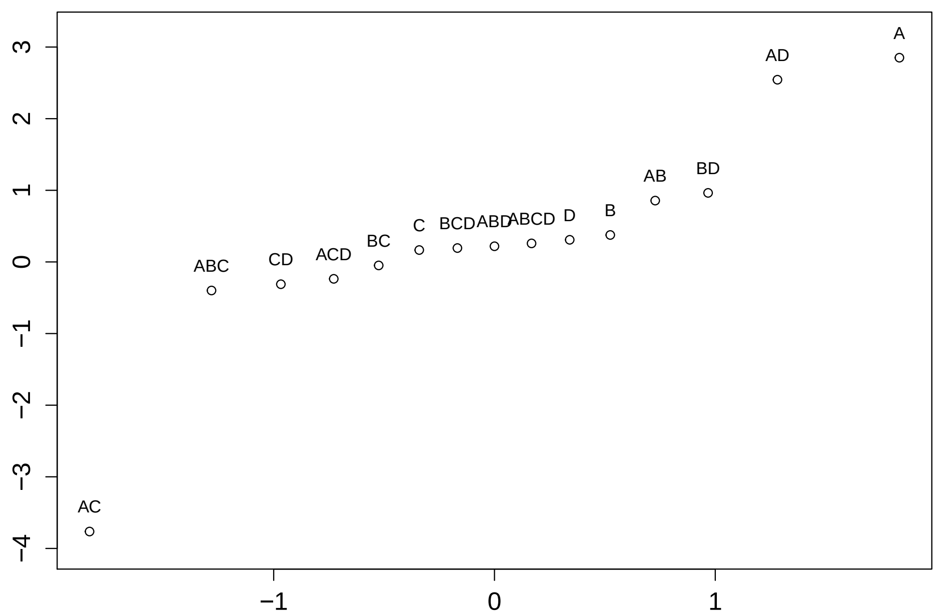
factor				
A	B	C	D	yield
-	-	-	-	8.54
+	-	-	-	11.08
-	+	-	-	6.90
+	+	-	-	12.03
-	-	+	-	12.13
+	-	+	-	8.93
-	+	+	-	11.32
+	+	+	-	8.08
-	-	-	+	5.57
+	-	-	+	13.75
-	+	-	+	5.55
+	+	-	+	16.16
-	-	+	+	9.14
+	-	+	+	9.60
-	+	+	+	9.69
+	+	+	+	12.02

(a) Please calculate the estimated effect of interaction AC, and its corresponding Sum of Squares, SS_{AC}.

(b) The QQ plot of the estimated effects is given on the next page. Please indicate the potentially important effects. Explain briefly.

(c) Given the QQ plot, the data analyst at the facility decides to fit the data in SAS by including in the `model` statement: the main effects A, C, D, and the interactions AC and AD. What is the degree of freedom of ***Residuals*** in the ANOVA table resulted from fitting this model? Explain briefly.

(d) The data analyst then fits a regression model by creating covariates taking value 1 when the corresponding factor is $+$, and taking value -1 when the factor is $-$. What is the fitted coefficient $\hat{\beta}_{AC}$ corresponding to the interaction AC?


A
AD
AB
BD
B
D
ABCD
ABD
BCD
C
BC
ACD
CD
ABC
AC
3
2
1
0
−1
−2
−3
−4
−1
0
1