Statistics Qualifying Exam

12:00 pm - 4:00 pm, Tuesday, May 6, 2025

Answer questions with showing all of your work.
This is closed-note/book. A calculator is allowed.

1. Let $Y_1 = \frac{1}{2}(X_1 - X_2)$, where X_1 and X_2 have the joint pdf

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} \frac{1}{4}\exp\left(-\frac{x_1+x_2}{2}\right), & 0 < x_1 < \infty,\ 0 < x_2 < \infty \\ 0, & \text{elsewhere} \end{cases}.$$

   Find the pdf of Y_1.

2. Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\theta)$ with $\lambda, \theta > 0$. X and Y are independent.
   (a) Find the moment generating function (MGF) of X.
   (b) Find the distribution of $X + Y$. Show all your steps.
   (c) Show that the conditional distribution of $X \mid X + Y$ is binomial. Clearly specify all parameters and justify your answer.

3. Let $X_1, \ldots, X_n$ be independently and indentically distributed (i.i.d) with probability density function (pdf)

$$f(x;\theta) = \theta x^{\theta-1}, \quad 0 \le x \le 1, \quad 0 < \theta < \infty.$$

   (a) Find the maximum likelihood estimator (MLE) of θ, say $\hat{\theta}$.
   (b) Find the pdf of the MLE $\hat{\theta}$. Show all your steps.
   (c) Is the MLE $\hat{\theta}$ an efficient estimator of θ? Clearly justify your answer.
   (d) Find the uniformly minimum variance unbiased estimator (MVUE) of θ, say $\tilde{\theta}$.
   (e) Find the limiting distribution of $\sqrt{n}(\tilde{\theta} - \theta)$, as $n \to \infty$. Note that $\tilde{\theta}$ is the MVUE.

4. Suppose that $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ where $\sigma \in (0, \infty)$ is the unknown parameter but $\mu \in (-\infty, \infty)$ is assumed known. With pre-assigned $\alpha \in (0, 1)$, derive a level α likelihood Ration (LR) test for a null hypothesis $H_0 : \sigma^2 = \sigma_0^2\ (> 0)$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ in the implementable form.

5. A simple linear regression model is specified and fit:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \;\; \epsilon_i \text{ iid } \; N(0, \sigma^2),$$

for $i = 1, \ldots, n$. Use the partially complete output below to answer questions.

Use the following t-values:
$t_{0.025,40} = 2.021$, $t_{0.05,40} = 1.684$, $t_{0.025,5} = 2.776$, $t_{0.05,5} = 2.132$, $t_{0.025,43} = 2.017$, $t_{0.05,43} = 1.681$, $t_{0.025,6} = 2.447$, $t_{0.05,6} = 1.943$, $t_{0.025,44} = 2.015$, $t_{0.05,44} = 1.680$, $t_{0.025,3} = 3.182$, $t_{0.05,3} = 2.353$,

```
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value  Pr(>F)
X1         1 152044    ****   ***      ***
Residuals 43  *****    6035

---

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   99.595    ****      ****    ***
X1            -5.433    ****      ****    ***
---
```

(a) Compute the coefficient of determination R^2, and explain its meaning (in the context of the simple linear regression).

(b) Conduct an F test to decide whether or not there is a 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. Show all the steps.

6. You have been asked to determine the pricing of the restaurants dinner menu such that it is competitively positioned with other high-end Italian restaurants in the target area. In particular, your role in the team is to analyze the pricing data that have been collected in order to produce a regression model to predict the price of dinner. Actual data from surveys of customers of 168 Italian restaurants in the target area are available. The data are in the form of the average of customer views on

   - Y = Cost = the price (in $US) of dinner (including one drink & a tip)
   - X_1 = Food = customer rating of the food (out of 30)
   - X_2 = Décor = customer rating of the decor (out of 30)
   - X_3 = Service = customer rating of the service (out of 30)
   - X_4 = East = dummy variable = 1 (0) if the restaurant is east (west) of Fifth Avenue

   (a) Write down the expression for the FITTED regression lines for the two different locations of east or west of Fifth Avenue (no $\hat{\beta}$'s, use numbers). Find also ?????(A) (Estimate the coefficient of Service).

   (b) Interpret estimates $\hat{\beta}_1$. Words only, no Y's or X's in your interpretation.

   (c) Assume modeling assumptions hold true. After fitting the regression model, the researchers concluded that the null hypothesis that $\beta_2 \geq 1.5$ was rejected at significance level $\alpha = 0.05$ against the alternative hypothesis that $\beta_2 < 1.5$. One of the researchers concluded: "At significance level $\alpha = 0.05$, we reject the null hypothesis." Do you agree with this assessment?

   (d) Suppose you have already decided to drop the variable X_4 from the model (i.e., X_4 has been removed). Given this, is it worth including X_3 in the model that already contains X_1 and X_2? Explain. Use $\alpha = 0.05$, and give hypotheses, test statistic, which $F(a; df_1, df_2)$ or $t(a; df)$ you used, (range of) p-values, and conclusions, in order to get full credits.

   (e) Derive the distribution (including the parameters) of $\hat{\beta}_2 - \hat{\beta}_1$ given that all assumptions of the multiple regression hold. Explain all steps.

A summary of the multiple linear regression for Qn. #6 is given below.

Use the t-table given to you or the given F-values:
$F_{0.95,2,163} = 3.051471$, $F_{0.95,3,163} = 2.660061$, $F_{0.95,2,162} = 2.7764$, $F_{0.95,2,164} = 3.051127$,

```
Call:
lm(formula = Cost ~ Food + Decor + Service + East)

Residuals:
     Min       1Q   Median       3Q      Max 
-14.0465  -3.8837   0.0373   3.3942  17.7491 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -24.023800   4.708359  -5.102 9.24e-07 ***
Food          1.538120   0.368951   4.169 4.96e-05 ***
Decor         1.910087   0.217005   8.802 1.87e-15 ***
Service       ?????(A)   0.396232  -0.007   0.9945    
East          2.068050   0.946739   2.184   0.0304 *  
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: ??? (B) on 163 degrees of freedom
Multiple R-squared:  0.6279,    Adjusted R-squared:  0.6187 
F-statistic: 68.76 on 4 and 163 DF,  p-value: < 2.2e-16

SS type I
Analysis of Variance Table

Response: Cost
           Df Sum Sq Mean Sq  F value  Pr(>F)    
Food        1 5670.3  5670.3 172.2269 < 2e-16 ***
Decor       1 3223.7  3223.7  97.9142 < 2e-16 ***
Service     1    3.9     3.9   0.1192 0.73037    
East        1  157.1   157.1   4.7716 0.03036 *  
Residuals 163 5366.5    32.9                     
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1
```

7. An engineer is interested in the effects of cutting speed (A), tool geometry (B), and cutting angle (C) on the life (in hours) of a machine tool. Two levels of each factor are chosen. Suppose that a full replicate of the experiment cannot all be run using the same bar stock.

   (a) Set up a design to run the treatment combinations in two blocks of four treatment combinations each with the three-way interaction ABC effects confounded.

   (b) Assume the data obtained are listed as below. Estimate the factor effect, construct the ANOVA table and suggest a refined model.

   (c) Construct the ANOVA table for the refined model suggested in Part (b). Discuss if there exists any significant effects. Use $\alpha = 0.05$ in each of the F-tests.

Treatment Combination	(1)	a	b	ab	c	ac	bc	abc
I	22	32	35	55	44	40	60	39

8. An engineer is designing a battery for use in a device that will be subject to some extreme variations in temperatures. She decides to test three plate materials at three temperature levels. Because there are two factors at three levels, this design is sometimes called a 3^2 factorial design. Four batteries are tested at each combination of plate materials and temperature, and all 36 tests are run in random order. The experiment and the resulting observed battery life data are given in the table below. A longer life is preferred. See the attached SAS codes and output.

   The overall mean battery life is **105.53** as estimated from the experimental data below.

	Temperature ($^\circ F$)		
Material Type	**15**	**70**	**125**
1	130, 155 74, 180	34, 40 80, 75	20, 70 82, 58
2	150, 188 159, 126	136, 122 106, 115	25, 70 58, 45
3	138, 110 168, 160	174, 120 150, 139	96, 104 82, 60

   (a) Write a statistical model for this experiment and explain each of the terms. Specify all the model assumptions and constraints.

   (b) Provide a 95% confidence interval estimate of of the treatment mean under the treatment combination (temperature, material type) = ($70^\circ F$, 3).

   (c) Construct the ANOVA table based on the given output from SAS.

      i. Clearly specify the sources of sum of squares, the degrees of freedom, the mean squares, and the values of corresponding F statistics.

      ii. State your findings based on the ANOVA table. Use α=0.05 for each F test. Clearly specify your decision using the p-value rule and draw the context specific conclusion from all the necessary test(s).

   (d) Given Temperature = $70^\circ F$, carry out Tukey multiple comparison on the material type's effect. Use the family error rate α=0.05.

A SAS Code and its partially completed output for Qn. #8 below.

```
SAS CODES

data battery;
 input temp type @;
  do rep =1 to 4;
   input life @;
   output;
  end;
  datalines;
15 1 130 155 74 180
15 2 150 188 159 126
15 3 138 110 168 160
70 1 34 40 80 75
70 2 136 122 106 115
70 3 174 120 150 139
125 1 20 70 82 58
125 2 25 70 58 45
125 3 96 104 82 60
; run;

proc glm data=battery;
 class temp type;
 model life = temp|type;
means temp|type;
run;
```

==

SAS OUTPUT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	***	59416.22222	7427.02778	***	<.0001
Error	***	***	***		
Corrected Total	***	77646.97222			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
temp	***	39118.72222	***	***	<.0001
type	***	10683.72222	***	***	0.0020
temp*type	***	****	***	***	0.0186

Level of temp	N	-------------life------------ Mean	Std Dev
15	12	144.833333	31.6940870
70	12	107.583333	42.8834750
125	12	64.166667	25.6721757

Level of type	N	-------------life------------ Mean	Std Dev
1	12	83.166667	48.5888751
2	12	108.333333	49.4723676
3	12	125.083333	35.7655455

Level of temp	Level of type	N	-------------life------------ Mean	Std Dev
15	1	4	134.750000	45.3532432
15	2	4	155.750000	25.6173769
15	3	4	144.000000	25.9743463
70	1	4	57.250000	23.5990819
70	2	4	119.750000	12.6589889
70	3	4	145.750000	22.5444006
125	1	4	57.500000	26.8514432
125	2	4	49.500000	19.2613603
125	3	4	85.500000	19.2786583

TABLE II Percentage Points of the *t* Distribution

ν \ α	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551

ν = Degrees of freedom.

TABLE V Percentage Points of the Studentized Range Statistic (*Continued*)

$q_{0.05}(p, f)$

f	p: 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18.1	26.7	32.8	37.2	40.5	43.1	45.4	47.3	49.1	50.6	51.9	53.2	54.3	55.4	56.3	57.2	58.0	58.8	59.6
2	6.09	8.28	9.80	10.89	11.73	12.43	13.03	13.54	13.99	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.36	16.57	16.77
3	4.50	5.88	6.83	7.51	8.04	8.47	8.85	9.18	9.46	9.72	9.95	10.16	10.35	10.52	10.69	10.84	10.98	11.12	11.24
4	3.93	5.00	5.76	6.31	6.73	7.06	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.67	8.80	8.92	9.03	9.14	9.24
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49	6.65	6.79	6.92	7.04	7.14	7.24	7.34	7.43	7.51	7.59
7	3.34	4.16	4.68	5.06	5.35	5.59	5.80	5.99	6.15	6.29	6.42	6.54	6.65	6.75	6.84	6.93	7.01	7.08	7.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.65
10	3.15	3.88	4.33	4.66	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.12	6.20	6.27	6.34	6.41	6.47
11	3.11	3.82	4.26	4.58	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.14	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.46	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	6.00	6.06	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64	5.72	5.79	5.86	5.92	5.98	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.79	5.85	5.91	5.96
16	3.00	3.65	4.05	4.34	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.62	4.02	4.31	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.55	5.61	5.68	5.74	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.83	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.26	4.47	4.64	4.79	4.92	5.04	5.14	5.23	5.32	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.24	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.50	5.56	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.50	5.55	5.59
30	2.89	3.48	3.84	4.11	4.30	4.46	4.60	4.72	4.83	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.48
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74	4.82	4.90	4.98	5.05	5.11	5.17	5.22	5.27	5.32	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24