Université
Libanaise
Faculté d'Information
Br I

الجامعة اللبنانية
كلية الاعلام
الفرع الاول

امتحان الفصل الثاني
للعام الجامعي 2026/2025

المرحلة: إجازة Data Science	المادة: STAT2
السنة المنهجية: 2 الاولى	المدة: ساعتان
الاستاذ: د.حسين نصار	الدورة: 1

Question 1: Multiple Choice Questions

1. **An e-commerce startup wants to test if their new checkout process takes less than 90 seconds on average. The population standard deviation is unknown, and they measure a sample of 20 transactions. Which statistical test is the most appropriate to use?**
   a) Z-test
   b) 1-Sample T-test
   c) 2-Sample T-test
   d) One-Way ANOVA

2. **A digital marketer runs an A/B test and concludes that Design B leads to a significantly higher conversion rate than Design A, when in reality, there is no actual difference between the two designs. What type of error has occurred?**
   a) Type I Error
   b) Type II Error
   c) Standard Error
   d) Margin of Error

3. **A regional retail chain knows its historical average customer transaction is $35 with a known population standard deviation of $5. They sample 100 recent transactions to see if a weekend promotion increased spending. Which test should be used?**
   a) Z-test
   b) 1-Sample T-test
   c) Chi-Square Test
   d) Pearson Correlation

4. **An analyst finds a Pearson correlation coefficient of $r = -0.85$ between the price of a product and the quantity sold. What does this indicate?**
   a) A strong negative linear relationship; as price increases, quantity sold tends to decrease.
   b) A weak negative linear relationship; price has little effect on sales.
   c) A strong positive linear relationship; as price increases, quantity sold tends to increase.
   d) A causal relationship; increasing the price directly causes a drop in sales.

5. ***Before running an Independent 2-Sample T-test to compare the average daily sales between two store locations, an***

analyst must ensure certain statistical assumptions are met. Which of the following is NOT a required assumption for this test?

a) The two samples are completely independent of each other.
b) The data in both groups are approximately normally distributed (or the sample sizes are sufficiently large).
c) The population standard deviations for both store locations are exactly known.
d) The dependent variable being measured (daily sales) is continuous.

6. **A marketing manager wants to compare the average click-through rates (CTR) of digital ads placed on three different platforms: Instagram, Facebook, and TikTok. What is the appropriate statistical test to use?**
   a) 1-Sample T-test
   b) 2-Sample T-test
   c) Chi-Square Test of Independence
   d) One-Way ANOVA

7. **A linear regression model predicting monthly revenue from SEO budget yields an R-squared (R^2) value of 0.60. How should this be interpreted?**
   a) Revenue will increase by $0.60 for every dollar spent on SEO.
   b) 60% of the variation in monthly revenue can be explained by the SEO budget.
   c) 60% of the data points fall exactly on the regression line.
   d) There is a 60% probability that SEO budget increases revenue.

8. **A researcher wants to determine if there is an association between a customer's age group (Gen Z, Millennial, Gen X) and their preferred social media platform for shopping (TikTok, Instagram, Facebook). Which test is the most appropriate?**
   a) 2-Sample T-test
   b) One-Way ANOVA
   c) Chi-Square Test of Independence
   d) Simple Linear Regression

9. **In comparing the effectiveness of two promotional campaigns, a calculated t-statistic is 2.50. The critical t-value for a significance level of 0.05 is 2.05. What is the correct conclusion?**
   a) Reject the null hypothesis; the campaigns perform significantly differently.
   b) Fail to reject the null hypothesis; there is not enough statistical evidence to say the campaigns differ.
   c) Accept the null hypothesis; the campaigns perform the same.
   d) The second campaign is 2.5 times better than the first.

10. **A regression equation modeling the relationship between the number of influencers hired (x) and total website visits (y) is $y = 1500 + 400x$. What does the value 400 represent in this context?**
    a) The store expects 400 website visits if zero influencers are hired.
    b) 40% of the variation in website visits is explained by the influencers.
    c) For each additional influencer hired, website visits are expected to increase by 400.
    d) For every 400 website visits, the store needs to hire one influencer.

Question 2: Evaluating Ad Creative Performance

Scenario:
An online pet supply brand ran two different Meta Ad creatives (Video A and Image B) for a new line of products. The digital marketing manager wants to determine if there is a statistically significant difference in the average daily clicks generated by the two ads. The campaign ran for 14 days, and the daily clicks were recorded.

The summary statistics for the daily clicks are as follows:

Ad A (Video): Sample size (n) = 14 days, Sample Mean ($\bar{x}_1$) = 125 clicks, Sample Standard Deviation (s_1) = 15 clicks

Ad B (Image): Sample size (n) = 14 days, Sample Mean ($\bar{x}_2$) = 142 clicks, Sample Standard Deviation (s_2) = 18 clicks

Questions:

1. **(2 points)** State the null hypothesis (H_0) and the alternative hypothesis (H_a).
2. **(4 points)** Calculate the pooled variance and the resulting t-statistic for this test. (Show your work).
3. **(2 points)** Based on your calculated t-statistic and the provided critical t-value, do you reject or fail to reject the null hypothesis?
4. **(2 points)** What is your final business recommendation to the marketing manager regarding which ad creative to scale?

Question 3: Optimizing Email Promotions (One-Way ANOVA)

Scenario:

An e-commerce manager is testing three different email promotional strategies to see if they impact the Average Order Value (AOV). The three strategies tested are: Group 1 (15% Off Discount), Group 2 (Buy One Get One 50% Off), and Group 3 (Free Shipping). The manager randomly assigns 15 different days to each of the three promotional strategies (Total N = 45 days). The data is run through a statistical software, yielding a partially completed ANOVA table.

Use a significance level of $\alpha = 0.05$.

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F-Statistic
Between Groups	1800 SST	[A] k-1 42	[C]	[E]
Within Groups	8400 SSE	[B] n-k 2	[D]	
Total	10200 SS Total	44 n-1 44		

Questions:

1. **(1 points)** State the null hypothesis (H_0) and the alternative hypothesis (H_a).
2. **(5 points)** Calculate the missing values in the ANOVA table [A, B, C, D, and E]. (Show your work for the degrees of freedom).
3. **(2 points)** Compare your calculated F-statistic [E] to the critical F-value. Do you reject or fail to reject the null hypothesis?
4. **(2 points)** As a data analyst, what do you tell the e-commerce manager about the impact of the different email strategies on Average Order Value?

Question 4: Analyzing Checkout Conversion

Scenario:

An e-commerce manager wants to know if a user's decision to complete a purchase (Conversion) is dependent on the type of checkout process they are assigned: "Guest Checkout" vs. "Forced Account Creation." They run an A/B test with 200 users and record the following observed frequencies:

Checkout Type	Completed Purchase	Abandoned Cart	Row Totals
Guest Checkout	60	40	**100**

Checkout Type	Completed Purchase	Abandoned Cart	Row Totals
Account Creation	30	70	**100**
Column Totals	**90**	**110**	**Grand Total = 200**

Questions:

1. **(1 points)** State the null hypothesis (H□) and the alternative hypothesis (H_a).
2. **(3 points)** Calculate the **Expected Frequencies** for all four cells in the table. (Show your work).
3. **(3 points)** Calculate the final **Chi-Square test statistic (χ^2)**. (Show the formula and your calculation steps).
4. **(1 points)** Calculate the degrees of freedom for this table and use the provided distribution table to find the **critical χ^2 value**.
5. **(2 points)** Compare your calculated test statistic to the critical value. Do you reject or fail to reject the null hypothesis? What should the manager do with the checkout process?

Question 5: Predicting Retail Sales

Scenario:

The manager of a local minimarket wants to understand the relationship between their weekly social media advertising spend (x, in dollars) and their total weekly revenue (y, in dollars). They track the data over a 10-week period (n = 10) and calculate the following summary statistics:

- Average weekly ad spend = \$50 $\bar{x}$
- Average weekly revenue = \$2,500 $\bar{y}$
- Standard deviation of ad spend = \$10 S_x
- Standard deviation of revenue = \$150 S_y
- Pearson correlation coefficient = 0.8333 r
- Standard Error of the slope = 2.5

Questions:

1. **(5 points)** Write out the final estimated regression equation ($\hat{y} = a + bx$).
2. **(4 points)** Verify the hypothesis that there is a significant linear relationship between ad spend and revenue. State the null (H□) and alternative (H_a) hypotheses, calculate the test statistic, and state your final conclusion.
3. **(1 points)** If the manager plans to spend \$75 on social media ads next week, what is the predicted weekly revenue?

TWO-SAMPLE TEST OF MEANS—KNOWN σ	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ (11–2)		$S_b = \frac{S_{y \cdot x}}{\sqrt{\sum (x_i - \bar{x})^2}}$
TWO-SAMPLE TEST OF MEANS—UNKNOWN σ'S	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ (11–4)	TEST FOR THE SLOPE	$t = \frac{b - 0}{S_b}$ with $n - 2$ degrees of freedom [13–6]
POOLED VARIANCE	$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ (11–3)	CORRELATION COEFFICIENT	$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y}$ [13–1]

DEGREES OF FREEDOM FOR UNEQUAL VARIANCE TEST	$df = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$	[11–6]
TEST STATISTIC FOR NO DIFFERENCE IN MEANS, UNEQUAL VARIANCES	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	[11–5]
PAIRED *t* TEST	$t = \frac{\bar{d}}{s_d/\sqrt{n}}$	(11–7)

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{.865\sqrt{15-2}}{\sqrt{1-.865^2}} = 6.216$$

$$s_d = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}}$$

1. t-Distribution Critical Values (Includes Standard Normal Z)

One-Tail α	0.10	0.05	0.025	0.01	0.005
Two-Tail α	0.20	0.10	0.05	0.02	0.01
df = 1	3.078	6.314	12.706	31.821	63.657
df = 8	1.397	1.860	2.306	2.896	3.355
df = 10	1.372	1.812	2.228	2.764	3.169
df = 20	1.325	1.725	2.086	2.528	2.845
df = 25	1.316	1.708	2.060	2.485	2.787
df = 26	1.315	1.706	2.056	2.479	2.779
df = 30	1.310	1.697	2.042	2.457	2.750
Z (df = ∞)	**1.282**	**1.645**	**1.960**	**2.326**	**2.576**

2. Chi-Square (χ^2) Distribution Critical Values (Right-Tail)

Degrees of Freedom (df)	α = 0.10	α = 0.05	α = 0.025	α = 0.01
1	2.706	3.841	5.024	6.635
2	4.605	5.991	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086

3. F-Distribution Critical Values (α = 0.05)

Denominator df (df)	Numerator df (df)				
	1	2	3	4	5
10	4.965	4.103	3.708	3.478	3.326
20	4.351	3.493	3.098	2.866	2.711
30	4.171	3.316	2.922	2.690	2.534
40	4.085	3.232	2.839	2.606	2.449
42	4.073	3.220	2.827	2.594	2.438
50	4.034	3.183	2.790	2.557	2.400