DSBA Statistics **Spring Midterm** **25 March 2025**

Group: **Name:** **Problem 1 (30 pts)**

Six rival vendors at **Tony's Trinket Emporium** have been locked in a heated battle for the title of "Market Monarch." Their daily earnings (in £) were recorded over five days. By week's end, the total revenue was:

- **A (Tony's Tacos)**: £1200
- **B (Bessie's Biscuits)**: £980
- **C (Carla's Cat Sweaters)**: £550
- **D (Dave's Dubious Antiques)**: £420
- **E (Ethel's Existential Coffee)**: £360
- **F (Fred's Forgotten Stall)**: *Data "eaten by a very hungry spreadsheet"*

A frazzled intern started an ANOVA table but got distracted by a suspiciously loud debate about whether pineapple belongs on pizza. Your mission:

Source	**DF**	**Sum of Squares**	**Mean Square**	F-Stat	F_{crit} (5%)
Day					3.01
Vendor		245,800		3.85	
Error			4,200		
Total		398,500			

(a) **(2 pts)** Fill in the missing values.

(b) **(2 pts)** Is there enough evidence to declare a *statistically significant* showdown between vendors' daily earnings? What about days?

(c) **(2 pts)** Build a 90% confidence interval for the daily earnings gap between **A (Tacos)** and **E (Coffee)**.

(d) **(2 pts)** Construct 95% simultaneous confidence intervals for **A vs. E** (Tacos vs. Coffee)

(e) **(2 pts)** Construct 95% simultaneous confidence intervals for **B vs. C** (Biscuits vs. Cat Sweaters)

"In the realm of data, there are no shortcuts—only Z-scores and p-values."
— Anonymous Statistician

A tech startup, **DataWizards Inc.**, claims their new algorithm improves the conversion rate of website visitors into paying customers. To test this, they ran an A/B test:

- Version A (Control): Shown to 500 visitors. 92 converted.
- Version B (Algorithm): Shown to 480 visitors. 112 converted.

The CEO insists the algorithm is a "game-changer," but the marketing team suspects foul play. Your job: Investigate whether the difference in conversion rates is statistically significant or just *corporate wizardry*.

(a) Hypotheses Haunting. State the null and alternative hypotheses. *Is the CEO's claim a trick or treat?*

(b) Verify the conditions for using the two-proportion z-test.

(c) Calculate the pooled proportion and the z-test statistic.

(d) Using $\alpha = 0.05$, determine the critical value and conclude whether to reject H_0.

(e) Construct a 99% confidence interval for the difference in proportions $(p_B - p_A)$. Interpret it.

(f) If the true difference in proportions is 3%, how would doubling the sample size affect the test's power?

A pet food company, **Paws and Reflect**, claims there is **no association** between type of pet ownership and food preference. To test this, they surveyed 300 customers and recorded the following observed frequencies:

Food Preference	**Dog Owners**	**Cat Owners**	**No Pets**
Organic Food	45	30	25
Non-Organic Food	30	40	35
No Preference	25	30	40

(a) State the null and alternative hypotheses.

(b) Calculate the expected frequencies for all cells.

(c) Compute the chi-squared test statistic.

(d) Determine the degrees of freedom, critical value at $\alpha = 0.05$, and conclude whether to reject H_0.

"In statistics, variance isn't just a number—it's a lifestyle."
— Anonymous Quality Control Engineer

A manufacturer of precision springs claims the variance in their spring lengths is **exactly 0.04 mm²**. A rival company, **TensionMasters**, suspects the variance is actually **higher**. To investigate, a random sample of 20 springs is tested. The total length of all springs is **200.0 mm**, and the sum of their squared lengths is **2005.0 mm²**. Assume spring lengths are normally distributed.

(a) State the null and alternative hypotheses. Is the rival company stirring drama or detecting truth?

(b) What conditions must be met for this test? Explain why they are satisfied here.

(c) Calculate the sample variance.

(d) Compute the χ^2 test statistic.

(e) Find the critical value at $\alpha = 0.05$. Should we reject H_0?

(f) Estimate the p-value range for your test statistic using the χ^2 table below.

A researcher collected data on two variables, x and y, for 12 observations. The data is provided below:

x	2	3	4	5	6	7	8	9	10	11	12	13
y	4	5	6	8	7	9	10	11	12	13	14	15

(a) Estimate the Ordinary Least Squares (OLS) linear regression line $\hat{y} = b_0 + b_1 x$. Show all steps, including sums of x, y, xy, and x^2. Use a calculator for computations.

(b) Interpret the slope coefficient b_1 in the context of the data.

(c) Calculate the Spearman rank correlation coefficient between x and y. Show all steps, including ranking and differences.

(d) Based on your results, comment on the relationship between x and y.

A school district is evaluating two teaching methods (Method X and Method Y). Independent random samples of students were taught using each method, with the following results:

Statistic	**Method X (n=15)**	**Method Y (n=20)**
Sample Mean ($\bar{x}$)	78	75
Sample Standard Deviation (s)	4	5

Test whether the population mean scores differ at $\alpha = 0.05$. Assume normality and **unequal variances**.

(a) State the null and alternative hypotheses.

(b) Verify the conditions for conducting this test.

(c) Calculate t-test statistic. Show all steps, including the standard error.

(d) Compute number of degrees of freedom

(e) Find the critical value(s) and conclude whether to reject H_0.

(f) Estimate the p-value range using the t-distribution table below. Interpret its significance.