Math 119 Name: ________________________

FINAL

(LESSONS 33-43: HYPOTHESIS TESTING; CORRELATION AND REGRESSION)
MATH 119 – SPRING 2022 – KUNIYUKI
120 POINTS TOTAL

No notes or books allowed. A scientific calculator is allowed. Simplify as appropriate.
You do not have to reduce fractions. For example, 10/20 does not have to be rewritten as ½.

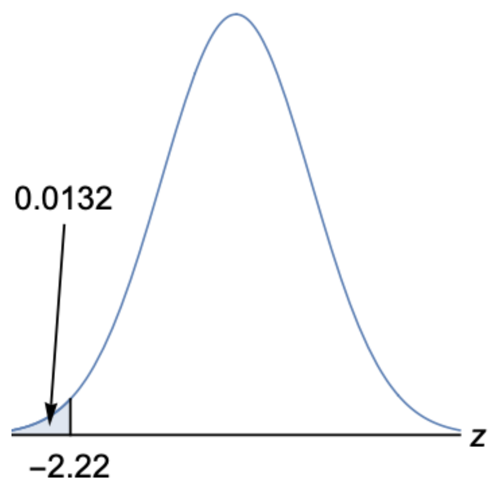
THE FORMULA SHEET IS THE LAST SHEET.

If you round off in the middle, round off to at least five significant digits.

1) (28 points). In a North Carolina primary election, if no one receives at least 30% of the vote, a runoff election between the top two finishers is expected. A challenger claims that the state senator they are running against has the approval of less than 30% of likely voters. Test the challenger's claim at the 0.05 significance level.
Warning: Write percents as decimals.

Let p = the proportion of likely voters who approve of the state senator.

Use these hints about the z distribution:



• a) Write the **setup** for the hypothesis test. The setup will include H_0, H_1, identifying which is the claim, and the significance level. (5 points)

• b) Is this test **two-tailed, right-tailed, or left-tailed**? (2 points)

We gather **sample data**. We randomly select 800 likely voters in a poll.
211 of them approve of the state senator.

• c) Find the **sample proportion** $\hat{p}$. Round off to three decimal places. (3 points)

• d) **Verify** that normal approximations are appropriate in this problem. (2 points)

• e) Compute the *z* **test statistic** for our sample and round it off to two decimal places. Show all work, as in class! (7 points)

• f) Find the corresponding ***P*-value**. (2 points)

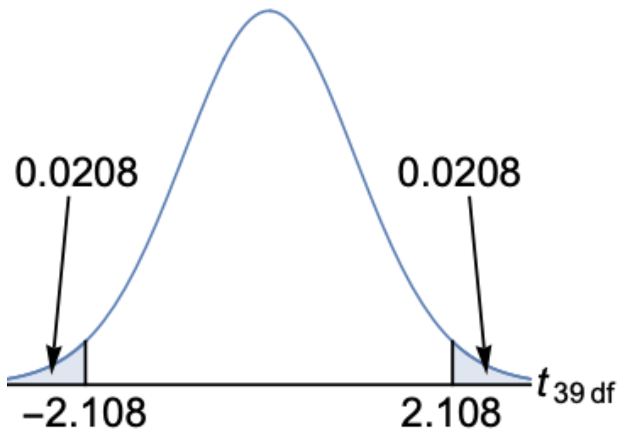
• g) **Decide** whether or not to reject H_0. (2 points)

• h) Write our **conclusion** relative to the claim. (5 points)

2) (23 points). A bottled water company claims that the average pH level of the water in its bottles is 7.00 pH units, which is neutral (neither acidic nor alkaline). Test this claim at the 0.10 significance level.

Let μ = the population mean pH level of the water in the company's bottles.

Use these hints about the *t* distribution on 39 degrees of freedom (df):



• a) Write the **setup** for the hypothesis test. The setup will include H_0, H_1, identifying which is the claim, and the significance level. (5 points)

• b) Is this test **two-tailed, right-tailed, or left-tailed**? (2 points)

We gather **sample data**. We randomly select 40 water bottles produced by the company. The sample mean pH level is 6.85 pH units and the sample standard deviation (SD) is 0.45 pH units.

• c) Compute the *t* **test statistic** for our sample and round it off to three decimal places. Show all work, as in class! (6 points)

• d) Find the corresponding ***P*-value**. (3 points)

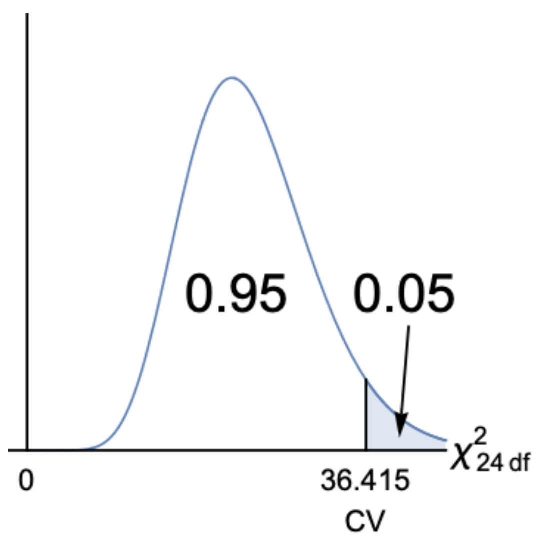
• e) **Decide** whether or not to reject H_0. (2 points)

• f) Write our **conclusion** relative to the claim. (5 points)

3) (19 points). The company from 2) is competing with a rival company that claims that the population standard deviation (SD) of the pH levels of water in the first company's bottles is more than 0.40 pH units. Test this claim at the 0.05 significance level. (The rival company does not have access to the sample data from 2).) Assume that the pH levels of water in the first company's bottles are approximately normally distributed. Use **the traditional (classical) method** of hypothesis testing.

Let σ = the population standard deviation (SD) of the pH levels of water in the first company's bottles.

Use these hints about the χ^2 distribution on 24 degrees of freedom (df):



• a) Write the **setup** for the hypothesis test. The setup will include H_0, H_1, identifying which is the claim, and the significance level. (5 points)

• b) Is this test **two-tailed, right-tailed, or left-tailed**? (2 points)

We gather **sample data**. We randomly select 25 of the first company's water bottles. The sample standard deviation (SD) of the pH levels is 0.43 pH units.

• c) Compute the χ^2 **test statistic** for our sample and round it off to three decimal places. Show all work, as in class! (5 points)

• d) **Decide** whether or not to reject H_0. (2 points)

• e) Write our **conclusion** relative to the claim. (5 points)

4) (2 points). When we reject a H_0 that is true, what type of error have we made?
Box in one:

Type I error　　　　Type II error

5) (21 points). According to an April AP poll, 56% of American adults favored mask requirements for public transportation in April, 24% opposed them, and 20% were undecided. Let's assume that these proportions were correct for April. Test the claim that the distribution of opinions on mask requirements for public transportation is the same this week as it was in April at the 0.05 significance level.

- a) Write the **setup** for the hypothesis test. The setup will include H_0, H_1, identifying which is the claim, and the significance level. (5 points)

  Let p_F = the proportion of American adults who favor mask requirements for public transportation this week.

  Let p_O = the proportion of American adults who oppose mask requirements for public transportation this week.

  Let p_U = the proportion of American adults who are undecided about mask requirements for public transportation this week.

- b) Is this test **two-tailed, right-tailed, or left-tailed**? (2 points)

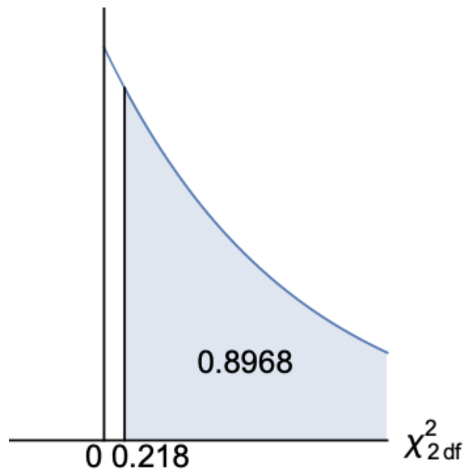
We gather **sample data**. We randomly sample 1085 American adults. (This was the sample size used in the AP poll.) Among them, 600 favor mask requirements for public transportation this week, 265 oppose them, and 220 are undecided.

- c) Write the **Observed (*O*) Table** and the **Expected (*E*) Table**. Don't round. Note that each of the *E* values is at least 5, so we may apply the methods of this Lesson. (7 points)

Opinion on mask requirements for public transportation	**Observed (*O*) Frequencies**	**Expected (*E*) Frequencies**
Favor		
Oppose		
Undecided		

• d) The χ^2 **test statistic** is about 0.218. **Decide** whether or not to reject H_0. (2 points)

Use these hints about the χ^2 distribution on 2 degrees of freedom (df):



• e) Write our **conclusion** relative to the claim. (5 points)

6) (14 points). A lung cancer patient sues a vaping company and claims that there is statistical dependence between use of its vaping product and incidents of lung cancer. Test the patient's claim at the 0.05 significance level.

- a) Write the **setup** for the hypothesis test. The setup will include H_0, H_1, identifying which is the claim, and the significance level. (5 points)

- b) Is this test **two-tailed, right-tailed, or left-tailed**? (2 points)

We gather **sample data** by studying a random sample of 3000 American adults. The results are summarized in the two-way **Observed (*O*) Table** below:

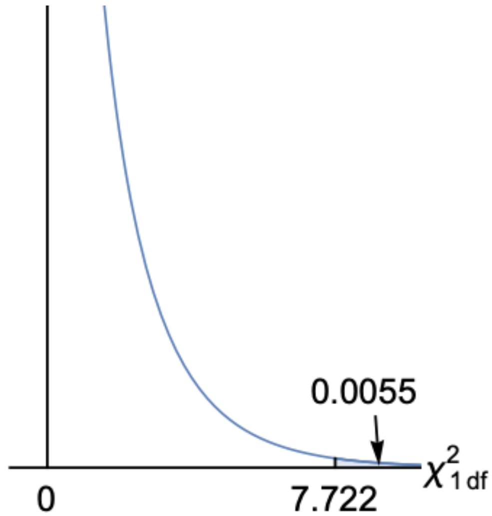
	Have lung cancer	Do not have lung cancer	**← Presence of lung cancer**
Do not use the product	200	2600	**2800**
Use the product	25	175	**200**
↑ Usage of product	**225**	**2775**	**3000**

The **Expected (*E*) Table** is below:

	Have lung cancer	Do not have lung cancer	**← Presence of lung cancer**
Do not use the product	210	2590	**2800**
Use the product	15	185	**200**
↑ Usage of product	**225**	**2775**	**3000**

• c) The χ^2 **test statistic** is about 7.722. We use 1 df. **Decide** whether or not to reject H_0. (2 points)

Use these hints about the χ^2 distribution on 1 degree of freedom (df):



• d) Write our **conclusion** relative to the claim. (5 points)

7) (2 points). Fill in the blank: If a regression line for sample data is given by

$\hat{y} = 30 + 8x$, then along the regression line, for every increase of 1 unit in x,

there is an increase of _____ units in y.

8) (1 point). A student scores two standard deviations below the mean on Midterm 1 in a math class. According to the principle of **regression to the mean**, which of the following is the most likely outcome for the student on Midterm 2 in that class? Box in one:

a) The student will score one standard deviation below the mean on Midterm 2.

b) The student will score two standard deviations below the mean on Midterm 2.

c) The student will score three standard deviations below the mean on Midterm 2.

9) (2 points). Given sample bivariate data involving two variables, x and y, we obtain $r = 0.9$ and find the corresponding least squares regression model $\hat{y} = b_0 + b_1 x$.
What proportion of the variance of y is accounted for by x and the regression model? Box in the best answer below, based on the class notes and homework:

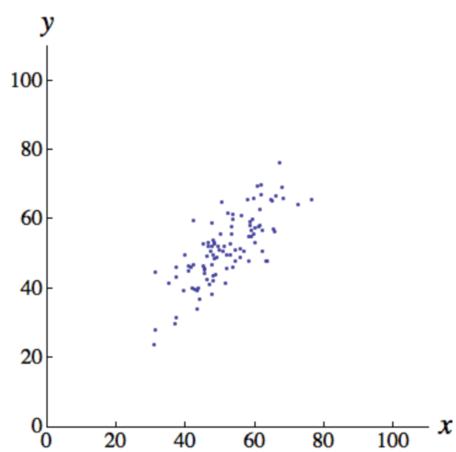
a) 9% b) 18% c) 81% d) 90% e) 99%

10) (8 points). (Matching)
For each variable, the average is 50 and the standard deviation is 10.

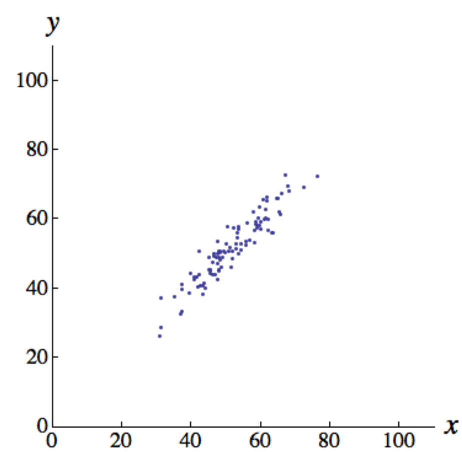
For one of the graphs below, $r = -0.90$.
For one of the graphs below, $r = 0.00$.
For one of the graphs below, $r = 0.80$.
For one of the graphs below, $r = 0.95$.

Fill in the blanks:

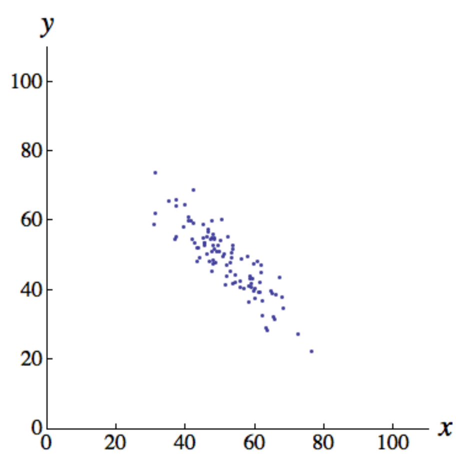
a) r for the graph below is ______.



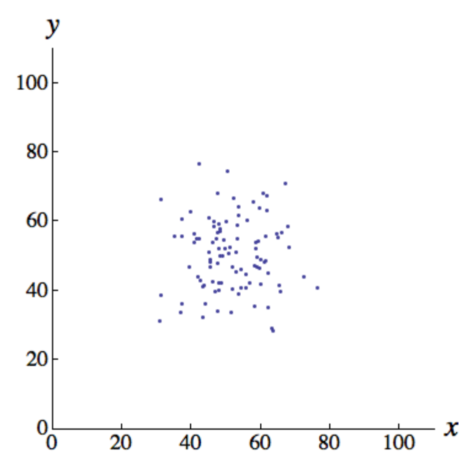
b) r for the graph below is ______.



c) r for the graph below is ______.



d) r for the graph below is ______.



MATH 119: FINAL FORMULA SHEETS

Sample Proportion of Successes

$$\hat{p} = \frac{x}{n}$$

Hypothesis Test for a Population Proportion or Probability p

(Assume $X \sim \text{Bin}(n, p)$. To justify a **normal approximation**, verify: $np \geq 5$, and $nq \geq 5$ under H_0, where $q = 1 - p$.)

The z Test Statistic for Tests for p

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

where p and q are obtained under H_0.

Hypothesis Test for a Population Mean μ (if σ is Unknown)

Assumptions:

- We are conducting a hypothesis test for a **population mean** μ.
- σ is unknown.
- Central Limit Theorem (CLT) applies:
  - § $X \overset{\text{approx.}}{\sim}$ Normal, or
  - § $n > 30$

The t Test Statistic for Tests for μ (if σ is unknown)

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

where μ is obtained under H_0.

We use the t distribution on $(n-1)$ **degrees of freedom (df)**.

(SEE NEXT PAGE!)

Hypothesis Test for a Population SD σ or Variance σ^2

Assumptions:

- We are conducting a hypothesis test for a **population SD** σ or **variance** σ^2.
- $X \overset{\text{approx.}}{\sim}$ Normal

The χ^2 Test Statistic for Tests for σ or σ^2

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

where σ^2 is obtained under H_0.

We use the χ^2 distribution on $(n-1)$ **degrees of freedom (df)**.

χ^2 Test for Goodness-of-Fit

- Expected frequency for category *i*: np_i, where p_i is obtained under H_0.
- We use the χ^2 distribution on $(k-1)$ **degrees of freedom (df)**, where *k* = the **number of categories**.
- Not needed for the Final: Test $\chi^2 = \sum \frac{(O-E)^2}{E}$

χ^2 Test for Independence

- We use the χ^2 distribution on $(r-1)(c-1)$ **degrees of freedom (df)**, where *r* = the **number of rows** and *c* = the **number of columns**.
- Not needed for the Final:

  Expected frequency for a cell = $\frac{(\text{row total})(\text{column total})}{n}$

- Not needed for the Final: Test $\chi^2 = \sum \frac{(O-E)^2}{E}$

Least squares regression line

Population data: $y = \beta_0 + \beta_1 x$ **Sample data:** $\hat{y} = b_0 + b_1 x$

Coefficient of Determination

r^2