SOLUTIONS TO THE FINAL

(LESSONS 32-43: HYPOTHESIS TESTING; CORRELATION AND REGRESSION)
MATH 119 – SPRING 2024 – 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.

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

1) Let p be the probability of heads for a magician's coin. We will test the claim that the coin is fair using a significance level of: $\alpha = 0.05$.

   H_0 states that the coin is **fair** (that is, $p = \frac{1}{2}$, or 0.5).

   H_1 states that the coin is **not fair** (that is, $p \neq \frac{1}{2}$, or 0.5).

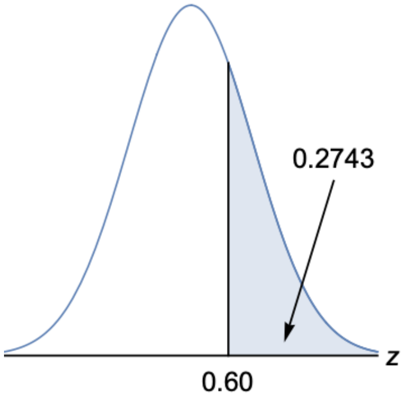
   We flip the coin 500 times and observe 261 heads (52.2%). A 95% confidence interval (CI) for p is $(0.478,\ 0.566)$. Based on this CI, do we **reject** or **not reject** H_0 in a two-tailed hypothesis test? (3 points)

   The 95% CI **does** contain 0.500 (which corresponds to a fair coin), so we **do not reject** H_0.

2) (28 points). In the 2026 Georgia Senate race, if no one receives a majority of the November vote, there will be a runoff election between the top two finishers. (A majority means more than 50%.) In October 2026, a senator claims that they are supported by a majority of likely voters in Georgia (and will thus be able to avoid a runoff). Test the senator's claim at the 0.05 significance level.
   Warning: Write percents as decimals.

   Let p = the proportion of likely Georgia voters supporting the senator.

   Use these hints about the z distribution:



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

$$H_0\!: p = 0.5, \text{ or } p \le 0.5$$

$$H_1\!: p > 0.5 \text{ (Claim)}$$

$$\alpha = 0.05$$

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

The test is **right-tailed**.

We gather **sample data**. We randomly select 900 likely Georgia voters in a poll. 459 of them support the senator.

- c) Find the **sample proportion** $\hat{p}$. Your answer will be exact; do not round off. (3 points)

$$\hat{p} = \frac{x}{n} = \frac{459}{900} = 0.510$$

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

$$\text{Under } H_0,\ np = (900)(0.5) = 450 \ge 5$$

$$\text{Under } H_0,\ nq = (900)(0.5) = 450 \ge 5$$

$$(\text{Note: } q = 1 - p = 1 - 0.5 = 0.5)$$

- e) Compute the z **test statistic** for our sample. Your answer will be exact; do not round off. Show all work, as in class! (7 points)

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.510 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{900}}} = 0.60$$

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

$$\text{Right-tailed } P\text{-value} \approx 0.2743$$

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

Do not reject H_0. (Note that 0.2743 > 0.05, so the *P*-value is **not low**.)

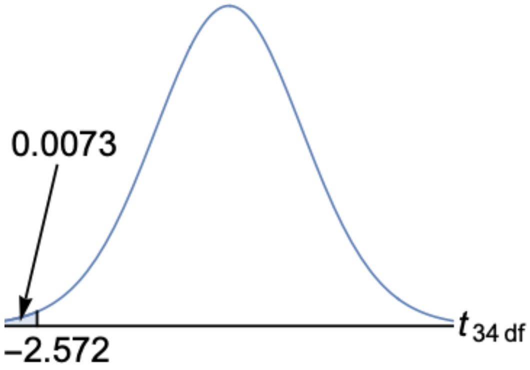
- h) Write our **conclusion** relative to the claim, as in class. (5 points)

  There is **insufficient** evidence **for** the claim that the senator is supported by a majority of likely voters in Georgia.

3) (23 points). McWendy's claims that its Healthy Burgers have on average less than 5.00 grams of saturated fat. Test this claim at the 0.01 significance level.

   Let μ = the population mean amount of saturated fat in McWendy's Healthy Burgers (in grams).

   Use these hints about the *t* distribution on 34 degrees of freedom (df):



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

  H_0: $\mu = 5.00$ grams, or $\mu \geq 5.00$ grams

  H_1: $\mu < 5.00$ grams (Claim)

  $\alpha = 0.01$

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

  The test is **left-tailed**.

We gather **sample data**. We randomly select 35 McWendy's Healthy Burgers. On average, the sample has 4.90 grams of saturated fat. The sample standard deviation (SD) is 0.23 grams of saturated fat.

- 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)

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{4.90 - 5.00}{\frac{0.23}{\sqrt{35}}} \approx -2.572$$

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

  Left−tailed P−value $\approx$ 0.0073

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

  Reject H_0. (Note that 0.0073 < 0.01, so the *P*-value is **low**.)

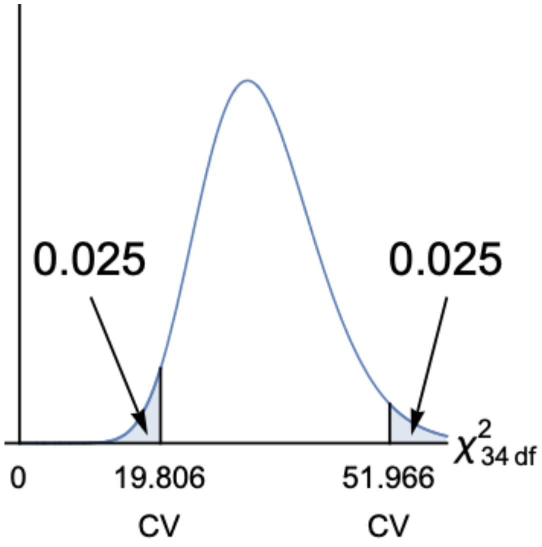
- f) Write our **conclusion** relative to the claim, as in class. (5 points)

  There is **sufficient** evidence **for** the claim that the average amount of saturated fat in McWendy's Healthy Burgers is less than 5.00 grams.

4) (20 points). The amounts of saturated fat in McWendy's Classic Burgers have a population standard deviation (SD) of 0.20 grams. Test the claim that the amounts of saturated fat in McWendy's Healthy Burgers (in grams) have the same population standard deviation (SD) as those in McWendy's Classic Burgers at the 0.05 significance level. Use **the traditional (classical) method** of hypothesis testing. Assume that the amounts of saturated fat in McWendy's Healthy Burgers are approximately normally distributed.

   Let σ = the population standard deviation (SD) of the amounts of saturated fat in McWendy's Healthy Burgers (in grams).

   Use these hints about the χ^2 distribution on 34 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)

  H_0: $\sigma = 0.20$ grams (Claim)

  H_1: $\sigma \neq 0.20$ grams

  $\alpha = 0.05$

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

  The test is **two-tailed**.

We gather **sample data**. We randomly select 35 McWendy's Healthy Burgers. The sample standard deviation (SD) is 0.23 grams of saturated fat. (We collect the same sample data for both Problem 2) and Problem 3).)

- c) Compute the χ^2 **test statistic** for our sample and round it off to three decimal places. Show all work, as in class! (6 points)

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(35-1)(0.23)^2}{(0.20)^2} \approx 44.965$$

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

  Do **not** reject H_0. (Note that 44.965 is **not** in the critical region, which corresponds to the shaded region in the figure.)

- e) Write our **conclusion** relative to the claim, as in class. (5 points)

  There is **insufficient** evidence **against** the claim that the amounts of saturated fat (in grams) in McWendy's Healthy Burgers have the same population standard deviation (SD) as those in McWendy's Classic Burgers.

5) (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

6) (21 points). According to an April 2022 AP poll, 56% of American adults favored mask requirements for public transportation, 24% opposed them, and 20% were undecided. Let's assume that these proportions were correct. Test the claim that the distribution of opinions on mask requirements for public transportation in April 2022 was the same among Fredonian adults at the 0.05 significance level.

- a) Write the **setup** for the hypothesis test, as in class. 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 Fredonian adults who favored mask requirements for public transportation in April 2022.

  Let p_O = the proportion of Fredonian adults who opposed mask requirements for public transportation in April 2022.

  Let p_U = the proportion of Fredonian adults who were undecided about mask requirements for public transportation in April 2022.

(cont.)

$$H_0: \left\{ \begin{matrix} p_F = 0.56 \\ p_O = 0.24 \\ p_U = 0.20 \end{matrix} \right\} \text{(Claim)}$$

H_1: The distribution is different.

$\alpha = 0.05$

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

  The test is **right-tailed**.

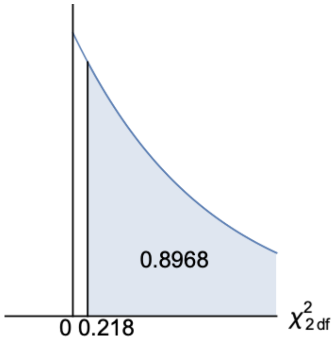
We gathered **sample data**. We randomly sampled 1085 Fredonian adults. (This was the sample size used in the AP poll.) Among them, 600 favored mask requirements for public transportation in April 2022, 265 opposed them, and 220 were undecided.

- 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	600	$np_F = (1085)(0.56)$ $= 607.6$
Oppose	265	$np_O = (1085)(0.24)$ $= 260.4$
Undecided	220	$np_O = (1085)(0.20)$ $= 217.0$

- 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):



Do **not** reject H_0. (Note that 0.8968 > 0.05, so the *P*-value is **not** low.)

- e) Write our **conclusion** relative to the claim, as in class. (5 points)

  There is **insufficient** evidence **against** the claim that the distribution of opinions on mask requirements for public transportation in April 2022 was the same among Fredonian adults.

7) (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, as in class. The setup will include H_0, H_1, identifying which is the claim, and the significance level. (5 points)

  H_0: Use of the vaping product is independent of incidents of lung cancer.

  H_1: Use of the vaping product is dependent on incidents of lung cancer. (Claim)

  $\alpha = 0.05$

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

  The test is **right-tailed**.

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:

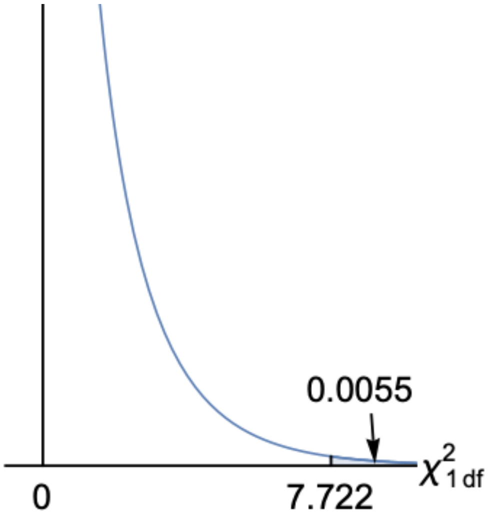
	Have lung cancer	Do not have lung cancer	**← Presence of lung cancer**
Do not use the product	200	2600	**2800**
Use the product	25	175	**200**
↑ Usage of product	**225**	**2775**	**3000**

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):



Reject H_0. (Note that 0.0055 < 0.05, so the *P*-value is **low**.)

- d) Write our **conclusion** relative to the claim, as in class. (5 points)

There is **sufficient** evidence **for** the claim that there is statistical dependence between use of the vaping product and incidents of lung cancer.

8) (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 8 units in y. (2 points)

8 is the slope of the line.

9) (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? Assume that the linear correlation coefficient between the Midterm 1 and Midterm 2 scores is $r = 0.5$. 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.

10) (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%

The coefficient of determination is: $r^2 = (0.9)^2 = 0.81 = 81\%$.

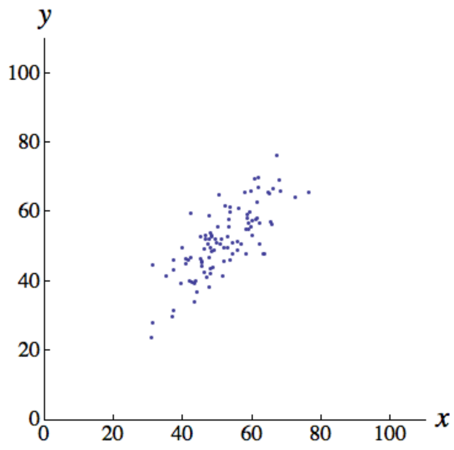
11) (4 points). (Matching)

For each variable, the average is 50 and the standard deviation is 10.

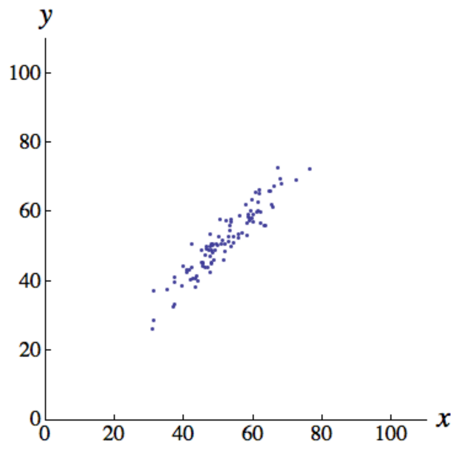
For one of the graphs below, $r = -0.90$.
For one of the graphs below, $r = 0.00$.
For one of the graphs below, $r = 0.80$.
For one of the graphs below, $r = 0.95$.

Fill in the blanks:

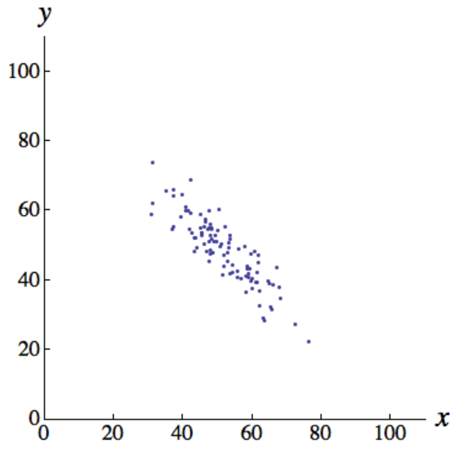
a) r for the graph below is $\underline{\ 0.80\ }$.



b) r for the graph below is $\underline{\ 0.95\ }$.



c) r for the graph below is $\underline{\ -0.90\ }$.



d) r for the graph below is $\underline{\ 0.00\ }$.