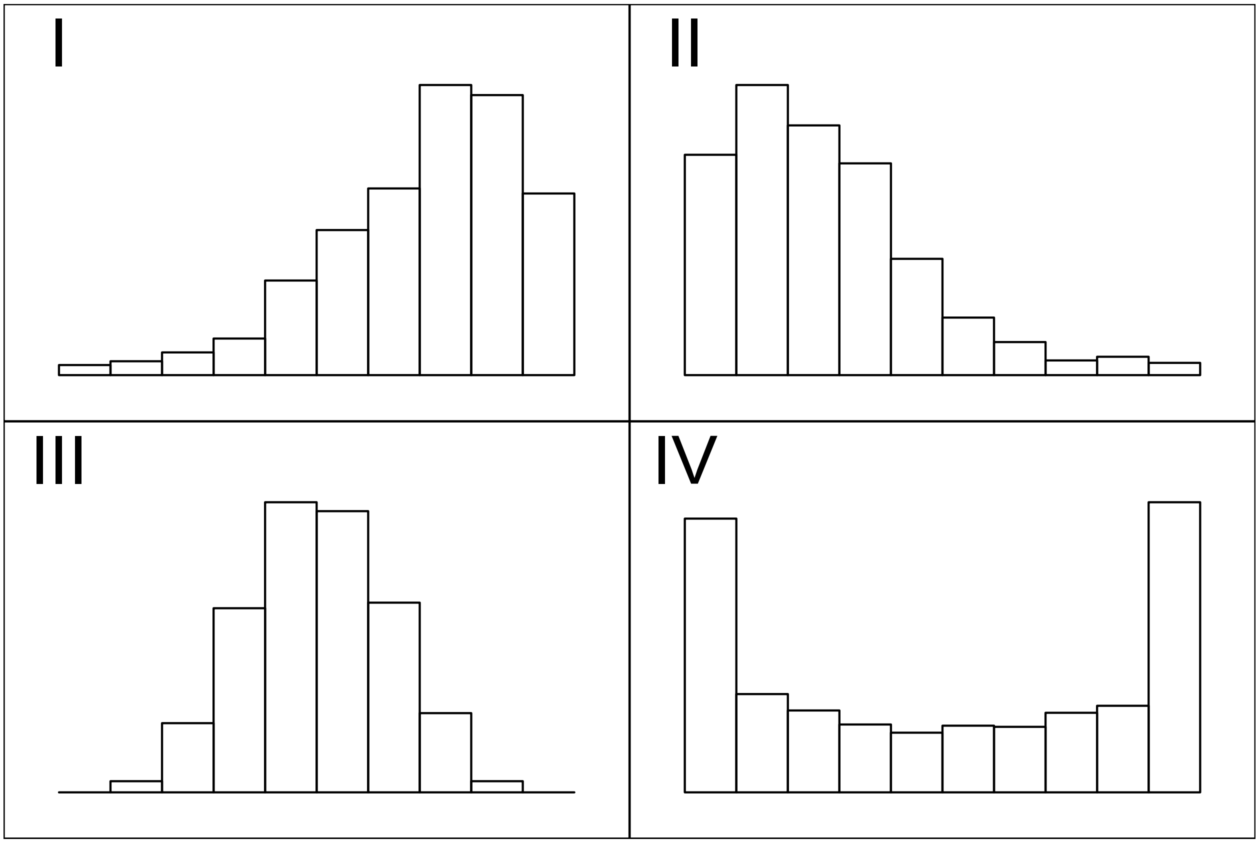
MAT-181 FINAL TAKE-HOME EXAM

This exam is to be taken without discussion or correspondance with any human. Please show work!

question	available points	earned points
1	10	
2	15	
3	10	
4	10	
5	10	
6	10	
7	15	
8	20	
EC	5	
EC	5	
Total	100	

1. (10 Points)



For each description below, choose which histogram best fits (I, II, III, or IV). Each histogram should be used once.

(a) The distribution of hours spent per week reading by adults. In this distribution, many people do not read much, and a similar number of people read a lot.

(b) The distribution of lengths of newborn babies

(c) The distribution of test scores on a very difficult exam, in which most students have poor to average scores, but a few did quite well.

(d) The distribution of ages at a skilled nursing facility, where most of the patients are elderly but a few are quite young.

Solution:

(a) IV

(b) III

(c) II

(d) I

2. (15 Points)

In a deck of strange cards, there are 626 cards. Each card has an image and a color. The amounts are shown in the table below.

	blue	pink	red	teal	yellow	Total
bike	20	38	10	37	47	152
flower	31	33	14	17	16	111
kite	46	48	34	12	32	172
wheel	24	49	50	39	29	191
Total	121	168	108	105	124	626

(a) What is the probability a random card is both a wheel and blue?

(b) What is the probability a random card is a flower given it is pink?

(c) Is a flower or a kite more likely to be pink?

(d) What is the probability a random card is red?

(e) What is the probability a random card is a wheel?

(f) What is the probability a random card is pink given it is a bike?

(g) What is the probability a random card is either a kite or teal (or both)?

Solution:

(a) P(wheel and blue) = 0.0383

(b) P(flower given pink) = 0.196

(c) P(pink given flower) = 0.297 and P(pink given kite) = 0.279, so a flower is more likely to be pink than a kite is.

(d) P(red) = 0.173

(e) P(wheel) = 0.305

(f) P(pink given bike) = 0.25

(g) P(kite or teal) = 0.423

3. (10 points)

   A farm produces 4 types of fruit: *A*, *B*, *C*, and *D*. The fruits' masses follow normal distributions, with parameters dependent on the type of fruit.

Type of fruit	Mean mass (g)	Standard deviation of mass (g)
A	84	5
B	104	8
C	139	12
D	134	11

   One specimen of each type is weighed. The results are shown below.

Type of fruit	Mass of specimen (g)
A	91.15
B	101.8
C	148.8
D	130.3

   Which specimen is the most unusually large (relative to others of its type)?

Solution: We compare the z-scores. The largest z-score corresponds to the specimen that is most unusually large.

Type of fruit	formula	z-score
A	$z = \frac{91.15-84}{5}$	1.43
B	$z = \frac{101.8-104}{8}$	-0.28
C	$z = \frac{148.8-139}{12}$	0.82
D	$z = \frac{130.3-134}{11}$	-0.34

Thus, the specimen of type A is the most unusually large.

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 44.3 millimeters and a standard deviation of 2.2 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 40.3 and 44 millimeters?

Solution:

$$\mu = 44.3$$
$$\sigma = 2.2$$
$$x_1 = 40.3$$
$$x_2 = 44$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{40.3 - 44.3}{2.2} = -1.82$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{44 - 44.3}{2.2} = -0.14$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.4443 - 0.0344 = 0.4099$$

5. (10 points)

   A species of duck is known to have a mean weight of 203.6 grams and a standard deviation of 56 grams. A researcher plans to measure the weights of 64 of these ducks sampled randomly. What is the probability the **sample mean** will be between 210.6 and 214.1 grams?

Solution:

$$n = 64$$
$$\mu = 203.6$$
$$\sigma = 56$$
$$SE = \frac{56}{\sqrt{64}} = 7$$
$$x_1 = 210.6$$
$$x_2 = 214.1$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{210.6 - 203.6}{7} = 1$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{214.1 - 203.6}{7} = 1.5$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.9332 - 0.8413 = 0.0919$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Dendroica striata*. She randomly samples 17 adults of *Dendroica striata*, resulting in a sample mean of 15.37 grams and a sample standard deviation of 4.09 grams. Determine a 95% confidence interval of the true population mean.

Solution: We are given the sample size, sample mean, sample standard deviation, and confidence level.

$$\begin{aligned} n &= 17 \\ \bar{x} &= 15.37 \\ s &= 4.09 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 17 - 1 \\ &= 16 \end{aligned}$$

Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 16$.

$$t^{\star} = 2.12$$

Use the formula for bounds (mean, σ unknown).

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} \\ &= 15.37 - 2.12 \times \frac{4.09}{\sqrt{17}} \\ &= 13.3 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 15.37 + 2.12 \times \frac{4.09}{\sqrt{17}} \\ &= 17.5 \end{aligned}$$

We are 95% confident that the population mean is between 13.3 and 17.5 grams.

$$CI = (13.3,\ 17.5)$$

7. (15 points)

   A student is taking a multiple choice test with 800 questions. Each question has 5 choices. You want to detect whether the student does significantly better than random guessing, so you decide to run a hypothesis test with a significance level of 0.05.

   Then, the student takes the test and gets 175 questions correct.

   (a) What kind of hypothesis test is appropriate?

   (b) State the hypotheses.

   (c) Determine the test statistic (z or t), draw a sketch, and determine the p-value.

   (d) Decide whether we reject or retain the null hypothesis.

   (e) Did the student do significantly better than random guessing?

Solution: This is a right-tail (one-tail) proportion test because we only care whether the student does better than random.

Determine the null population proportion.

$$p_0 = \frac{1}{5} = 0.2$$

State the hypotheses.

$$H_0 \quad \text{claims} \quad p = 0.2$$

$$H_A \quad \text{claims} \quad p > 0.2$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.2(1-0.2)}{800}} = 0.0141$$

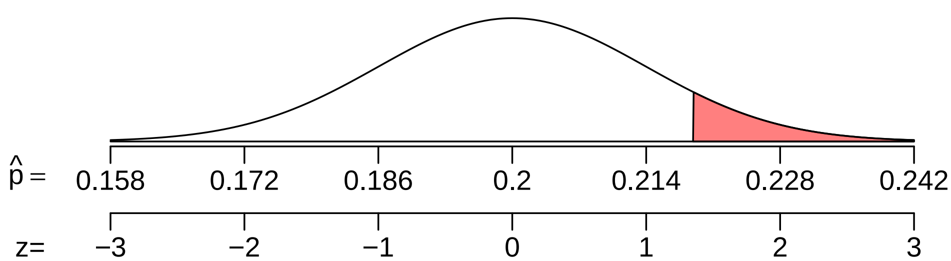
Determine the sample proportion.

$$\hat{p} = \frac{175}{800} = 0.219$$

Determine a z score. For simplicity, we ignore the continuity correction.

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.219 - 0.2}{0.0141} = 1.35$$

Make a sketch of the null's sampling distribution. The p-value is a right area.



To determine that right area, we use the z table.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.219) \\ &= P(Z > 1.35) \\ &= 1 - P(Z < 1.35) \\ &= 0.0885 \end{aligned}$$

Compare p-value to α (which is 0.05).

$$p\text{-value} > \alpha$$

Make the conclusion: we retain the null hypothesis.

We think the student might just be guessing.

(a) Right tail (one-tail) proportion test

(b) Hypotheses: H_0 claims $p = 0.2$ and H_A claims $p > 0.2$.

(c) The p-value is 0.0885

(d) We retain the null hypothesis.

(e) We think the student might just be guessing.

8. (20 points) [Note: this question uses 2 pages.]

   You have collected the following data:

x	y	xy
9.6	250	
7.6	480	
9.8	410	
7.6	870	
4.6	610	
9.5	450	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

   (a) Complete the table.

   (b) Calculate the correlation coefficient (r) using the formula below.

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

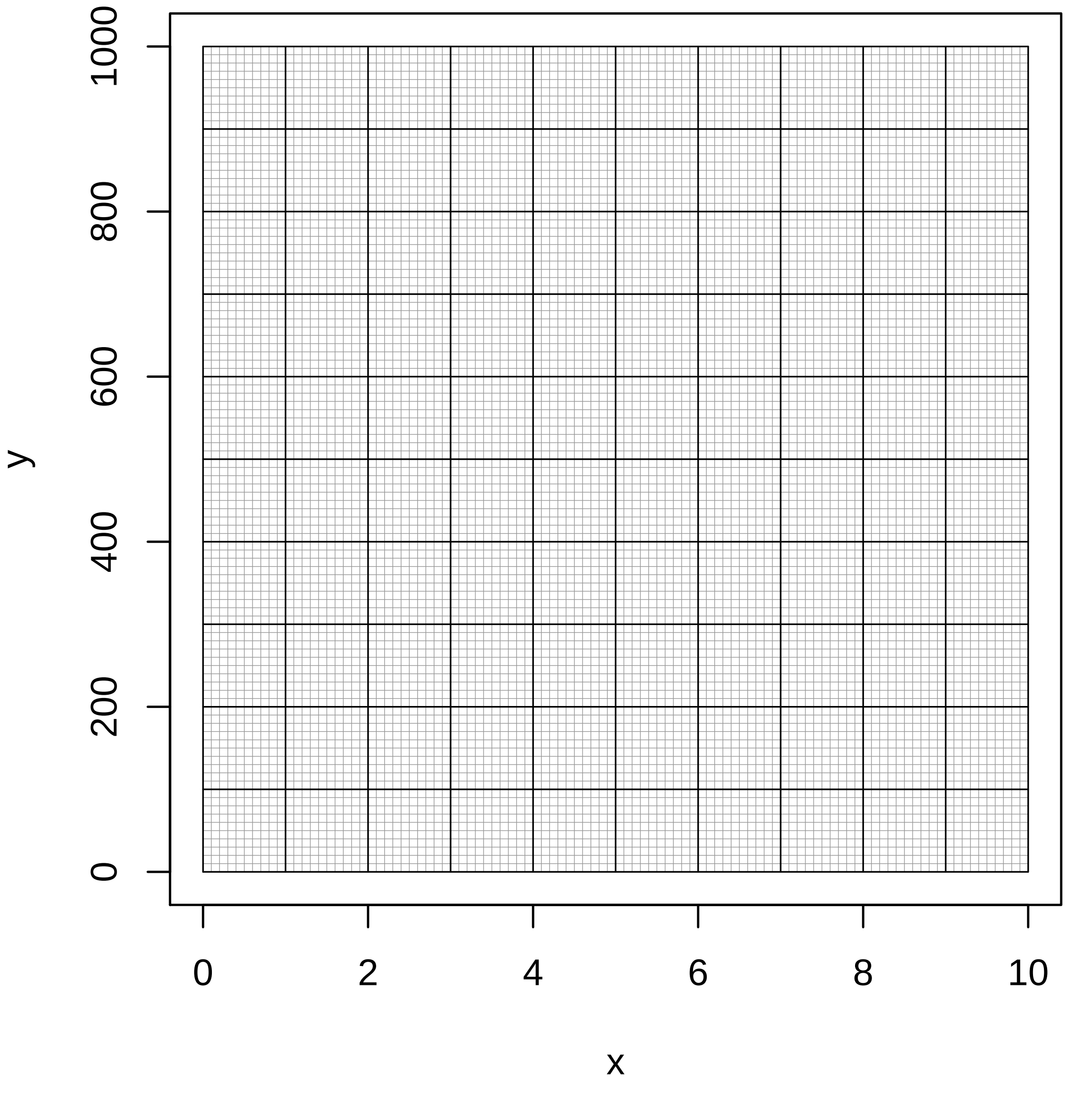
   (c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

$$b = r\frac{s_y}{s_x}$$

$$a = \bar{y} - b\bar{x}$$

   (d) Write the equation of the regression line (using the calculated values of a and b.)

(e) Please plot the data and a corresponding regression line.



Solution: Remember the formula for the correlation coefficient.

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

We calculate the necessary values.

x	y	xy
9.6	250	2400
7.6	480	3648
9.8	410	4018
7.6	870	6612
4.6	610	2806
9.5	450	4275
$\sum x = 48.7$	$\sum y = 3070$	$\sum x_i y_i = 23759$
$\bar{x} = 8.117$	$\bar{y} = 511.7$	
$s_x = 1.992$	$s_y = 210.6$	

$$r = \frac{23759 - (6)(8.117)(511.7)}{(6-1)(1.992)(210.6)} = -0.554$$

If you didn't round any of the steps up to here, you'd get an exact value which is pretty close to our value.

$$r_{\text{exact}} = -0.5526082$$

The regression line has the form

$$y = a + bx$$

So, a is the y-intercept and b is the slope. We have formulas to determine them:

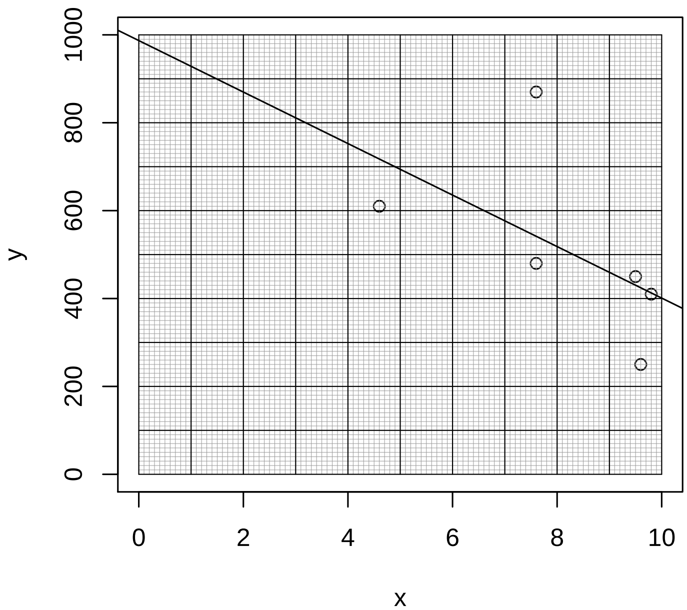
$$b = r\frac{s_y}{s_x} = -0.554 \cdot \frac{210.6}{1.992} = -58.6$$

$$a = \bar{y} - b\bar{x} = 512 - (-58.6)(8.12) = 987$$

Our regression line:

$$y = 987 + (-58.6)x$$

Make a plot.


y
x
0
200
400
600
800
1000
0
2
4
6
8
10

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.71$. If 128 trials occur, what is the probability of getting more than 79 but less than 98 successes?

   In other words, let $X \sim \text{Bin}(n = 128,\ p = 0.71)$ and find $P(79 < X < 98)$.

   Use a normal approximation along with the continuity correction.

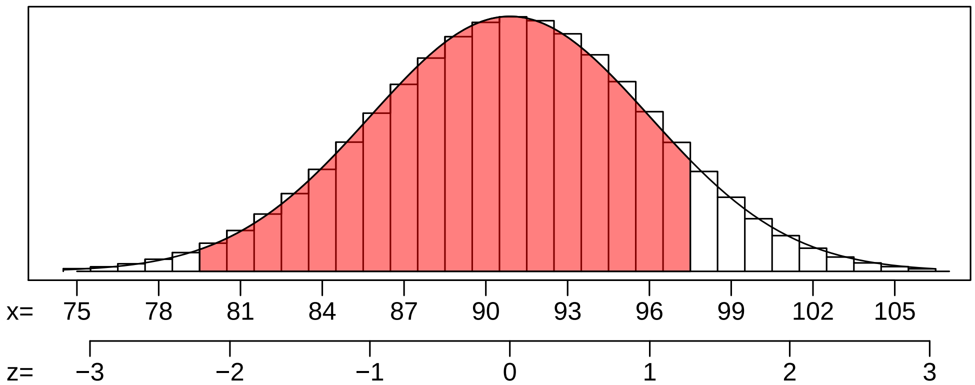
Solution: Find the mean.

$$\mu = np = (128)(0.71) = 90.88$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(128)(0.71)(1-0.71)} = 5.1337$$

Make a sketch, specifically try to picture whether you need to add or subtract 0.5 for the continuity correction.



Find the z scores.

$$z_1 = \frac{79.5 - 90.88}{5.1337} = -2.22$$

$$z_2 = \frac{97.5 - 90.88}{5.1337} = 1.29$$

Find the percentiles (from z-table).

$$\ell_1 = 0.0132$$

$$\ell_2 = 0.9015$$

Calculate the probability.

$$P(79 < X < 98) = 0.9015 - 0.0132 = 0.888$$

10. (Extra credit: 5 points)

A null hypothesis claims a population has a mean $\mu = 250$. You decide to run two-tail test on a sample of size $n = 12$ using a significance level $\alpha = 0.02$.

You then collect the sample:

299.2	239	261.7	239.3	303.6
302.7	286.2	271.2	233	267.5
288.8	250			

(a) Determine the p-value.

(b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 250$$

$$H_A \quad \text{claims} \quad \mu \neq 250$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 270.183$$

$$s = 25.964$$

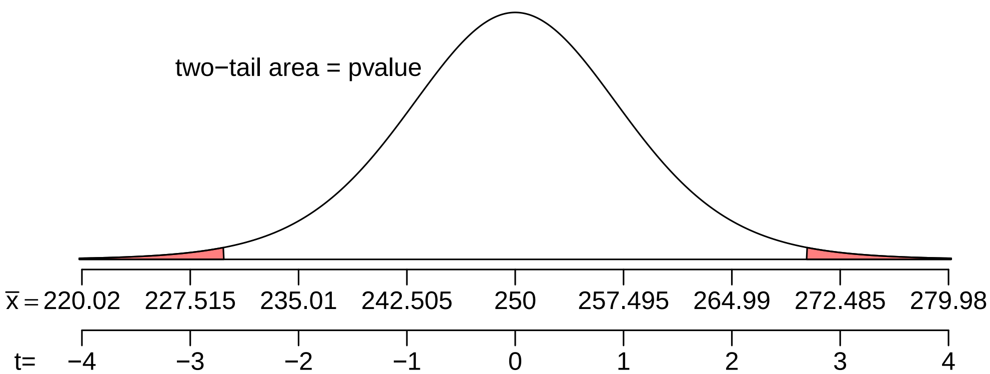
Determine the degrees of freedom.

$$df = 12 - 1 = 11$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{25.964}{\sqrt{12}} = 7.495$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{270.183 - 250}{7.495} = 2.69$$

Find the p-value.

$$p\text{-value} = P(|T| > 2.69)$$

We can't get an exact value with our table, but we can determine an interval that contains the p-value. (Look at row with $df = 11$.)

$$P(|T| > 2.72) = 0.02$$

$$P(|T| > 2.33) = 0.04$$

Basically, because t is between 2.72 and 2.33, we know the p-value is between 0.02 and 0.04.

$$0.02 < p\text{-value} < 0.04$$

Compare the p-value and the significance level ($\alpha = 0.02$).

$$p\text{-value} > \alpha$$

No, we do not reject the null hypothesis.

(a) $0.02 < p\text{-value} < 0.04$

(b) No, we do not reject the null hypothesis.