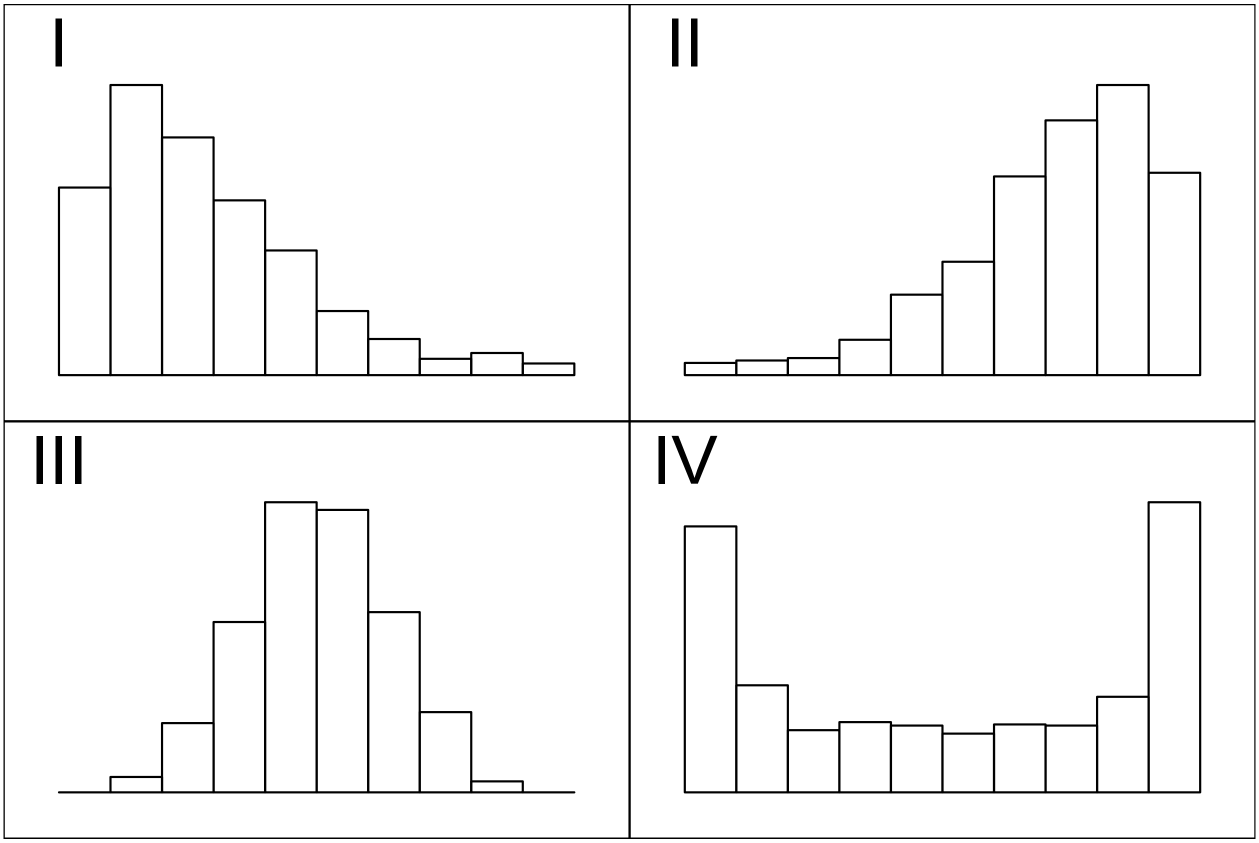
MAT-181 FINAL TAKE-HOME EXAM

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 that students studied for an exam when about half of students studied a lot and a similar number of students studied very little.

(b) The distribution of quiz scores on an easy quiz. Most students did very well, but a few did poorly.

(c) The distribution of annual income for school employees where a high percentage of employees are entry-level teachers and only a few are high-paid administrators.

(d) The distribution of weights of newborn babies

Solution:

(a) IV

(b) II

(c) I

(d) III

2. (15 Points)

In a deck of strange cards, there are 486 cards. Each card has an image and a color. The amounts are shown in the table below.

	blue	orange	pink	red	Total
bike	29	22	34	21	106
cat	35	18	48	45	146
dog	19	10	30	49	108
quilt	16	40	42	28	126
Total	99	90	154	143	486

(a) What is the probability a random card is a cat?

(b) What is the probability a random card is a dog given it is orange?

(c) Is a bike or a cat more likely to be red?

(d) What is the probability a random card is either a cat or red (or both)?

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

(f) What is the probability a random card is both a cat and red?

(g) What is the probability a random card is red given it is a bike?

Solution:

(a) $P(\text{cat}) = 0.3$

(b) $P(\text{dog given orange}) = 0.111$

(c) $P(\text{red given bike}) = 0.198$ and $P(\text{red given cat}) = 0.308$, so a cat is more likely to be red than a bike is.

(d) $P(\text{cat or red}) = 0.502$

(e) $P(\text{orange}) = 0.185$

(f) $P(\text{cat and red}) = 0.0926$

(g) $P(\text{red given bike}) = 0.198$

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	73	13
B	60	7
C	114	6
D	94	12

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

Type of fruit	Mass of specimen (g)
A	61.43
B	56.01
C	112.6
D	104.1

Which specimen is the most unusually far (in either direction) from average (relative to others of its type)?

Solution: We compare the absolute z-scores. The largest absolute z-score corresponds to the specimen that is most unusually far from average.

Type of fruit	formula	absolute z-score
A	$z = \frac{\lvert 61.43-73\rvert}{13}$	0.89
B	$z = \frac{\lvert 56.01-60\rvert}{7}$	0.57
C	$z = \frac{\lvert 112.6-114\rvert}{6}$	0.24
D	$z = \frac{\lvert 104.1-94\rvert}{12}$	0.84

Thus, the specimen of type A is the most unusually far from average.

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 122.8 millimeters and a standard deviation of 5.7 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 125.4 and 127.2 millimeters?

Solution:

$$\mu = 122.8$$
$$\sigma = 5.7$$
$$x_1 = 125.4$$
$$x_2 = 127.2$$

$$\begin{aligned} z_1 &= \frac{x_1 - \mu}{\sigma} = \frac{125.4 - 122.8}{5.7} = 0.46 \\ z_2 &= \frac{x_2 - \mu}{\sigma} = \frac{127.2 - 122.8}{5.7} = 0.77 \\ P(x_1 < X < x_2) &= P(z_1 < Z < z_2) = 0.7794 - 0.6772 = 0.1022 \end{aligned}$$

5. (10 points)

   A species of duck is known to have a mean weight of 215.3 grams and a standard deviation of 20 grams. A researcher plans to measure the weights of 100 of these ducks sampled randomly. What is the probability the **sample mean** will be between 215.3 and 219.3 grams?

Solution:

$$
\begin{aligned}
n &= 100 \\
\mu &= 215.3 \\
\sigma &= 20 \\
SE &= \frac{20}{\sqrt{100}} = 2 \\
x_1 &= 215.3 \\
x_2 &= 219.3
\end{aligned}
$$

$$
\begin{aligned}
z_1 &= \frac{x_1 - \mu}{SE} = \frac{215.3 - 215.3}{2} = 0 \\
z_2 &= \frac{x_2 - \mu}{SE} = \frac{219.3 - 215.3}{2} = 2 \\
P(x_1 < \overline{X} < x_2) &= P(z_1 < Z < z_2) = 0.9772 - 0.5 = 0.4772
\end{aligned}
$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Vireo philadelphicus*. She randomly samples 20 adults of *Vireo philadelphicus*, resulting in a sample mean of 13.5 grams and a sample standard deviation of 1.52 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.

$$\begin{aligned} n &= 20 \\ \bar{x} &= 13.5 \\ s &= 1.52 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 20 - 1 \\ &= 19 \end{aligned}$$

Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 19$.

$$t^{\star} = 2.09$$

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

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} \\ &= 13.5 - 2.09 \times \frac{1.52}{\sqrt{20}} \\ &= 12.8 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 13.5 + 2.09 \times \frac{1.52}{\sqrt{20}} \\ &= 14.2 \end{aligned}$$

We are 95% confident that the population mean is between 12.8 and 14.2 grams.

$$CI = (12.8,\ 14.2)$$

7. (15 points)

   A student is taking a multiple choice test with 300 questions. Each question has 4 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 89 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.

$$p_0 = \frac{1}{4} = 0.25$$

State the hypotheses.

$$H_0 \quad \text{claims} \quad p = 0.25$$
$$H_A \quad \text{claims} \quad p > 0.25$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.25(1-0.25)}{300}} = 0.025$$

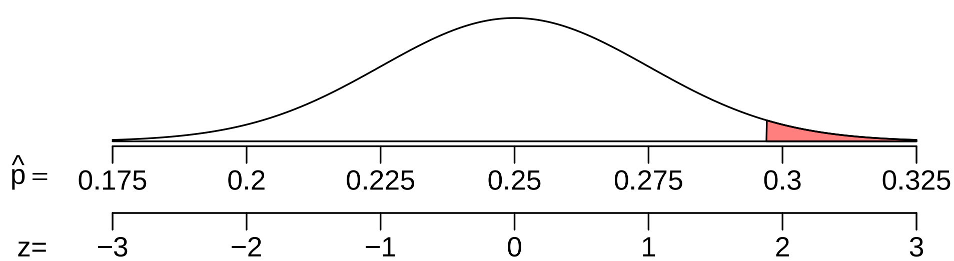
Determine the sample proportion.

$$\hat{p} = \frac{89}{300} = 0.297$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.297 - 0.25}{0.025} = 1.88$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.297) \\ &= P(Z > 1.88) \\ &= 1 - P(Z < 1.88) \\ &= 0.0301 \end{aligned}$$

Compare p-value to α (which is 0.05).

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

Make the conclusion: we reject the null hypothesis.

We think the student did better than random guessing typically allows.

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

(b) Hypotheses: H_0 claims $p = 0.25$ and H_A claims $p > 0.25$.

(c) The p-value is 0.0301

(d) We reject the null hypothesis.

(e) We think the student did better than random guessing typically allows.

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

   You have collected the following data:

x	y	xy
7.9	60	
5	57	
2	91	
7.2	54	
5.5	41	
3.5	72	
4.5	52	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

   (a) Complete the table.

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

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

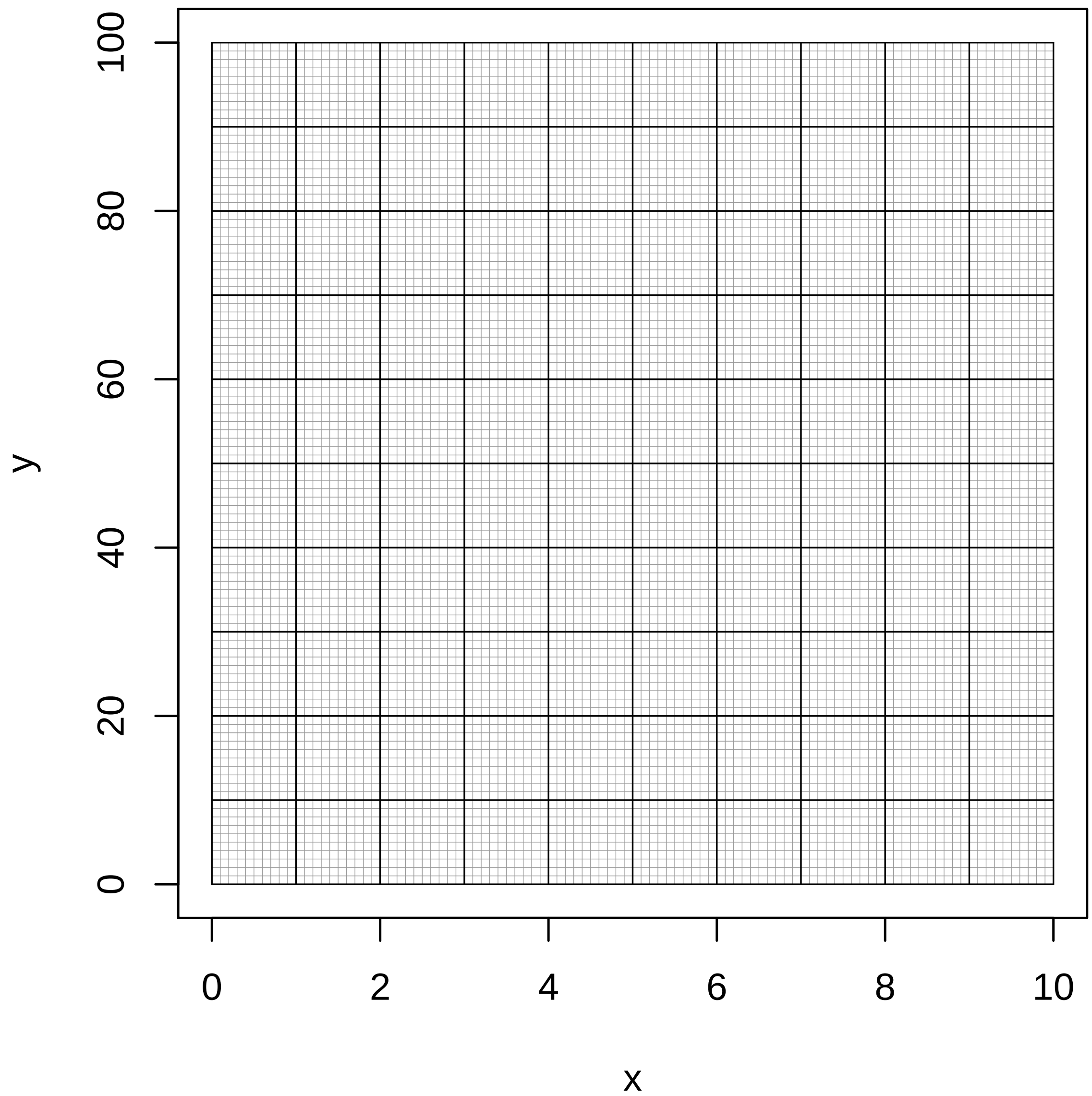
   (c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

   (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
7.9	60	474
5	57	285
2	91	182
7.2	54	388.8
5.5	41	225.5
3.5	72	252
4.5	52	234
$\sum x = 35.6$	$\sum y = 427$	$\sum x_i y_i = 2041.3$
$\bar{x} = 5.086$	$\bar{y} = 61$	
$s_x = 2.039$	$s_y = 16.17$	

$$r = \frac{2041.3 - (7)(5.086)(61)}{(7-1)(2.039)(16.17)} = -0.659$$

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.

$$r_{\text{exact}} = -0.6587923$$

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:

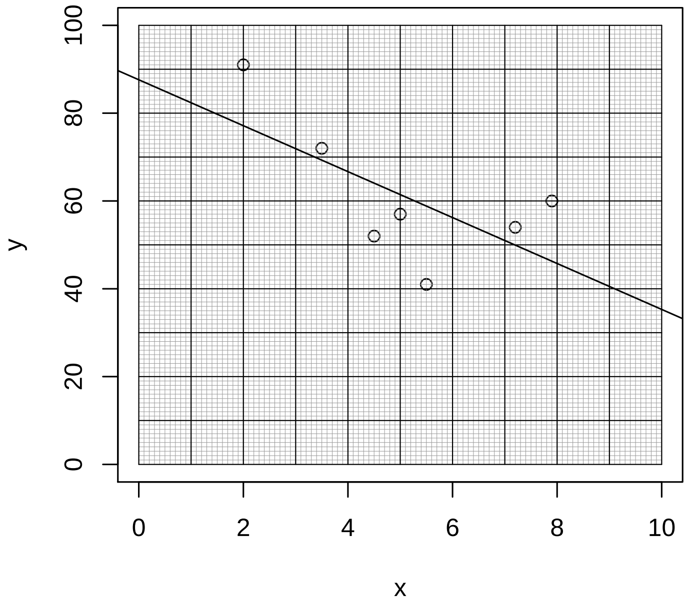
$$b = r\frac{s_y}{s_x} = -0.659 \cdot \frac{16.17}{2.039} = -5.23$$

$$a = \bar{y} - b\bar{x} = 61 - (-5.23)(5.09) = 87.6$$

Our regression line:

$$y = 87.6 + (-5.23)x$$

Make a plot.



0
20
40
60
80
100
y
0
2
4
6
8
10
x

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.81$. If 195 trials occur, what is the probability of getting more than 145 but at most 166 successes?

   In other words, let $X \sim \text{Bin}(n = 195,\ p = 0.81)$ and find $P(145 < X \leq 166)$.

   Use a normal approximation along with the continuity correction.

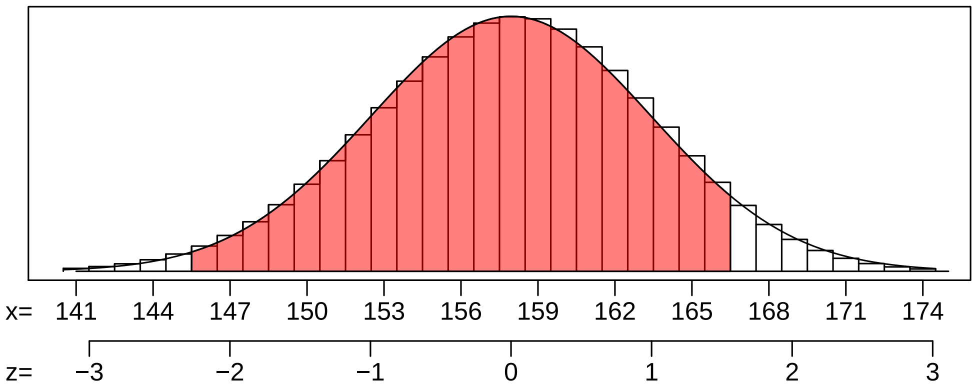
Solution: Find the mean.

$$\mu = np = (195)(0.81) = 157.95$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(195)(0.81)(1-0.81)} = 5.4782$$

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



Find the z scores.

$$z_1 = \frac{145.5 - 157.95}{5.4782} = -2.27$$

$$z_2 = \frac{166.5 - 157.95}{5.4782} = 1.56$$

Find the percentiles (from z-table).

$$\ell_1 = 0.0116$$

$$\ell_2 = 0.9406$$

Calculate the probability.

$$P(145 < X \leq 166) = 0.9406 - 0.0116 = 0.929$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 120$. You decide to run two-tail test on a sample of size $n = 8$ using a significance level $\alpha = 0.02$.

    You then collect the sample:

121.1	127.3	123.1	118.2	124.1
123	121.4	123.3		

    (a) Determine the p-value.

    (b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \text{ claims } \mu = 120$$
$$H_A \text{ claims } \mu \neq 120$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 122.688$$
$$s = 2.625$$

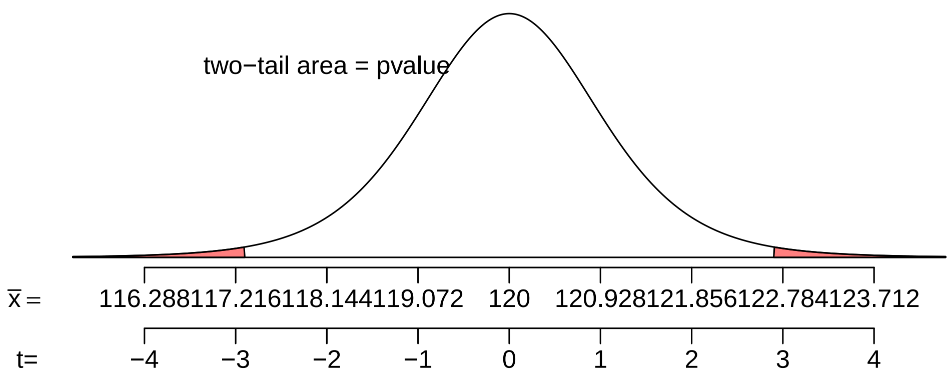
Determine the degrees of freedom.

$$df = 8 - 1 = 7$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.625}{\sqrt{8}} = 0.928$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{122.688 - 120}{0.928} = 2.9$$

Find the p-value.

$$p\text{-value} = P(|T| > 2.9)$$

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 $df = 7$.)

$$P(|T| > 3) = 0.02$$
$$P(|T| > 2.52) = 0.04$$

Basically, because t is between 3 and 2.52, 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.