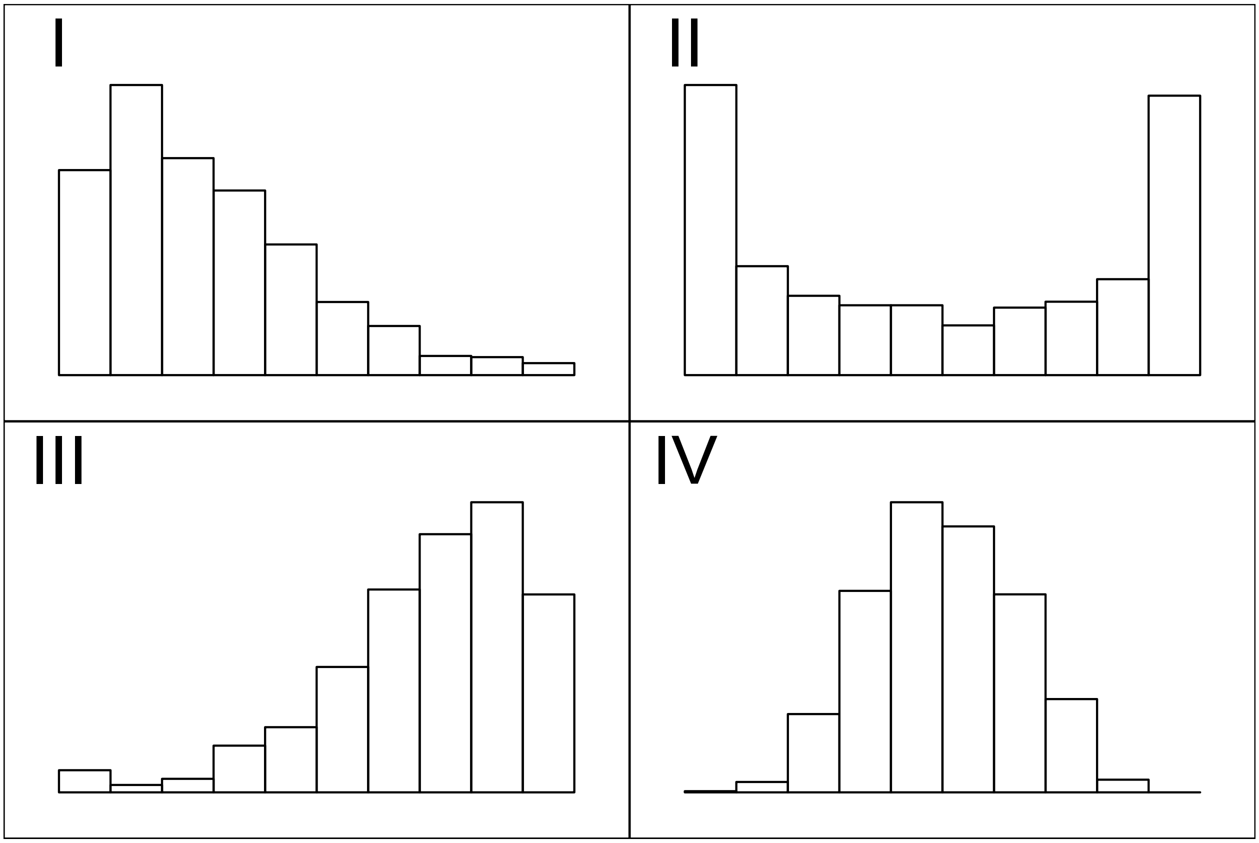
MAT-181 FINAL TAKE-HOME EXAM

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

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

(c) The distribution of heights of adult women

(d) 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.

Solution:

(a) I

(b) III

(c) IV

(d) II

2. (15 Points)

In a deck of strange cards, there are 374 cards. Each card has an image and a color. The amounts are shown in the table below.

	blue	gray	red	Total
horn	39	20	14	73
jigsaw	11	24	30	65
needle	19	34	36	89
pig	15	43	13	71
shovel	12	26	38	76
Total	96	147	131	374

(a) What is the probability a random card is either a jigsaw or gray (or both)?

(b) What is the probability a random card is both a shovel and gray?

(c) Is a horn or a shovel more likely to be red?

(d) What is the probability a random card is blue given it is a jigsaw?

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

(f) What is the probability a random card is a jigsaw given it is red?

(g) What is the probability a random card is gray?

Solution:

(a) P(jigsaw or gray) = 0.503

(b) P(shovel and gray) = 0.0695

(c) P(red given horn) = 0.192 and P(red given shovel) = 0.5, so a shovel is more likely to be red than a horn is.

(d) P(blue given jigsaw) = 0.169

(e) P(pig) = 0.19

(f) P(jigsaw given red) = 0.229

(g) P(gray) = 0.393

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	92	9
B	96	11
C	103	6
D	77	4

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

Type of fruit	Mass of specimen (g)
A	95.87
B	105.6
C	107
D	78.6

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	$z = \frac{\lvert 95.87-92 \rvert}{9}$	0.43
B	$z = \frac{\lvert 105.6-96 \rvert}{11}$	0.87
C	$z = \frac{\lvert 107-103 \rvert}{6}$	0.67
D	$z = \frac{\lvert 78.6-77 \rvert}{4}$	0.4

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

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 161.5 millimeters and a standard deviation of 3.7 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 162.9 and 166.6 millimeters?

Solution:

$$\mu = 161.5$$
$$\sigma = 3.7$$
$$x_1 = 162.9$$
$$x_2 = 166.6$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{162.9 - 161.5}{3.7} = 0.38$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{166.6 - 161.5}{3.7} = 1.38$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.9162 - 0.648 = 0.2682$$

5. (10 points)

   A species of duck is known to have a mean weight of 200.7 grams and a standard deviation of 67.5 grams. A researcher plans to measure the weights of 81 of these ducks sampled randomly. What is the probability the **sample mean** will be between 205.2 and 215.7 grams?

Solution:

$$n = 81$$
$$\mu = 200.7$$
$$\sigma = 67.5$$
$$SE = \frac{67.5}{\sqrt{81}} = 7.5$$
$$x_1 = 205.2$$
$$x_2 = 215.7$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{205.2 - 200.7}{7.5} = 0.6$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{215.7 - 200.7}{7.5} = 2$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.9772 - 0.7257 = 0.2515$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Piranga rubra*. She randomly samples 32 adults of *Piranga rubra*, resulting in a sample mean of 36.13 grams and a sample standard deviation of 7.1 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.

$$\begin{aligned} n &= 32 \\ \bar{x} &= 36.13 \\ s &= 7.1 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 32 - 1 \\ &= 31 \end{aligned}$$

Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 31$.

$$t^{\star} = 2.04$$

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

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} \\ &= 36.13 - 2.04 \times \frac{7.1}{\sqrt{32}} \\ &= 33.6 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 36.13 + 2.04 \times \frac{7.1}{\sqrt{32}} \\ &= 38.7 \end{aligned}$$

We are 95% confident that the population mean is between 33.6 and 38.7 grams.

$$CI = (33.6,\ 38.7)$$

7. (15 points)

   A student is taking a multiple choice test with 800 questions. Each question has 2 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 426 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.

$$p_0 = \frac{1}{2} = 0.5$$

State the hypotheses.

$$H_0 \quad \text{claims} \quad p = 0.5$$
$$H_A \quad \text{claims} \quad p > 0.5$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{800}} = 0.0177$$

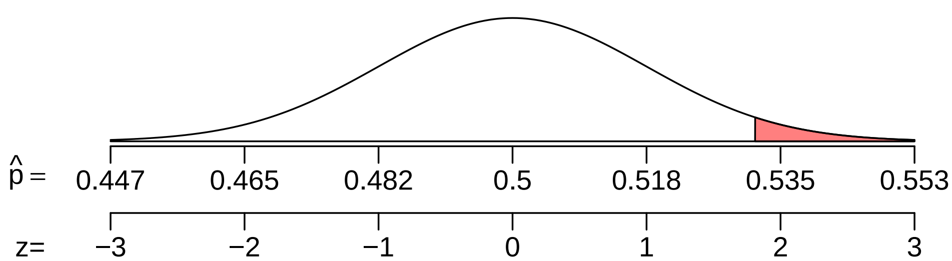
Determine the sample proportion.

$$\hat{p} = \frac{426}{800} = 0.532$$

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

$$z = \frac{\hat{p}-p_0}{\sigma_{\hat{p}}} = \frac{0.532-0.5}{0.0177} = 1.81$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.532) \\ &= P(Z > 1.81) \\ &= 1 - P(Z < 1.81) \\ &= 0.0351 \end{aligned}$$

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 $p = 0.5$ and H_A claims $p > 0.5$.
(c) The p-value is 0.0351
(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
2	71	
6.7	16	
8.7	9.1	
9.6	5.5	
9.3	26	
4.2	22	
7	11	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

   (a) Complete the table.

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

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

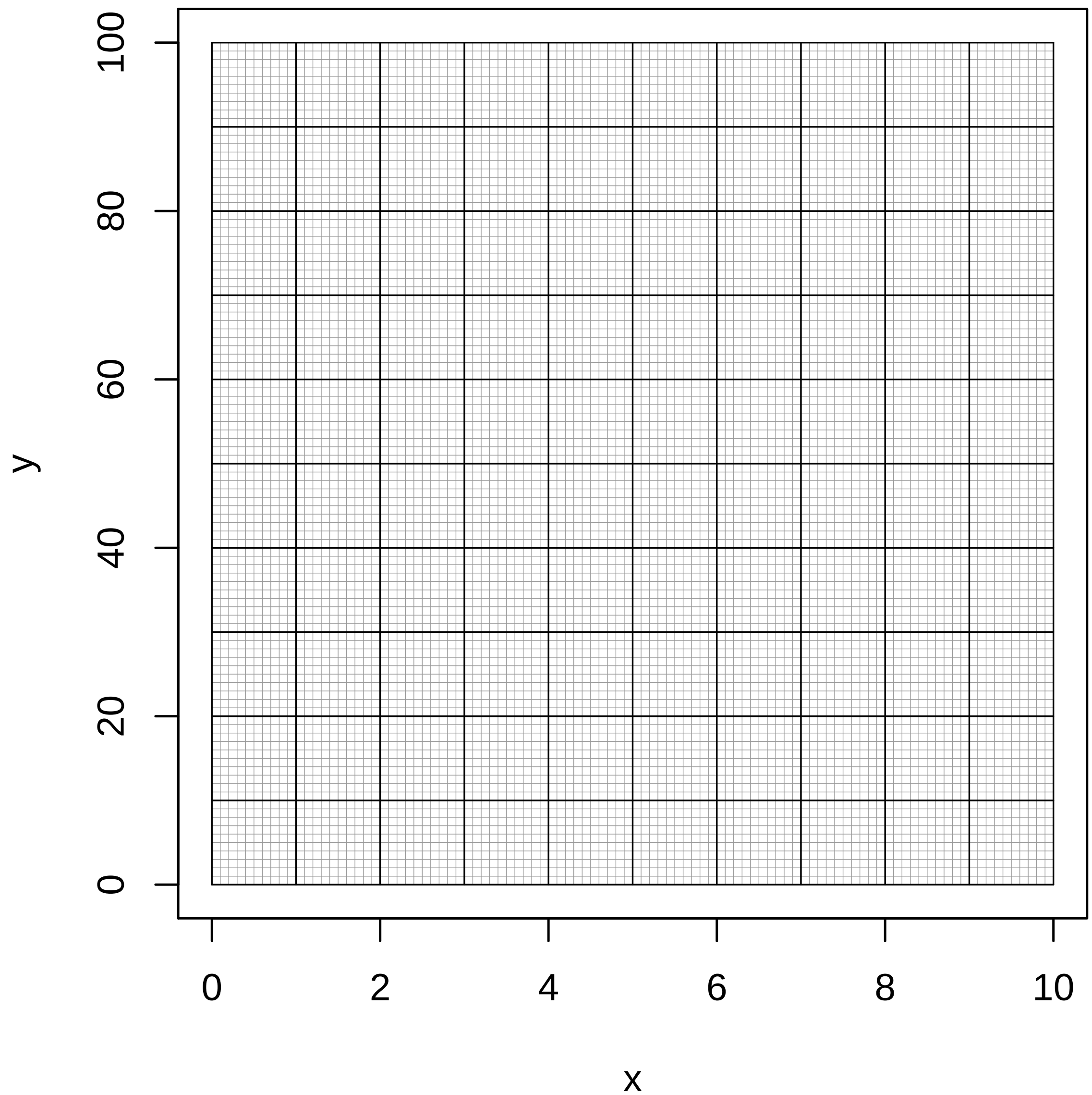
   (c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

   (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
2	71	142
6.7	16	107.2
8.7	9.1	79.17
9.6	5.5	52.8
9.3	26	241.8
4.2	22	92.4
7	11	77
$\sum x = 47.5$	$\sum y = 160.6$	$\sum x_i y_i = 792.37$
$\bar{x} = 6.786$	$\bar{y} = 22.94$	
$s_x = 2.815$	$s_y = 22.38$	

$$r = \frac{792.37 - (7)(6.786)(22.94)}{(7-1)(2.815)(22.38)} = -0.787$$

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.

$$r_{\text{exact}} = -0.7866275$$

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:

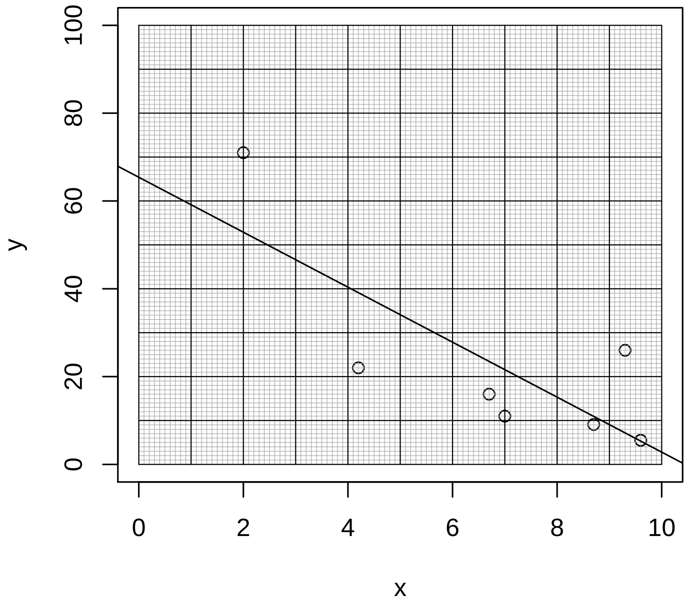
$$b = r\frac{s_y}{s_x} = -0.787 \cdot \frac{22.38}{2.815} = -6.26$$

$$a = \bar{y} - b\bar{x} = 22.9 - (-6.26)(6.79) = 65.4$$

Our regression line:

$$y = 65.4 + (-6.26)x$$

Make a plot.



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

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.19$. If 185 trials occur, what is the probability of getting at least 37 but at most 43 successes?

   In other words, let $X \sim \text{Bin}(n = 185,\ p = 0.19)$ and find $P(37 \leq X \leq 43)$.

   Use a normal approximation along with the continuity correction.

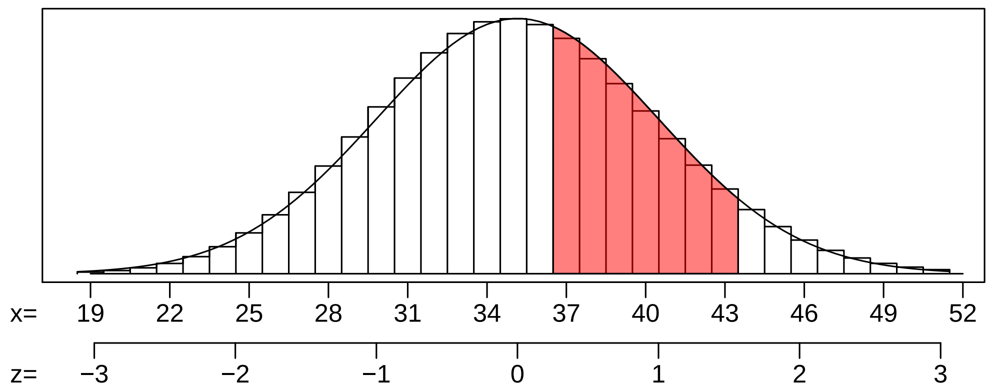
Solution: Find the mean.

$$\mu = np = (185)(0.19) = 35.15$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(185)(0.19)(1-0.19)} = 5.3359$$

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



Find the z scores.

$$z_1 = \frac{36.5 - 35.15}{5.3359} = 0.25$$

$$z_2 = \frac{43.5 - 35.15}{5.3359} = 1.56$$

Find the percentiles (from z-table).

$$\ell_1 = 0.5987$$

$$\ell_2 = 0.9406$$

Calculate the probability.

$$P(37 \leq X \leq 43) = 0.9406 - 0.5987 = 0.3419$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 80$. You decide to run two-tail test on a sample of size $n = 9$ using a significance level $\alpha = 0.02$.

    You then collect the sample:

79.6	80.9	82.7	79.4	83
85.3	81	82.5	81.5	

    (a) Determine the p-value.

    (b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \text{ claims } \mu = 80$$

$$H_A \text{ claims } \mu \neq 80$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 81.767$$

$$s = 1.841$$

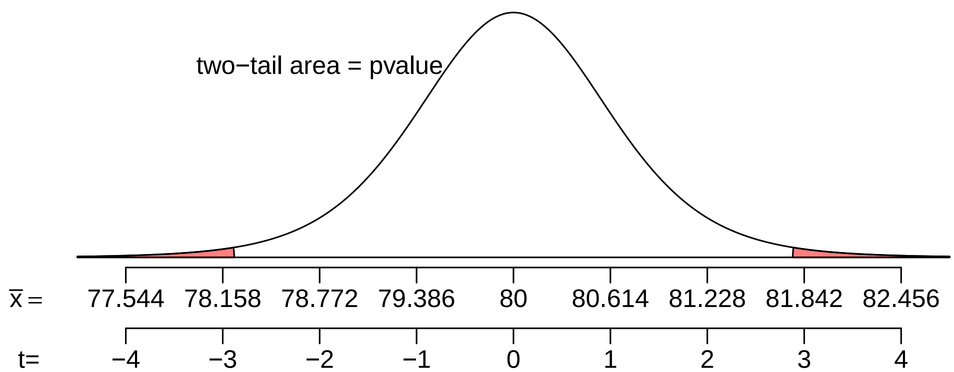
Determine the degrees of freedom.

$$df = 9 - 1 = 8$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{1.841}{\sqrt{9}} = 0.614$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{81.767 - 80}{0.614} = 2.88$$

Find the p-value.

$$p\text{-value} = P(|T| > 2.88)$$

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 $df = 8$.)

$$P(|T| > 2.9) = 0.02$$

$$P(|T| > 2.45) = 0.04$$

Basically, because t is between 2.9 and 2.45, 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.