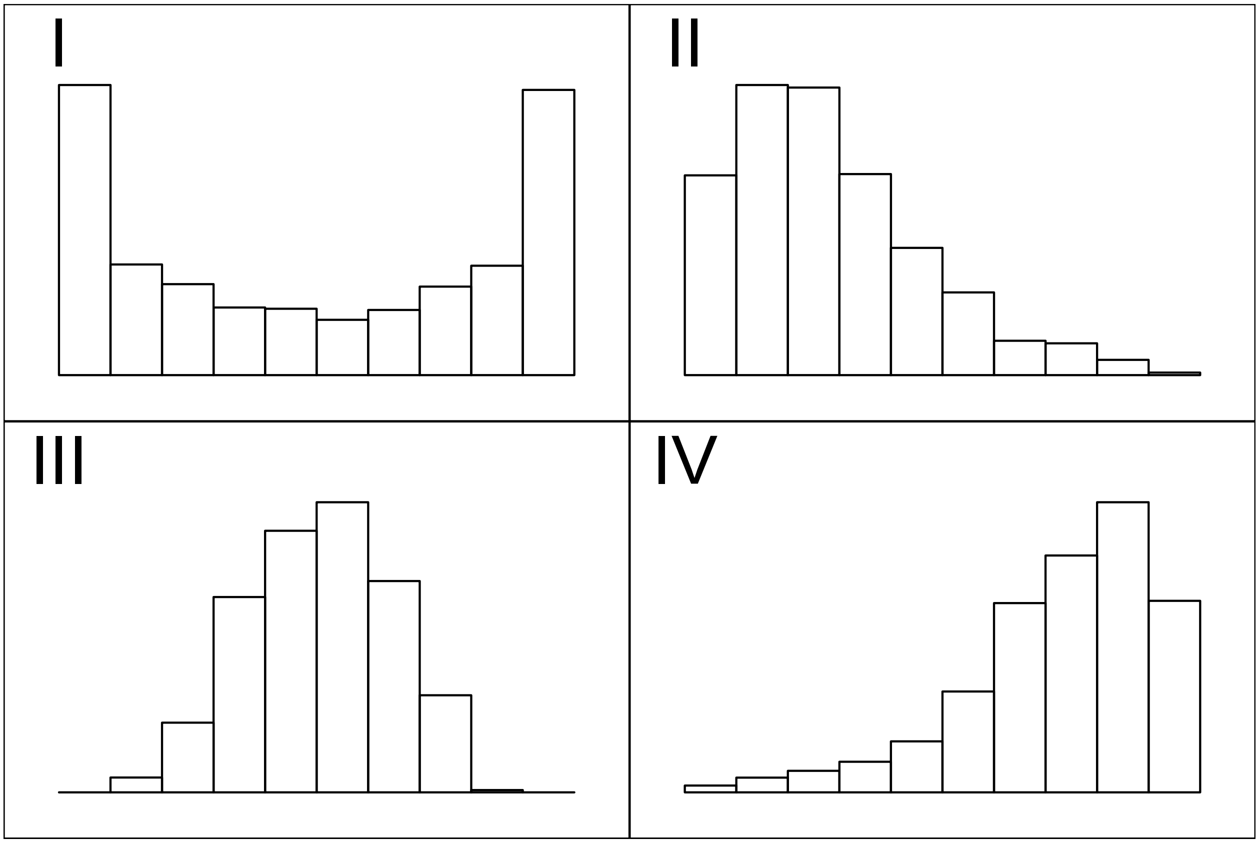
# MAT-181 FINAL TAKE-HOME EXAM

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 heights of adult men

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

(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 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) III
(b) IV
(c) II
(d) I

2. (15 Points)

   In a deck of strange cards, there are 309 cards. Each card has an image and a color. The amounts are shown in the table below.

| | gray | red | teal | violet | Total |
|---|---|---|---|---|---|
| dog | 22 | 38 | 37 | 21 | 118 |
| jigsaw | 29 | 45 | 15 | 14 | 103 |
| quilt | 28 | 24 | 25 | 11 | 88 |
| Total | 79 | 107 | 77 | 46 | 309 |

   (a) What is the probability a random card is both a jigsaw and violet?

   (b) What is the probability a random card is a quilt given it is teal?

   (c) Is a dog or a jigsaw more likely to be red?

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

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

   (f) What is the probability a random card is either a quilt or teal (or both)?

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

**Solution:**

(a) $P$(jigsaw and violet) = 0.0453

(b) $P$(quilt given teal) = 0.325

(c) $P$(red given dog) = 0.322 and $P$(red given jigsaw) = 0.437, so a jigsaw is more likely to be red than a dog is.

(d) $P$(gray given jigsaw) = 0.282

(e) $P$(quilt) = 0.285

(f) $P$(quilt or teal) = 0.453

(g) $P$(violet) = 0.149

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* | 123 | 14 |
| *B* | 138 | 8 |
| *C* | 103 | 10 |
| *D* | 146 | 6 |

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

| Type of fruit | Mass of specimen (g) |
|---|---|
| *A* | 101.3 |
| *B* | 136.1 |
| *C* | 106 |
| *D* | 136.3 |

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

**Solution:** We compare the $z$-scores. The smallest $z$-score corresponds to the specimen that is most unusually small.

| Type of fruit | formula | $z$-score |
|---|---|---|
| $A$ | $z = \frac{101.3-123}{14}$ | -1.55 |
| $B$ | $z = \frac{136.1-138}{8}$ | -0.24 |
| $C$ | $z = \frac{106-103}{10}$ | 0.3 |
| $D$ | $z = \frac{136.3-146}{6}$ | -1.62 |

Thus, the specimen of type $D$ is the most unusually small.

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 129 millimeters and a standard deviation of 6.2 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 114.4 and 125.8 millimeters?

**Solution:**

$$\mu = 129$$
$$\sigma = 6.2$$
$$x_1 = 114.4$$
$$x_2 = 125.8$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{114.4 - 129}{6.2} = -2.35$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{125.8 - 129}{6.2} = -0.52$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.3015 - 0.0094 = 0.2921$$

5. (10 points)

   A species of duck is known to have a mean weight of 164.3 grams and a standard deviation of 130 grams. A researcher plans to measure the weights of 169 of these ducks sampled randomly. What is the probability the **sample mean** will be between 141.3 and 180.8 grams?

**Solution:**

$$n = 169$$
$$\mu = 164.3$$
$$\sigma = 130$$
$$SE = \frac{130}{\sqrt{169}} = 10$$
$$x_1 = 141.3$$
$$x_2 = 180.8$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{141.3 - 164.3}{10} = -2.3$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{180.8 - 164.3}{10} = 1.65$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.9505 - 0.0107 = 0.9398$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Denrdoica magnolia*. She randomly samples 16 adults of *Denrdoica magnolia*, resulting in a sample mean of 9.33 grams and a sample standard deviation of 1.67 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.

$$\begin{aligned} n &= 16 \\ \bar{x} &= 9.33 \\ s &= 1.67 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 16 - 1 \\ &= 15 \end{aligned}$$

Determine the critical $t$ value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 15$.

$$t^{\star} = 2.13$$

Use the formula for bounds (mean, $\sigma$ unknown).

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} & \qquad UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 9.33 - 2.13 \times \frac{1.67}{\sqrt{16}} & &= 9.33 + 2.13 \times \frac{1.67}{\sqrt{16}} \\ &= 8.44 & &= 10.2 \end{aligned}$$

We are 95% confident that the population mean is between 8.44 and 10.2 grams.

$$CI = (8.44,\ 10.2)$$

7. (15 points)

A student is taking a multiple choice test with 700 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 159 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.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.2(1-0.2)}{700}} = 0.0151$$

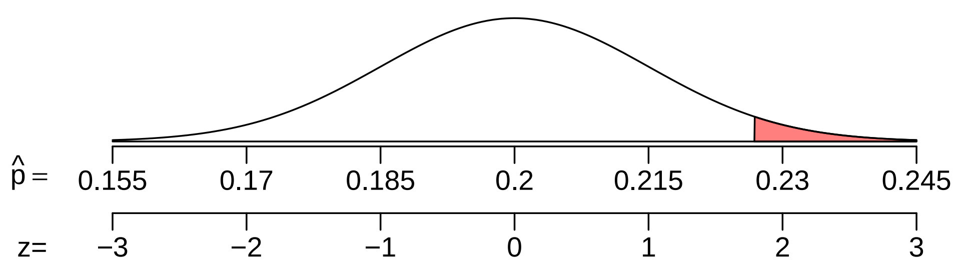
Determine the sample proportion.

$$\hat{p} = \frac{159}{700} = 0.227$$

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

$$z = \frac{\hat{p}-p_0}{\sigma_{\hat{p}}} = \frac{0.227-0.2}{0.0151} = 1.79$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.227) \\ &= P(Z > 1.79) \\ &= 1 - P(Z < 1.79) \\ &= 0.0367 \end{aligned}$$

Compare $p$-value to $\alpha$ (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 $p = 0.2$ and $H_A$ claims $p > 0.2$.

(c) The $p$-value is 0.0367

(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:

| $x$ | $y$ | $xy$ |
|---|---|---|
| 640 | 25 | |
| 900 | 52 | |
| 550 | 42 | |
| 600 | 35 | |
| 610 | 53 | |
| 400 | 44 | |
| 280 | 28 | |
| 570 | 19 | |
| 120 | 25 | |
| $\sum x =$ | $\sum y =$ | $\sum xy =$ |
| $\bar{x} =$ | $\bar{y} =$ | |
| $s_x =$ | $s_y =$ | |

(a) Complete the table.

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

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

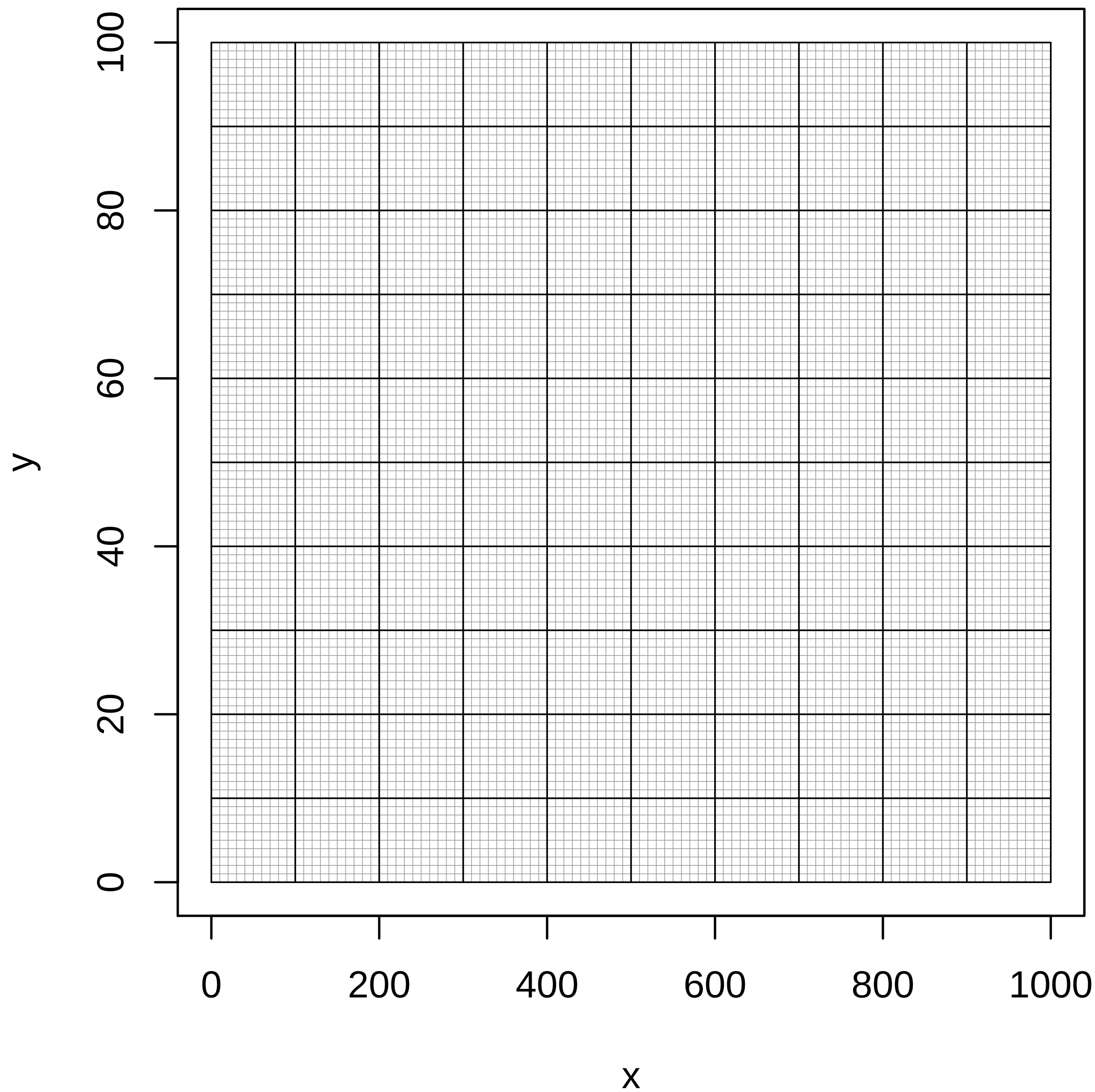
(c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters ($b$ and $a$) using the formulas below.

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

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

(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.

| $x$ | $y$ | $xy$ |
|---|---|---|
| 640 | 25 | 16000 |
| 900 | 52 | 46800 |
| 550 | 42 | 23100 |
| 600 | 35 | 21000 |
| 610 | 53 | 32330 |
| 400 | 44 | 17600 |
| 280 | 28 | 7840 |
| 570 | 19 | 10830 |
| 120 | 25 | 3000 |
| $\sum x = 4670$ | $\sum y = 323$ | $\sum x_i y_i = 178500$ |
| $\bar{x} = 518.9$ | $\bar{y} = 35.89$ | |
| $s_x = 226$ | $s_y = 12.45$ | |

$$r = \frac{178500 - (9)(518.9)(35.89)}{(9-1)(226)(12.45)} = 0.484$$

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.

$$r_{\text{exact}} = 0.483971$$

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:

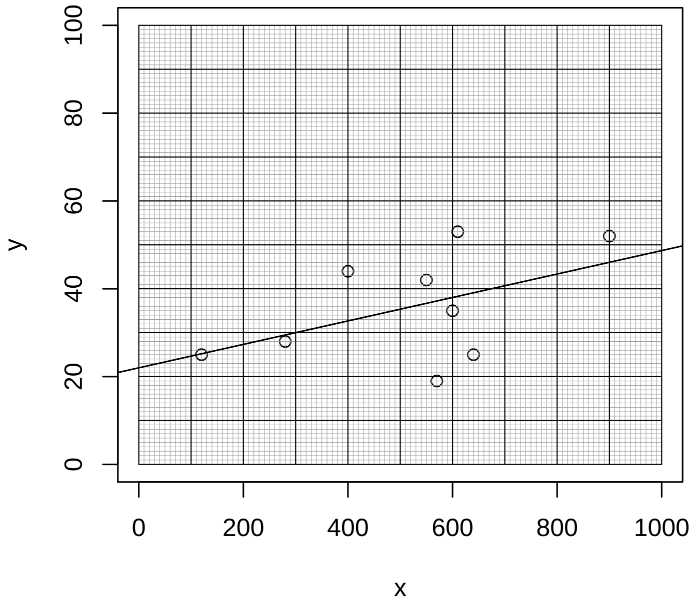
$$b = r\frac{s_y}{s_x} = 0.484 \cdot \frac{12.45}{226} = 0.0267$$

$$a = \bar{y} - b\bar{x} = 35.9 - (0.0267)(519) = 22$$

Our regression line:

$$y = 22 + (0.0267)x$$

Make a plot.


y
x
0
20
40
60
80
100
0
200
400
600
800
1000

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.49$. If 129 trials occur, what is the probability of getting at least 51 but at most 77 successes?

   In other words, let $X \sim \text{Bin}(n = 129,\ p = 0.49)$ and find $P(51 \leq X \leq 77)$.

   Use a normal approximation along with the continuity correction.

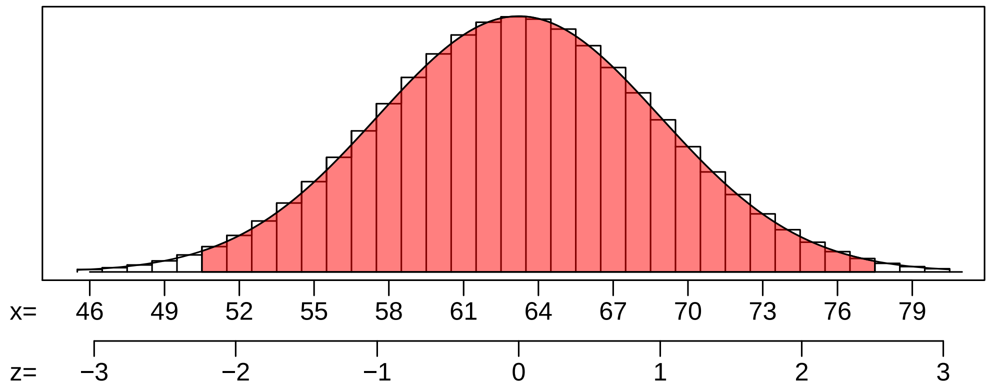
**Solution:** Find the mean.

$$\mu = np = (129)(0.49) = 63.21$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(129)(0.49)(1-0.49)} = 5.6778$$

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




Find the $z$ scores.

$$z_1 = \frac{50.5 - 63.21}{5.6778} = -2.24$$

$$z_2 = \frac{77.5 - 63.21}{5.6778} = 2.52$$

Find the percentiles (from $z$-table).

$$\ell_1 = 0.0125$$

$$\ell_2 = 0.9941$$

Calculate the probability.

$$P(51 \leq X \leq 77) = 0.9941 - 0.0125 = 0.9816$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 160$. You decide to run two-tail test on a sample of size $n = 8$ using a significance level $\alpha = 0.1$.

    You then collect the sample:

| | | | | |
|---|---|---|---|---|
| 176 | 169.1 | 160.3 | 162.8 | 165.3 |
| 151.6 | 166.2 | 169.1 | | |

    (a) Determine the $p$-value.

    (b) Do you reject the null hypothesis?

**Solution:** State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 160$$

$$H_A \quad \text{claims} \quad \mu \neq 160$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 165.05$$

$$s = 7.197$$

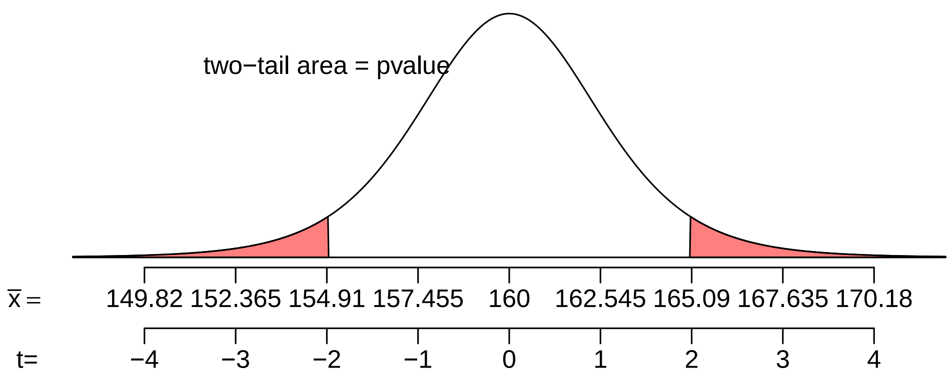
Determine the degrees of freedom.

$$df = 8 - 1 = 7$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{7.197}{\sqrt{8}} = 2.545$$

Make a sketch of the null's sampling distribution.




Find the $t$ score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{165.05 - 160}{2.545} = 1.98$$

Find the $p$-value.

$$p\text{-value} = P(|T| > 1.98)$$

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 $df = 7$.)

$$P(|T| > 2.36) = 0.05$$

$$P(|T| > 1.89) = 0.1$$

Basically, because $t$ is between 2.36 and 1.89, we know the $p$-value is between 0.05 and 0.1.

$$0.05 < p\text{-value} < 0.1$$

Compare the $p$-value and the significance level ($\alpha = 0.1$).

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

Yes, we reject the null hypothesis.

(a) $0.05 < p\text{-value} < 0.1$

(b) Yes, we reject the null hypothesis.