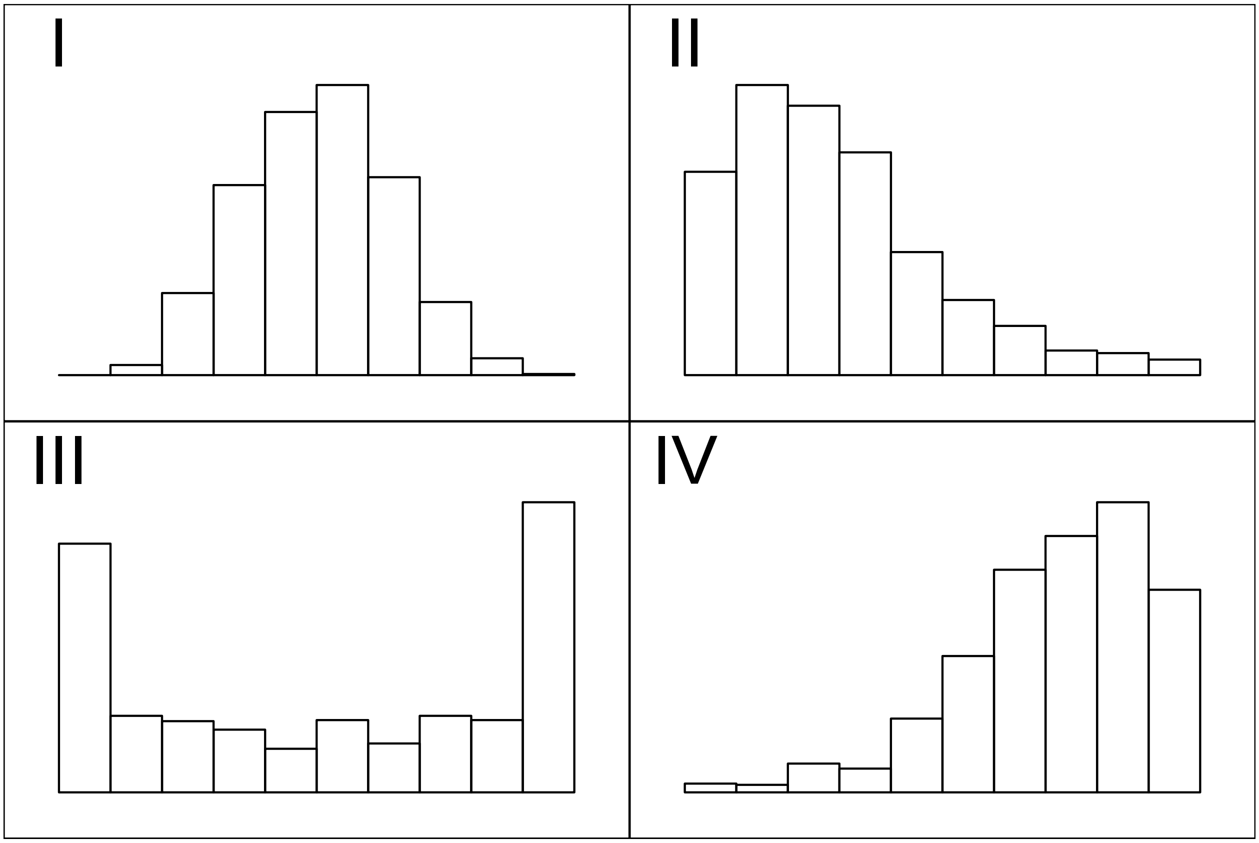
# MAT-181 FINAL TAKE-HOME EXAM

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 weights of newborn babies

(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 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.

(d) 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.

**Solution:**

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

2. (15 Points)

In a deck of strange cards, there are 415 cards. Each card has an image and a color. The amounts are shown in the table below.

| | blue | gray | green | pink | teal | Total |
|---|---|---|---|---|---|---|
| jigsaw | 45 | 12 | 10 | 38 | 33 | 138 |
| tree | 19 | 29 | 30 | 35 | 28 | 141 |
| wheel | 21 | 48 | 27 | 24 | 16 | 136 |
| Total | 85 | 89 | 67 | 97 | 77 | 415 |

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

(b) Is a jigsaw or a wheel more likely to be blue?

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

(d) What is the probability a random card is green given it is a wheel?

(e) What is the probability a random card is either a wheel or green (or both)?

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

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

**Solution:**

(a) $P$(wheel and green) = 0.0651

(b) $P$(blue given jigsaw) = 0.326 and $P$(blue given wheel) = 0.154, so a jigsaw is more likely to be blue than a wheel is.

(c) $P$(wheel) = 0.328

(d) $P$(green given wheel) = 0.199

(e) $P$(wheel or green) = 0.424

(f) $P$(tree given pink) = 0.361

(g) $P$(pink) = 0.234

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* | 113 | 5 |
| *B* | 134 | 4 |
| *C* | 133 | 13 |
| *D* | 99 | 7 |

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

| Type of fruit | Mass of specimen (g) |
|---|---|
| *A* | 112.2 |
| *B* | 136.6 |
| *C* | 138.7 |
| *D* | 97.11 |

   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 |
|---|---|---|
| $A$ | $z = \frac{\lvert 112.2-113 \rvert}{5}$ | 0.17 |
| $B$ | $z = \frac{\lvert 136.6-134 \rvert}{4}$ | 0.66 |
| $C$ | $z = \frac{\lvert 138.7-133 \rvert}{13}$ | 0.44 |
| $D$ | $z = \frac{\lvert 97.11-99 \rvert}{7}$ | 0.27 |

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 69.7 millimeters and a standard deviation of 9.5 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 53.1 and 89.8 millimeters?

**Solution:**

$$\mu = 69.7$$
$$\sigma = 9.5$$
$$x_1 = 53.1$$
$$x_2 = 89.8$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{53.1 - 69.7}{9.5} = -1.75$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{89.8 - 69.7}{9.5} = 2.12$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.983 - 0.0401 = 0.9429$$

5. (10 points)

   A species of duck is known to have a mean weight of 171.6 grams and a standard deviation of 72 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 158.1 and 162.6 grams?

**Solution:**

$$n = 64$$
$$\mu = 171.6$$
$$\sigma = 72$$
$$SE = \frac{72}{\sqrt{64}} = 9$$
$$x_1 = 158.1$$
$$x_2 = 162.6$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{158.1 - 171.6}{9} = -1.5$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{162.6 - 171.6}{9} = -1$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.1587 - 0.0668 = 0.0919$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Archilochus colubris*. She randomly samples 14 adults of *Archilochus colubris*, resulting in a sample mean of 3.35 grams and a sample standard deviation of 0.686 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.

$$\begin{aligned} n &= 14 \\ \bar{x} &= 3.35 \\ s &= 0.686 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 14 - 1 \\ &= 13 \end{aligned}$$

Determine the critical $t$ value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 13$.

$$t^{\star} = 2.16$$

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

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} \\ &= 3.35 - 2.16 \times \frac{0.686}{\sqrt{14}} \\ &= 2.95 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 3.35 + 2.16 \times \frac{0.686}{\sqrt{14}} \\ &= 3.75 \end{aligned}$$

We are 95% confident that the population mean is between 2.95 and 3.75 grams.

$$CI = (2.95,\ 3.75)$$

7. (15 points)

   A student is taking a multiple choice test with 600 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 170 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.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.25(1-0.25)}{600}} = 0.0177$$

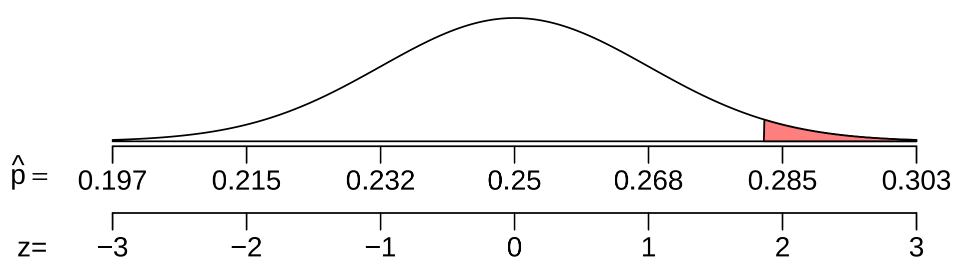
Determine the sample proportion.

$$\hat{p} = \frac{170}{600} = 0.283$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.283 - 0.25}{0.0177} = 1.86$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.283) \\ &= P(Z > 1.86) \\ &= 1 - P(Z < 1.86) \\ &= 0.0314 \end{aligned}$$

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 $p = 0.25$ and $H_A$ claims $p > 0.25$.

(c) The $p$-value is 0.0314

(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:

| $x$ | $y$ | $xy$ |
|---|---|---|
| 280 | 68 | |
| 890 | 78 | |
| 110 | 39 | |
| 770 | 77 | |
| 480 | 61 | |
| 360 | 62 | |
| $\sum x =$ | $\sum y =$ | $\sum xy =$ |
| $\bar{x} =$ | $\bar{y} =$ | |
| $s_x =$ | $s_y =$ | |

   (a) Complete the table.

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

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

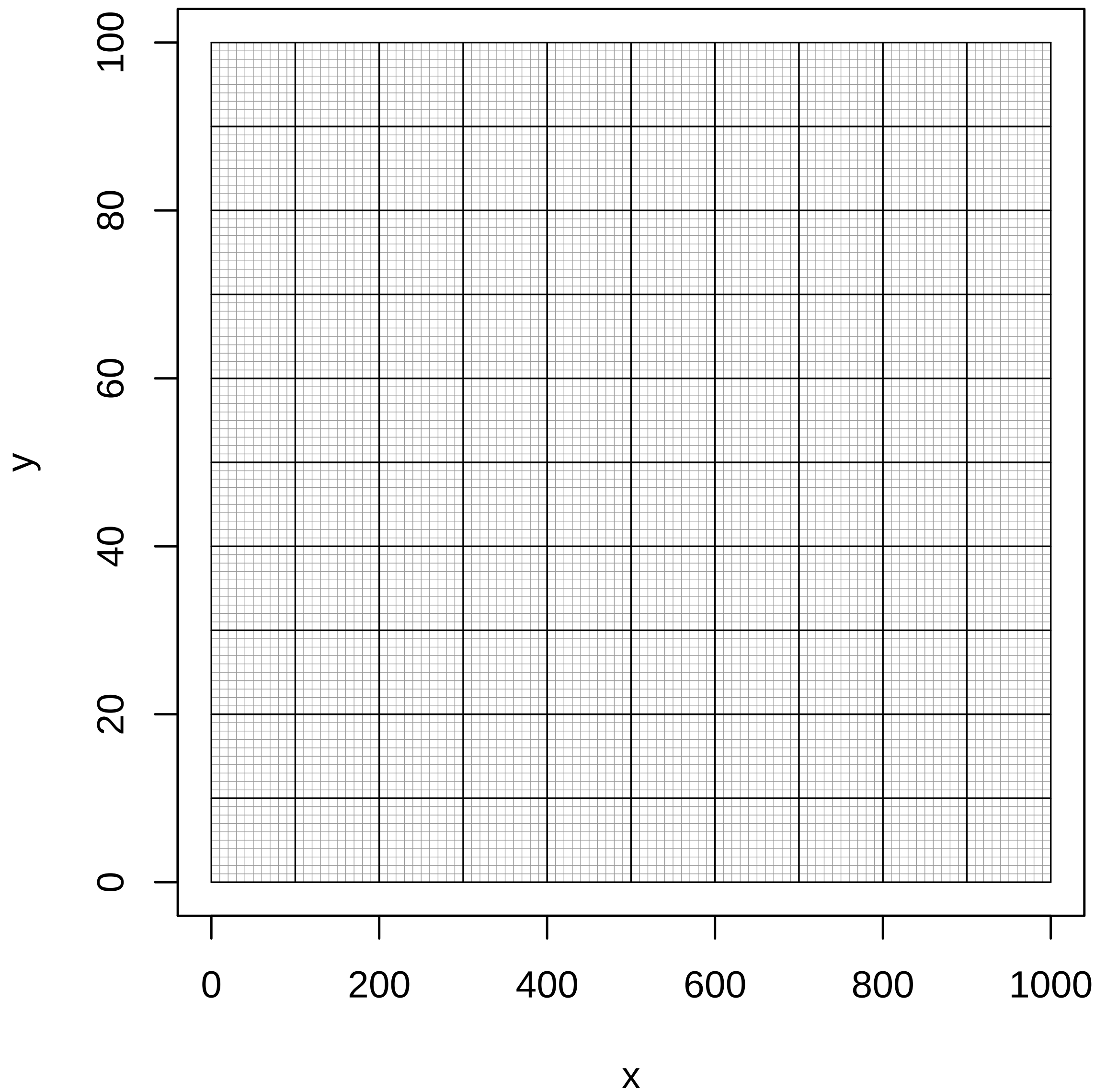
   (c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters ($b$ and $a$) using the formulas below.

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

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

   (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.

| $x$ | $y$ | $xy$ |
|---|---|---|
| 280 | 68 | 19040 |
| 890 | 78 | 69420 |
| 110 | 39 | 4290 |
| 770 | 77 | 59290 |
| 480 | 61 | 29280 |
| 360 | 62 | 22320 |
| $\sum x = 2890$ | $\sum y = 385$ | $\sum x_i y_i = 203640$ |
| $\bar{x} = 481.7$ | $\bar{y} = 64.17$ | |
| $s_x = 297.8$ | $s_y = 14.27$ | |

$$r = \frac{203640 - (6)(481.7)(64.17)}{(6-1)(297.8)(14.27)} = 0.855$$

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.

$$r_{\text{exact}} = 0.8561333$$

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:

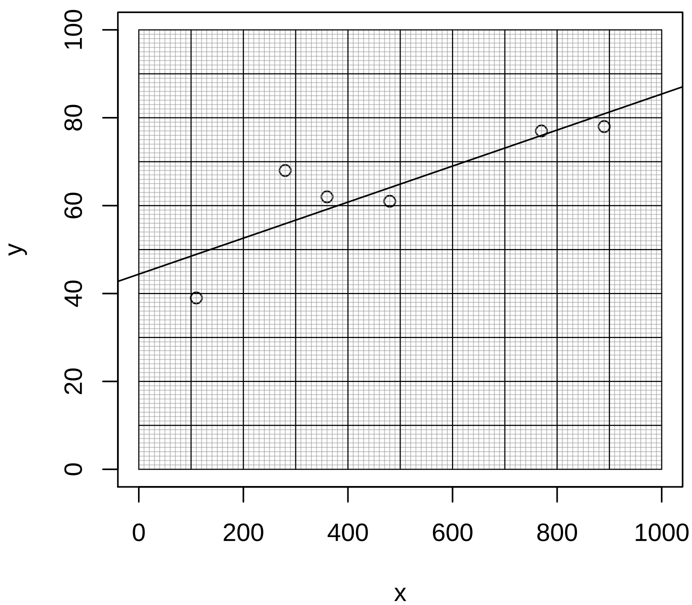
$$b = r\frac{s_y}{s_x} = 0.855 \cdot \frac{14.27}{297.8} = 0.041$$

$$a = \bar{y} - b\bar{x} = 64.2 - (0.041)(482) = 44.4$$

Our regression line:

$$y = 44.4 + (0.041)x$$

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 $p = 0.39$. If 75 trials occur, what is the probability of getting more than 21 but less than 36 successes?

   In other words, let $X \sim \text{Bin}(n = 75,\ p = 0.39)$ and find $P(21 < X < 36)$.

   Use a normal approximation along with the continuity correction.

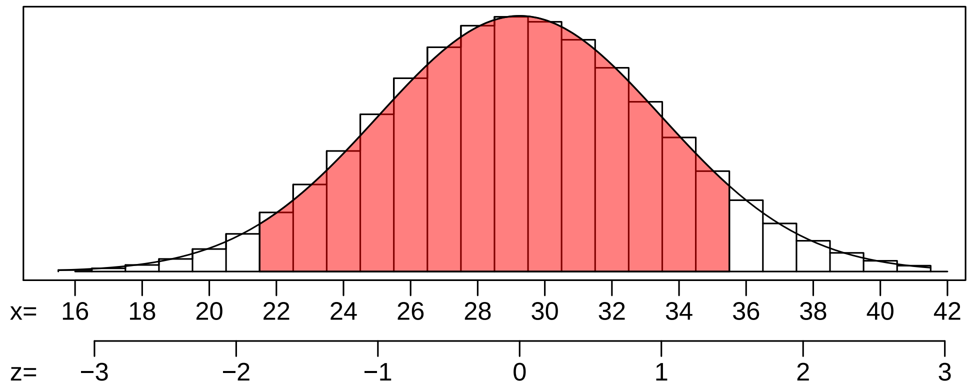
**Solution:** Find the mean.

$$\mu = np = (75)(0.39) = 29.25$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(75)(0.39)(1-0.39)} = 4.224$$

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




Find the $z$ scores.

$$z_1 = \frac{21.5 - 29.25}{4.224} = -1.83$$

$$z_2 = \frac{35.5 - 29.25}{4.224} = 1.48$$

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

$$\ell_1 = 0.0336$$

$$\ell_2 = 0.9306$$

Calculate the probability.

$$P(21 < X < 36) = 0.9306 - 0.0336 = 0.897$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 150$. You decide to run two-tail test on a sample of size $n = 9$ using a significance level $\alpha = 0.05$.

    You then collect the sample:

| | | | | |
|---|---|---|---|---|
| 134.7 | 204.4 | 157.5 | 138.7 | 154.9 |
| 216.2 | 217.3 | 225.1 | 162.3 | |

    (a) Determine the $p$-value.

    (b) Do you reject the null hypothesis?

**Solution:** State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 150$$

$$H_A \quad \text{claims} \quad \mu \neq 150$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 179.011$$

$$s = 36.28$$

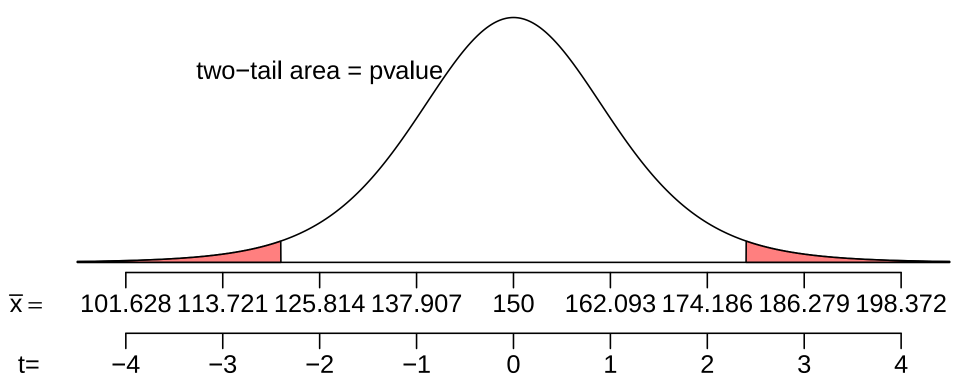
Determine the degrees of freedom.

$$df = 9 - 1 = 8$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{36.28}{\sqrt{9}} = 12.093$$

Make a sketch of the null's sampling distribution.




Find the $t$ score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{179.011 - 150}{12.093} = 2.4$$

Find the $p$-value.

$$p\text{-value} = P(|T| > 2.4)$$

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 $df = 8$.)

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

$$P(|T| > 2.31) = 0.05$$

Basically, because $t$ is between 2.45 and 2.31, we know the $p$-value is between 0.04 and 0.05.

$$0.04 < p\text{-value} < 0.05$$

Compare the $p$-value and the significance level ($\alpha = 0.05$).

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

Yes, we reject the null hypothesis.

(a) $0.04 < p\text{-value} < 0.05$
(b) Yes, we reject the null hypothesis.