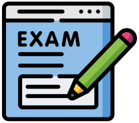
IMS 20: Inference for comparing two independent means

Libraries Needed

```r
library(openintro)
library(tidyverse)
```

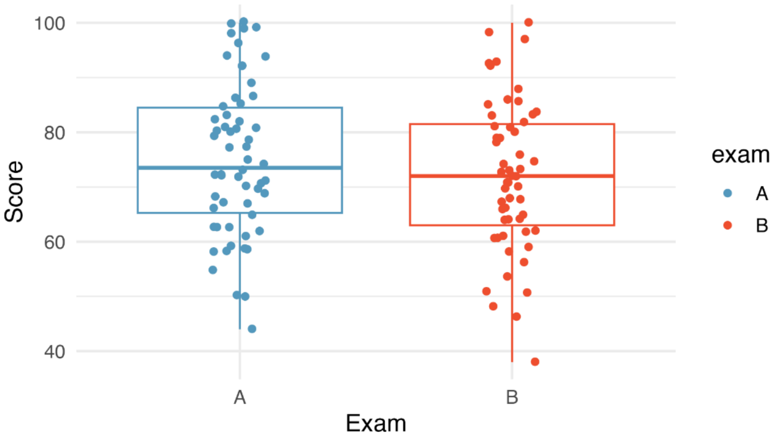
Example

An instructor decided to run two slight variations of the same exam. Anticipating complaints from students who took Version B, they would like to evaluate whether the difference observed in the groups is so large that it provides convincing evidence that Version B was more difficult (on average) than Version A.

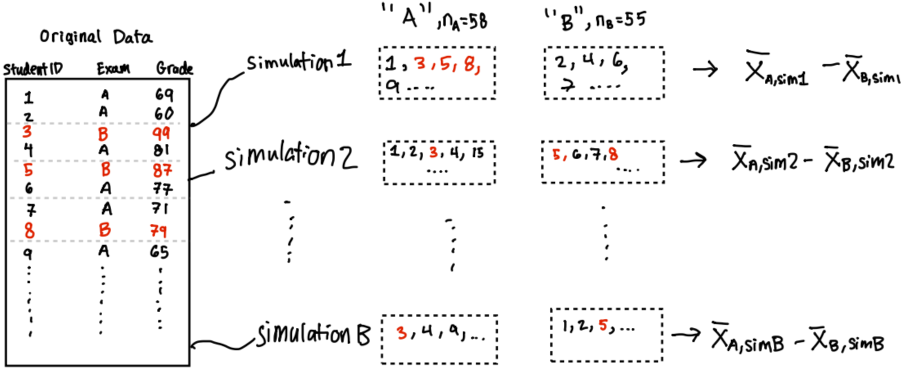


Group	n	Mean	SD	Min	Max
A	58	75.1	13.9	44	100
B	55	72.0	13.8	38	100

Example



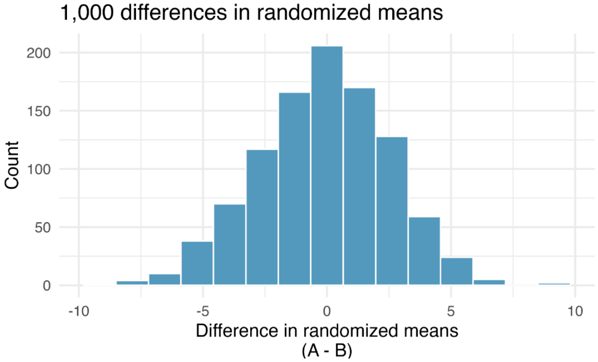
Randomization Test for Means



Randomization Test for Means

Distribution of the $\overline{x}_{A,sim} - \overline{x}_{B,sim}$ values.

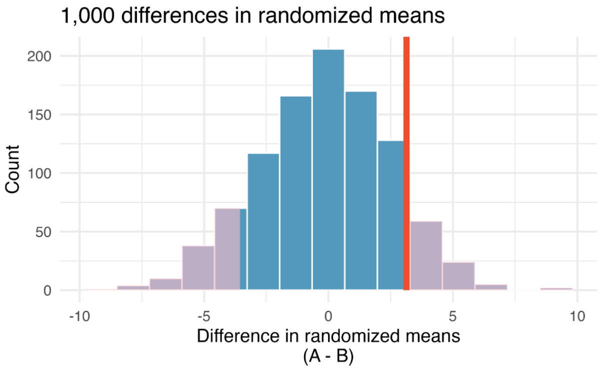
In other words: the distribution of possible values for $\overline{x}_A - \overline{x}_B$ if there was no difference between the tests.



Randomization Test for Means

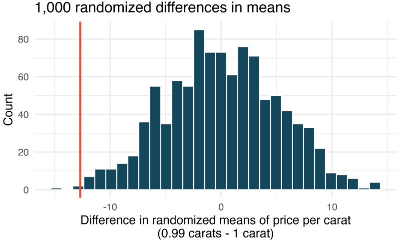
Calculate the p-value for a randomization test:

- Look at the value we actually observed → $\overline{x}_A - \overline{x}_B = 3.1$
- Find the proportion of $\overline{x}_{A,sim} - \overline{x}_{B,sim}$ that were more extreme → 0.10
- Multiple by 2 → p-value = 0.20



Practice Problem

[IMS 20.3] Diamonds, randomization test. We have data on two random samples of diamonds: one with diamonds that weigh 0.99 carats and one with diamonds that weigh 1 carat. Each sample has 23 diamonds. The randomization distribution (with 1,000 repetitions) below describes the null distribution of the difference in sample means (of price per carat) if there really was no difference in the population from which these diamonds came.



Conduct a hypothesis test using $\alpha = 0.10$ to evaluate if there is a difference between the prices per carat of diamonds that weigh 0.99 carats and diamonds that weigh 1 carat. Note that the observed difference marked on the plot with a vertical line is -12.7.

Mathematical model for testing the difference in means

Step 1) Frame the research question in terms of hypotheses.

$$H_0 : \mu_A - \mu_B = 0$$
$$H_A : \mu_A - \mu_B \neq 0$$

Step 2) Collect data (check conditions)

- **Independence**: Need independence on two levels: between and within groups.
- **Normality**: No extreme data values, data is approximately symmetric for both groups.

Step 3) Model the randomness that would occur if the null hypothesis was true

Standard Error for a difference in means

If we have independence between and with groups, and the data is approximately normal for both groups, the the standard error for the difference in means $(\overline{x}_A - \overline{x}_B)$ is

$$SE = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

For our example we would get

$$SE = \sqrt{\frac{13.9^2}{58} + \frac{13.8^2}{55}} \approx 2.6065$$

Step 3) Model the randomness that would occur if the null hypothesis was true

Test-statistic (T-score) for a difference of two means

$$T = \frac{\overline{x}_A - \overline{x}_B - 0}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

The distribution of T under H_0 is

$$T \overset{H_0}{\sim} t(df = \text{min}(n_A - 1, n_B - 1))$$

For our example $T \overset{H_0}{\sim} t(df = 54)$

Step 4) Analyze the data.

- Calculate the test statistic.
- Find the tail area.
- If it is a two-sided test, multiply the area by 2.

$$T = \frac{75.1 - 72 - 0}{\sqrt{\frac{13.9^2}{58} + \frac{13.8^2}{55}}} \approx 1.1893$$

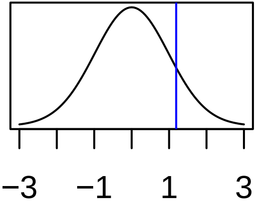
Step 4) Analyze the data.

THIS IS THE ONLY CODE BOX YOU NEED TO ADJUST:

```
my_df <- 54
null <- 0
my_se <-  sqrt(13.9^2/58 + 13.8^2/55)
my_mean_diff <- 75.1 - 72
my_test_stat <- (my_mean_diff - null)/my_se
```

To draw a picture with a vertical line at your test statistic (T), use this code:

```
curve(dt(x, df = my_df), to = -3, from = 3,
      yaxt="n", ylab = "", xlab = "")
abline(v = my_test_stat, col = "blue")
```



Step 4) Analyze the data.

Calculate the left tail area of your test statistic:

```
1-pt(my_test_stat, df= my_df)
```

```
[1] 0.1197542
```

P-value for this setting:

```
2*(1-pt(my_test_stat, df= my_df))
```

```
[1] 0.2395085
```

Step 5: Form a conclusion.

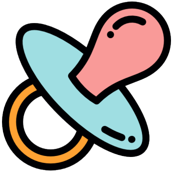
Same rules as before.

In this example: *The p-value is approximately 0.2395 which is greater than* $\alpha = 0.05$*. We conclude in favor of* H_0*. It seems reasonable that the two exams are the same level of rigor.*

Steps at a glance

1. Frame the research question in terms of hypotheses.
   - μ_A and μ_B
2. Collect data (check conditions)
   - Independence (beteween and within groups) and approx normal.
3. Model the randomness that would occur if the null hypothesis was true.
   - Calculate the test statistic T.
4. Analyze the data.
   - Use $T \sim t(df = \min(n_A - 1, n_B - 1))$ to get the p-value.
   - Calculate tail area with `pt()`
5. Form a conclusion.

Practice Problem



Every year, the US releases to the public a large data set containing information on births recorded in the country. We will work with a random sample of 1,000 cases from the data set released in 2014.

We can see that the average birth weight of babies born to smoker moms is lower than those born to nonsmoker moms.

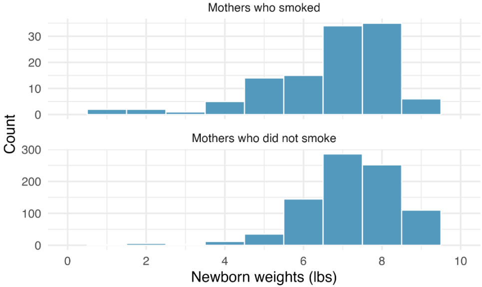
In the 1970s the tobacco industry claimed that *"babies born from women who smoke are... just as healthy as the babies born from women who do not smoke"*.

Is there is a significant difference in baby birth weight between smokers and non-smokers?

Practice Problem

Is there is a significant difference in baby birth weight between smokers and non-smokers?

Habit	n	Mean	SD
nonsmoker	867	7.27	1.23
smoker	114	6.68	1.60



Confidence Intervals

To create confidence intervals using for the difference of two means we check the same **independence** and **normality** conditions as a standard hypothesis test for the difference in two means.

Confidence Intervals for the difference of means

$$lower = (\overline{x}_A - \overline{x}_B) - t^*_{df} SE$$

$$upper = (\overline{x}_A - \overline{x}_B) + t^*_{df} SE$$

where $SE = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$

Using R

If we have access to the full data set, we can use `t.test()` to calculate our p-value, test statistic, and confidence interval!

```
smoking <- births14 |>
  filter(habit == "smoker")  |>
  select(weight)

nonsmoking <- births14 |>
  filter(habit == "nonsmoker")|>
  select(weight)

t.test(smoking, nonsmoking,
       mu = 0,
       alternative = "two.sided",
       conf.level = .95)
```

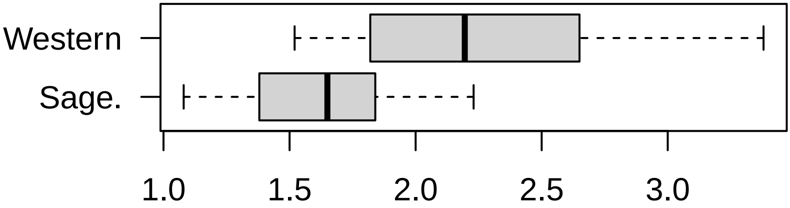
Using R

```
    Welch Two Sample t-test

data:  smoking and nonsmoking
t = -3.8166, df = 131.31, p-value = 0.0002075
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8998751 -0.2854852
sample estimates:
mean of x mean of y
 6.677193  7.269873
```

Practice Problem

Data on top speeds measured on a laboratory race track for 26 Sagebrush lizards and 22 Western fence lizards. Below is a boxplot and some summary statistics.



	n	mean	std. dev.
Western	22	2.3145	0.5555
Sage	26	1.6127	0.3241

The data is stored in a data set called `lizard_run` which is in the `openintro` package in R.

Practice Problem

Calculate a 90% confidence interval to test if the mean difference between the Sagebrush and Western fence lizards is 0.

Calculate the p-value for testing if the mean difference between the Sagebrush and Western fence lizards is 0. Use $\alpha = 0.10$ and state your conclusion.