STA220H1: The Practice of Statistics I

Elena Tuzhilina

March 21, 2023

Please turn on your videos :)

Announcements

1. Midterm 2 solution is posted.
2. The grades will be released next week.
3. We have three more lectures!

Agenda for today

- Recap: statistical testing, H_0 and H_a, process, p-value
- More on statistical testing: power, type I and II error, connection to confidence intervals
- Statistical testing for two samples

Statistical testing

Statistical tests use data to answer questions about the population.

We want to study the average height in Canada.

We take a sample of n people, record their heights and compute the sample mean and standard deviations:

$$x_1, \ldots, x_n \Rightarrow \bar{x}, s$$

How to use this information to check if the average height is equal to some pre-specified value?

Statistical testing

Step 1: state your **null** hypothesis and the **alternative** hypothesis.

$H_0 : \mu = \mu_0$

$H_a : \mu > \mu_0$ or $H_a : \mu < \mu_0$ or $H_a : \mu \neq \mu_0$

Step 2: summarize the data into a **test statistic**.

$$t_{obs} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Step 3: compute **p-value**.

p-value $= P(\overset{\sim t_{n-1}}{T} > t_{obs})$ or $P(T < t_{obs})$ or $P(|T| > |t_{obs}|)$

Step 4: draw the conclusion.

If *p-value* $< \alpha$, thus we can reject H_0 in favor of H_a.

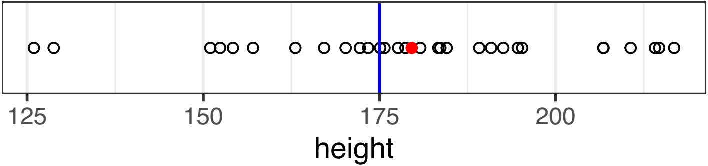
Else, we "do not have enough evidence to reject the null."

Statistical testing

Is the average height in Canada is 175 cm?

$\mu = 175$

```
sample = rnorm(n = 30, mean = 175, sd = 20)
```




```
mean(sample)
```

```
## [1] 179.5734
```

175

```
sd(sample)
```

```
## [1] 23.47429
```

Statistical testing

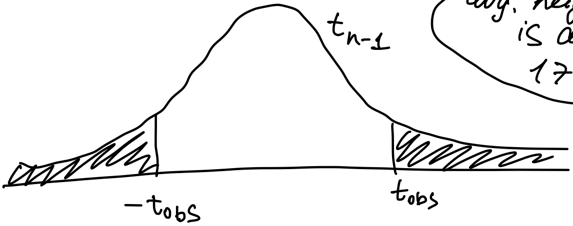
Step 1: $H_0 : \mu = 175$ and $H_a : \mu \neq 175$

Step 2: $t_{obs} = 1.0671 = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{179.6-175}{23.5/\sqrt{30}}$

Step 3: *p-value* $= P(|T| > |t_{obs}|) = 0.2947$

Step 4: *p-value* > 0.05, thus we cannot reject H_0

avg. heigh in Canada is different from 175cm



Statistical testing

```r
t.test(sample, mu = 175, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  sample
## t = 1.0671, df = 29, p-value = 0.2947
## alternative hypothesis: true mean is not equal to 175
## 95 percent confidence interval:
##  170.8080 188.3389
## sample estimates:
## mean of x
##  179.5734
```

Statistical testing: comments

Statistical testing can only check if H_0 is incorrect

- *p-value* $> \alpha$ **does not mean** that we can accept H_0

Is the average height in Canada is 175.5 cm?

$H_0 : \mu = 175.5$

$H_0 : \mu \neq 175.5$

Step 1: $H_0 : \mu = 175.5$ and $H_a : \mu \neq 175.5$

Step 2: $t_{obs} = 0.95045$

Step 3: *p-value* $= P(|T| > |t_{obs}|) = 0.3497$

Step 4: *p-value* > 0.05, thus we cannot reject H_0

Statistical testing: comments

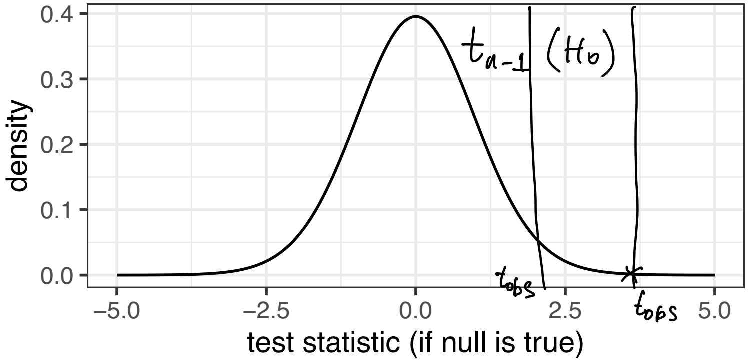
```r
t.test(sample, mu = 175.5, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  sample
## t = 0.95045, df = 29, p-value = 0.3497
## alternative hypothesis: true mean is not equal to 175.5
## 95 percent confidence interval:
##  170.8080 188.3389
## sample estimates:
## mean of x
##  179.5734
```

Statistical testing: comments

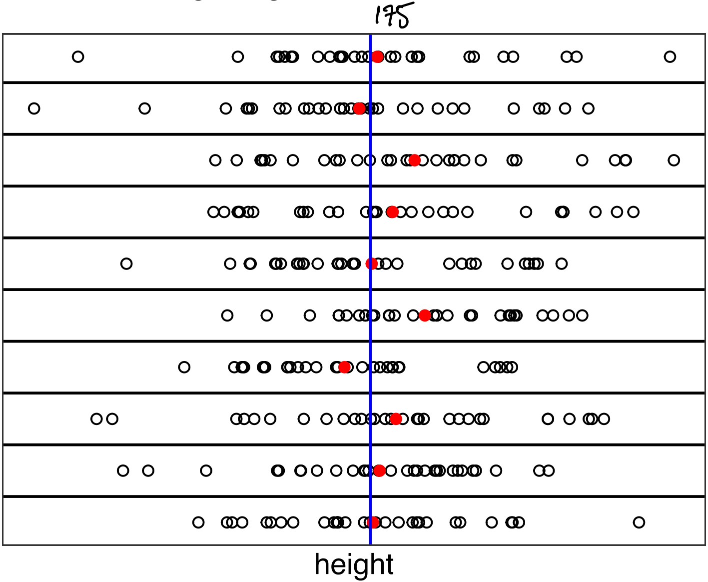
P-value measures how likely the observed data would be if H_0 is true.

- t_{obs} is computed under the assumption that H_0 is true
- *p-value* quantifies how strong is the evidence against H_0



Statistical testing: comments

What is the meaning of significance level $\alpha = 0.05$*?*



Statistical testing: comments

What is the meaning of significance level $\alpha = 0.05$?

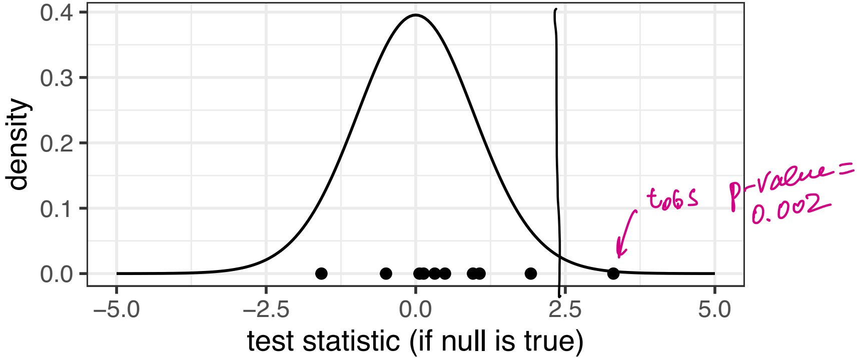
sample	tobs	pvalue
1	0.1315220	0.8962706
2	0.4887261	0.6287113
3	1.0671139	0.2947238
4	-1.5760942	0.1258517
5	3.3038490	$\boxed{0.0025415}$ $\rightarrow$ reject H_0
6	0.0624155	0.9506601
7	0.9570186	0.3464687
8	1.9237988	0.0642451
9	-0.4967206	0.6231308
10	0.3175223	0.7531223

$\mu = 175$

Statistical testing: comments

What is the meaning of significance level $\alpha = 0.05$?

- Different samples produce different test statistics (and conclusions)
- If the null hypothesis is true, we will incorrectly reject null 5% times



Statistical testing: type I error

$N(175, 20^2)$

Type I error: we rejected H_0 when H_0 was true.

- This happens with probability $\alpha = 0.05 \Rightarrow 5\%$
- By selecting smaller α we decrease the chance of type I error

$\alpha = 0.01 \Rightarrow 1\%$

What if we make α very small? $\alpha = 0.00001$?

- We will never reject H_0, even if H_a was true

Statistical testing: power

What happens if H_a is true? $\cancel{\mu = 175}$ $H_0: \mu = 175$

$H_a: \mu \neq 175$

```
sample = rnorm(n = 30, mean = 180, sd = 20)
```

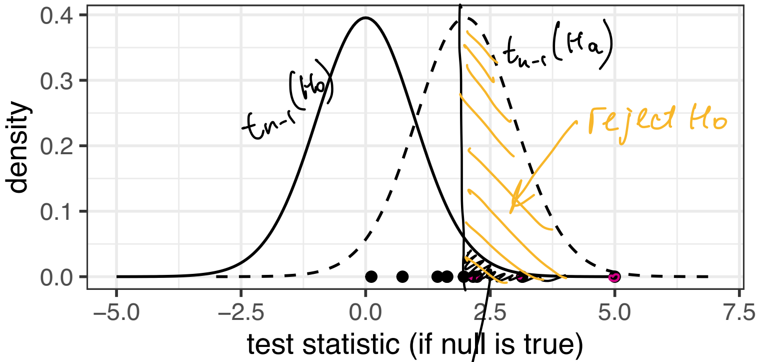
$\boxed{\mu = 180}$

sample	tobs	pvalue
1	1.6294412	0.1140357
2	1.9704657	0.0583987
3	2.2337574	0.0333752
4	0.1134225	0.9104772
5	4.9992181	0.0000254
6	1.4434491	0.1596104
7	2.1564185	0.0394791
8	3.1393408	0.0038726
9	0.7394223	0.4655943
10	1.6371530	0.1124061

reject H_0

Statistical testing: power

What happens if H_a is true?

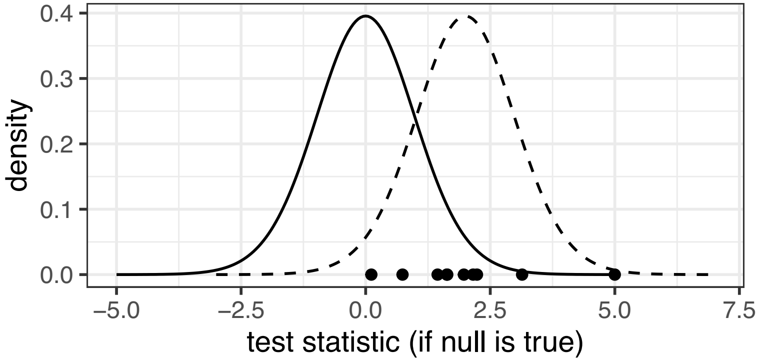


$\alpha = 0{,}05$

$\alpha \downarrow \Rightarrow$ TIE $\downarrow$ power $\downarrow$

Statistical testing: power

The **power** of the test is the probability of correctly rejecting H_0 when the alternative is true.



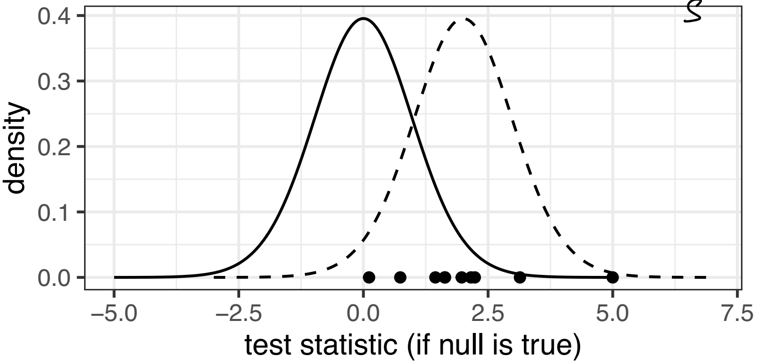
Statistical testing: power

- Smaller α means ~~higher~~ lower power
- Power is high if true μ is far from the value that we consider in H_0

$$t_{obs} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \rightarrow \text{large}$$

$$t_{obs} = \frac{(\bar{x} - \mu) \cdot \sqrt{n}}{s}$$

- Larger n implies higher power



Statistical testing: type II error

Type II error: we failed to reject H_0 when H_a was true.

- Type II error is usually denoted by β

$$\beta = 1 - \textit{power}$$

- Small α implies large β

$$\alpha \downarrow \Rightarrow \text{power} \downarrow \Rightarrow \beta \uparrow$$

Statistical testing: type I and II errors

$$P(\text{fail to reject } H_0 \mid H_0 \text{ true}) = 1 - P(\text{reject } H_0 \mid H_0 \text{ true})$$
$$= 1 - \alpha$$

$$\alpha = P(\textit{reject } H_0 | H_0 \textit{ is true})$$

$$\beta = P(\textit{fail to reject } H_0 | H_a \textit{ is true})$$

	H_0 is true	H_a is true
Reject H_0	Type I error (α)	Correct decision "power" ($1-\beta$)
Fail to reject H_0	Correct decision ($1-\alpha$)	Type II error (β)

Exercise

$$1-\alpha = P(\text{fail to reject } H_0 \mid \text{true } H_0)$$

H_0 : patient doesn't have disease

H_a : patient does have disease

Are these type I, type II errors or power?

- Test detected the disease but the patient is not sick $\underline{\text{I E}}$
- Test detected the disease and the patient is sick power
- Test did not detect the disease and the patient is sick II E

$$P(\text{rejected } H_0 \mid \text{true } H_0) = \alpha$$

$$P(\text{fail to reject } H_0 \mid \text{true } H_a) = \beta$$

$$P(\text{rejected } H_0 \mid \text{true } H_a) = 1-\beta$$

Statistical testing: additional discuttion

For more information watch some general advice about statistical tests video.

Statistical testing and confidence intervals

We want to study the average life expectancy in Canada μ.

We take a sample of n people, record their ages of death and compute the sample mean and standard deviation:

$$x_1, \dots, x_n \Rightarrow \bar{x}, s$$

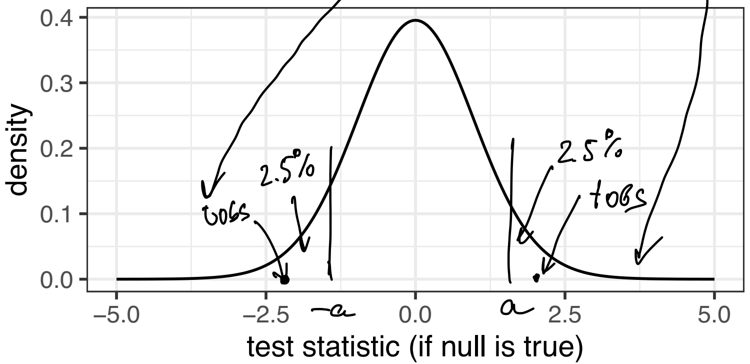
Statistical testing and confidence intervals

Suppose we tested two-sided hypothesis $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ and *p-value* < 0.05.

If a is a quantile that corresponds to 2.5% tail of t-distribution then

$t_{obs} > a$ $\quad\quad$ $t_{obs} < -a$

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} > a \quad \text{or} \quad \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < -a$$



Statistical testing and confidence intervals

$$\frac{\bar{x}-\mu_0}{s/\sqrt{n}} > a \Rightarrow \bar{x}-\mu_0 > a\cdot\frac{s}{\sqrt{n}} \Rightarrow \mu_0 < \bar{x} - a\cdot\frac{s}{\sqrt{n}}$$

(right tail)

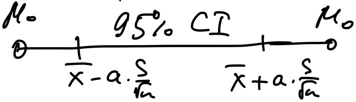
This is equivalent to

$\mu \in [a, b]$ 95%

$\mu \in [a, +\infty)$ 95%

$$\mu_0 < \bar{x} - a\cdot\frac{s}{\sqrt{n}} \qquad \text{or} \qquad \mu_0 > \bar{x} + a\cdot\frac{s}{\sqrt{n}}$$

Recall that 95% confidence interval is



$$\left[\bar{x} - a\cdot\frac{s}{\sqrt{n}}, \bar{x} + a\cdot\frac{s}{\sqrt{n}}\right]$$

$[\bar{x} - a\cdot\frac{s}{\sqrt{n}}, +\infty)$ a 5% instead of 2.5%

Thus CI does not cover μ_0!

t_{obs} is extreme $\Rightarrow$ CI does not cover μ_0

$H_0: \mu = \mu_0 \Rightarrow$ reject H_0

$H_a: \mu \neq \mu_0$

$\alpha = 0.05 \Rightarrow$ 95% CI

$\alpha = 0.1 \Rightarrow$ 90% CI

Statistical testing and confidence intervals

There is a connection between statistical testing and CI.

- If 95% CI does not cover μ_0, then we can reject $H_0 : \mu = \mu_0$ in favor of $H_a : \mu \neq \mu_0$ $\alpha = 0.05$
- If 95% CI covers μ_0, we do not have enough evidence to reject H_0. $\alpha = 0.05$

Exercise

We want to test if the average life expectancy in Canada is exactly 75. We collect a sample of size 30.

```
mean(ages)
```

```
## [1] 78.82458
```
$\simeq 78.8$

2.5% tail of t_{n-1} distribution

$$\left[\bar{x} - a \cdot \frac{s}{\sqrt{n}}, \bar{x} + a \cdot \frac{s}{\sqrt{n}}\right]$$

$$\left[78.8 - \cancel{a} \frac{9.2}{\sqrt{30}}, 78.8 + \frac{9.2}{\sqrt{30}}\right]$$

t_{29}

```
sd(ages)
```

```
## [1] 9.241208
```
$\simeq 9.2$

Find 95% confidence interval. Does it cover 75? What conclusion can we make?

$[75.518, 82.131]$

$[75.5, 82.1]$ $[75.37, 82.28]$

Exercise

```
t.test(ages, mu = 75, alternative = "two.sided",
       conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  ages
## t = 2.2668, df = 29, p-value = 0.03104
## alternative hypothesis: true mean is not equal to 75
## 95 percent confidence interval:
##  75.37386 82.27531
## sample estimates:
## mean of x
##  78.82458
```

reject H_0

< 0.05

$\not\ni 75 \Rightarrow$ reject $H_0: \mu = 75$ in favor of $H_a: \mu \neq 75$

Exercise

We want to test if the average life expectancy in Canada is exactly 75. We collect a sample of size 30.

```
mean(ages)
```

0.5% in one tail of t_{29}

$$[78.8 - \_\_\ \frac{9.2}{\sqrt{30}},\ 78.8 + \_\_\ \frac{9.2}{\sqrt{30}}]$$

```
## [1] 78.82458
```

```
sd(ages)
```

```
## [1] 9.241208
```

Now find 99% confidence interval. Does it cover 75? What conclusion can we make?

99% CI : [74.16, 83.44]   [74.18, 83.46]

Exercise

```
t.test(ages, mu = 75, alternative = "two.sided",
       conf.level = 0.99)
```

fail to reject Ho

```
##
##  One Sample t-test
##
## data:  ages
## t = 2.2668, df = 29, p-value = 0.03104 > 0.01
## alternative hypothesis: true mean is not equal to 75
## 99 percent confidence interval:
##  74.17399 83.47517 ∋ 75 ⇒ fail to reject H0
## sample estimates:
## mean of x
##  78.82458
```

Statistical testing for two groups

30 chickens were fed with sunflower seeds for 1 month. Their weight gain (in gramms) was recorded.

```
## [1] 267.9340 194.8606 219.3836 197.3097 131.1470
```

The same chickens were fed with corn for 1 month. The new weight gain was recorded.

```
## [1] 300.0809 178.0380 214.4870 181.4001 142.8363
```

Is any diet better for the weight gain?

Statistical testing for two groups

Let's first check the summary statistics.

```
mean(gain1)
```

```
## [1] 206.6387
```

```
sd(gain1)
```

```
## [1] 39.76608
```

```
mean(gain2)
```

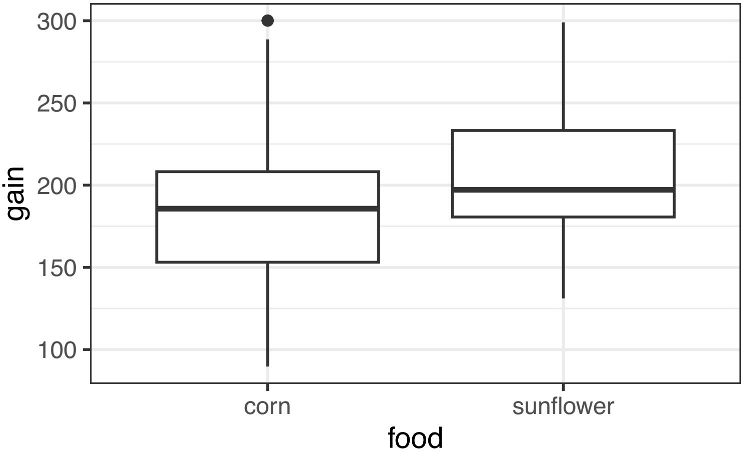
```
## [1] 185.5139
```

```
sd(gain2)
```

```
## [1] 48.01734
```

Statistical testing for two groups

Let's also compare the boxplots.



How to check if this difference is statistically significant?

Statistical testing for two groups

We are given two samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ with **matching observations**.

Do we observe significant difference in $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$?

- We create a sample that shows the difference in measurements:

$$d_1, \ldots, d_n \text{ where } d_i = \underset{\swarrow\, \text{St}}{x_i} - \underset{\swarrow\, \text{C}}{y_i}$$

- We perform statistical test on differences testing $H_0 : \mu_d = 0$ vs $H_a : \mu_d \neq 0$.

$\mu_d > 0$

$$d_1 \ldots d_n \to \bar{d},\ S_d \to t_{obs} = \frac{\bar{d} - \mu_d}{S_d/\sqrt{n}}$$

Statistical testing for two groups

Is any diet better for the weight gain?

chicken	diet1	diet2	difference
1	267.9340	300.08089	-32.146910
2	194.8606	178.03800	16.822614
3	219.3836	214.48697	4.896612
4	197.3097	181.40011	15.909640
5	131.1470	142.83634	-11.689317
6	179.2503	189.43962	-10.189343
7	180.2855	89.75207	90.533434
8	197.0343	253.27774	-56.243413
9	255.0013	187.66267	67.338602
10	238.1588	288.63058	-50.471796

Exercise

Find t_{obs} and *p-value*.

$H_0: \mu_d = 0$

$H_a: \mu_d \neq 0$

```
diff = gain1 - gain2
mean(diff)
```

```
## [1] 21.12483
```

$= \bar{d}$

```
sd(diff)
```

```
## [1] 46.39968
```

$= S_d$

$$t_{obs} = \frac{\bar{d} - \mu_d}{S_d / \sqrt{n}} = \frac{21.1 - 0}{46.4/\sqrt{30}} =$$

2.494

2.48

What will change if we want to check if corn is better than sunflower seed for gaining weight?

Exercise

```r
t.test(gain1, gain2, mu = 0, alternative = "two.sided",
       paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  gain1 and gain2
## t = 2.4937, df = 29, p-value = 0.0186 < 0.05
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##   3.798903 38.450752 ∌ 0 => reject Ho
## sample estimates:
## mean of the differences
##                21.12483
```

$H_0: \mu_d = 0$

Statistical testing for two groups

This procedure is called *paired t-test*

- Paired t-test is suitable only if there is a matching between $x_1, \dots, x_n$ and $y_1, \dots, y_n$ samples
- Samples $x_1, \dots, x_n$ and $y_1, \dots, y_n$ are **not independent**

Exercise

Is paired t-test suitable for this analysis?

- Test the difference in heart rates before and after drinking coffee $X_1 \ldots X_n$ – HRT before $Y_1 \ldots Y_n$ – HRT after
- Test the difference in COVID-19 death rates between young and old people $X_1 \ldots X_n$ $Y_1 \ldots Y_n$
- Test if protein-based diet will increase your sport performance
- Test if average IQ in Canada is higher than in the US

$$\underbrace{X_1 \ldots X_{50}}_{\text{Canada}} \qquad \underbrace{Y_1 \ldots Y_{50}}_{\text{US}}$$

Confidence intervals for two groups

Given two samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ with **matching observations** you can also find confidence interval for the average difference μ_d.

TO DO

1. Module 9. The Effective Use of Statistical Tests and Module 10. Comparing Two Groups
2. Quiz 10 due Monday (March 27) @ 11:59 PM (EST)
3. Practice Problem Set 10