Inference for numerical data

E. Pastucha

September 2024

Central Limit Theorem Recap

If:

- Samples are independent,
- Sample size is bigger or equal to 30,
- Population distribution is not strongly skewed.

Then:

Point estimate distribution can be approximated by Normal Distribution with mean equal to population mean and standard deviation equal to Standard Error.

Confidence Intervals

95% confidence interval

$$\textit{point estimate} \pm 1.96 \cdot SE$$

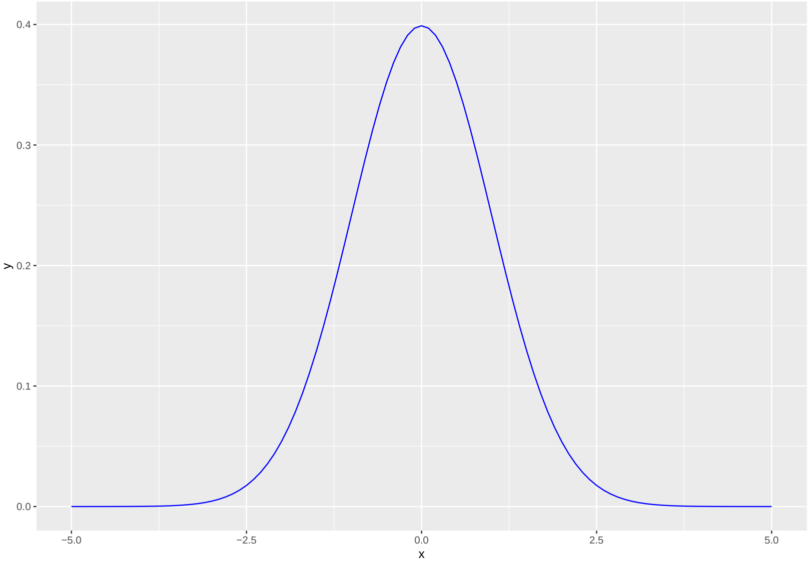
99% confidence interval

$$\textit{point estimate} \pm 2.58 \cdot SE$$

Normal distribution vs Student-t distribution

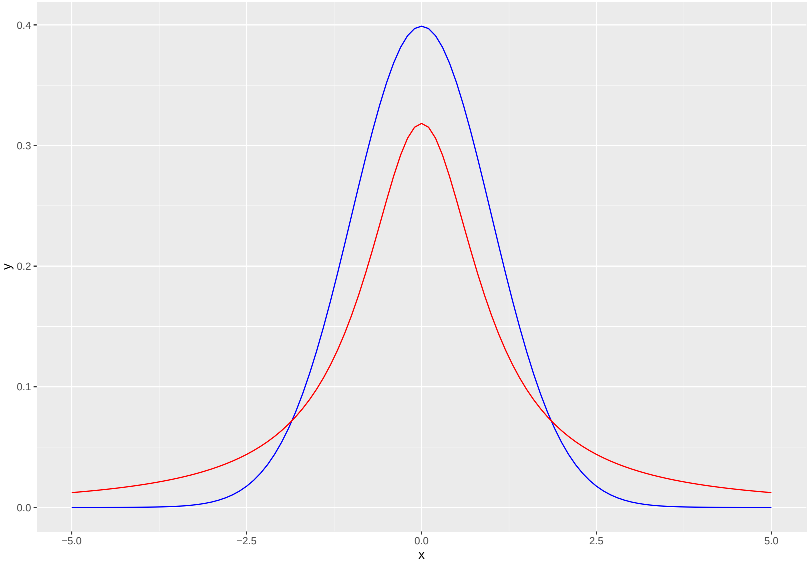
What if the sample is not even 30?

Normal distribution vs Student-t distribution



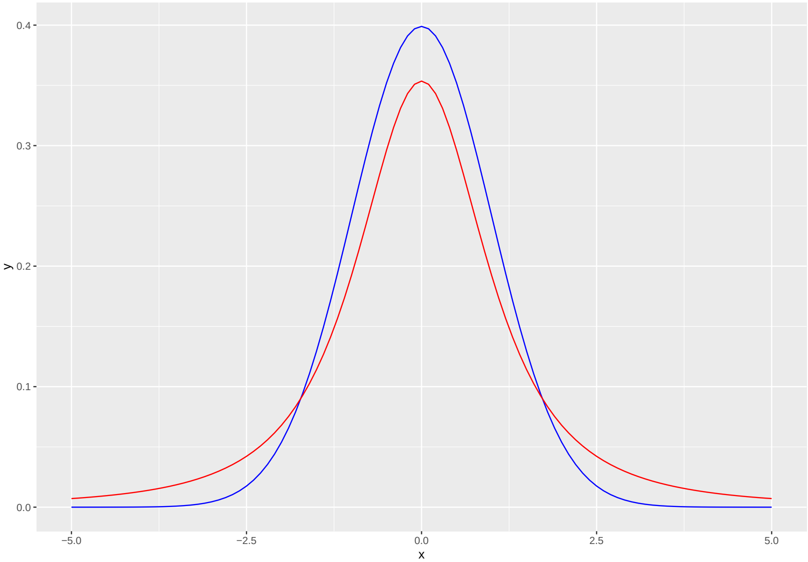
Normal distribution vs Student-t distribution

N(0,1) T(1)



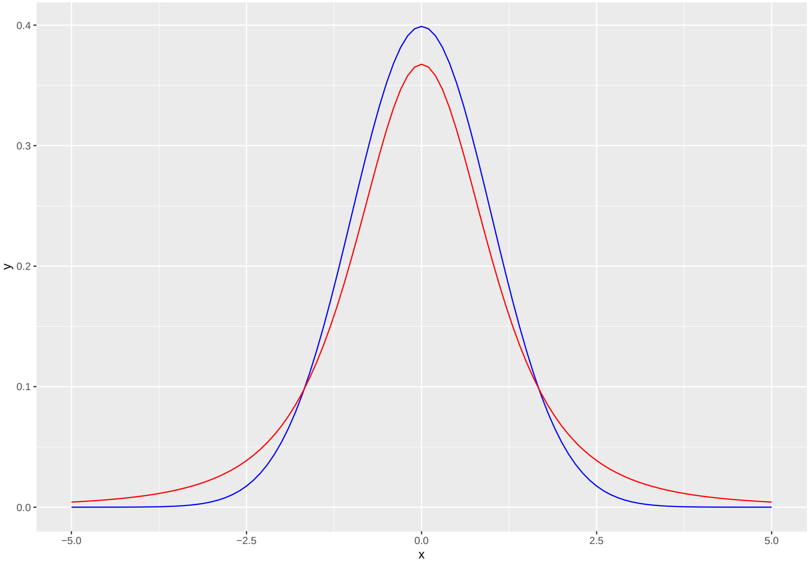
Normal distribution vs Student-t distribution

N(0,1) T(2)



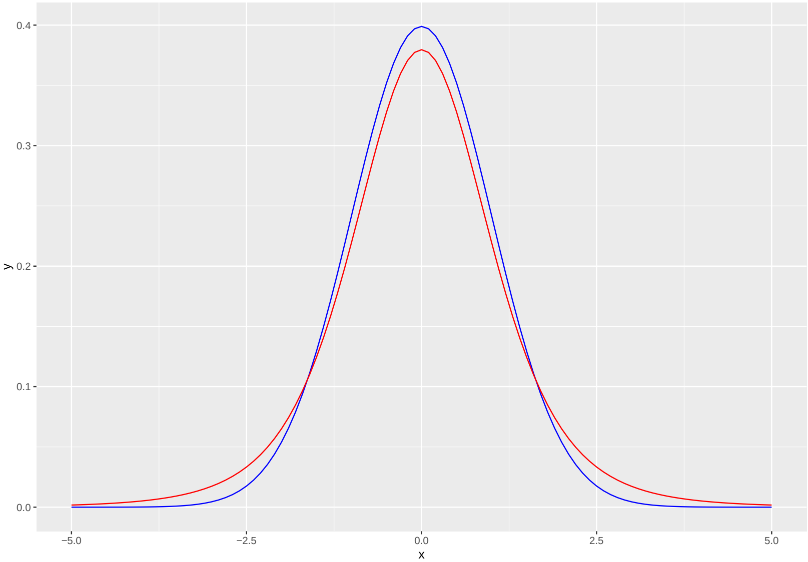
Normal distribution vs Student-t distribution

N(0,1) T(3)



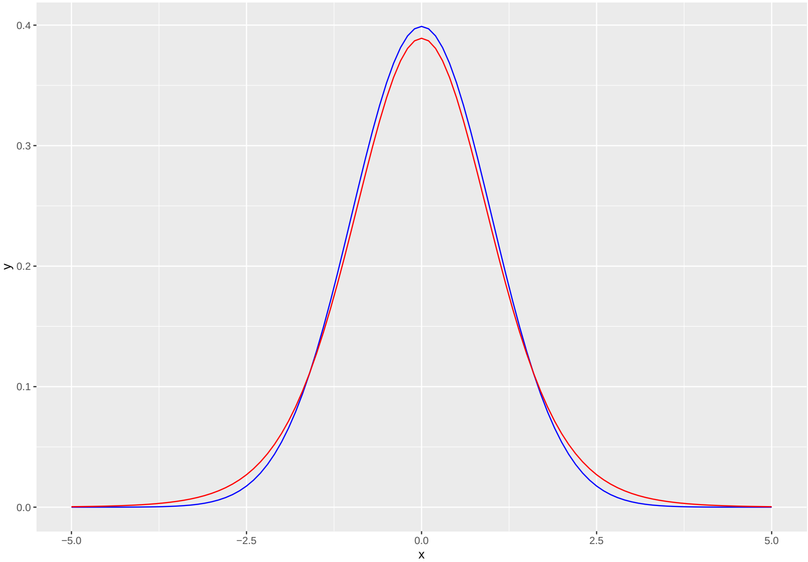
Normal distribution vs Student-t distribution

N(0,1) T(5)



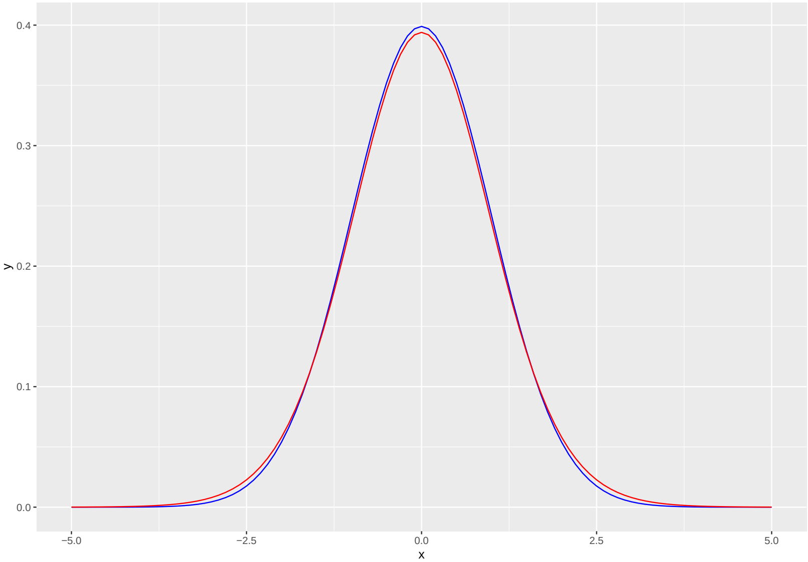
Normal distribution vs Student-t distribution

N(0,1) T(10)



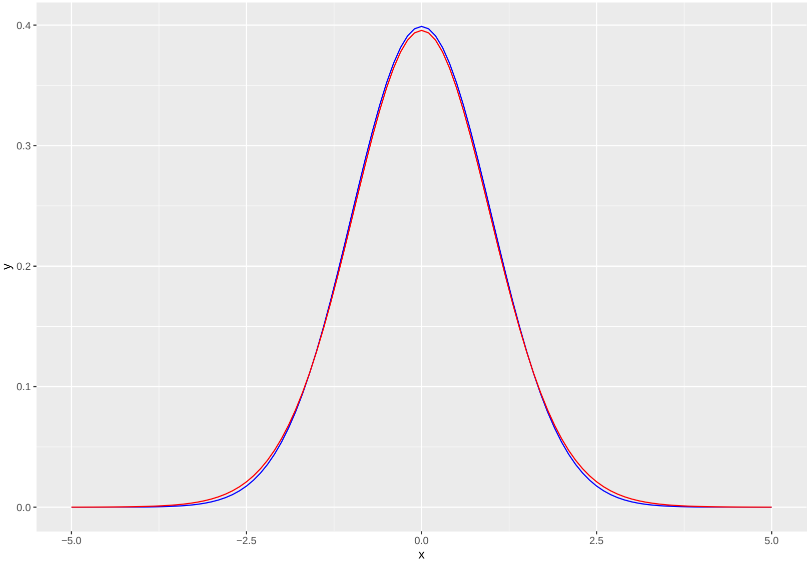
Normal distribution vs Student-t distribution

N(0,1) T(20)



Normal distribution vs Student-t distribution

N(0,1) T(30)



Normal distribution vs Student-t distribution

$Z\text{-}score = \frac{x-\bar{x}}{s} \longrightarrow T\text{-}score = \frac{x-\bar{x}}{s}$

To read out probability:

$mean = 0, SD = 1 \longrightarrow df = n - 1$

T-score is sometimes also called T-statistic.

Central Limit Theorem Recap

If:

- Samples are independent,
- ~~Sample size is bigger or equal to 30~~,
- Population distribution is not strongly skewed.

Then:

Normalized point estimate distribution can be approximated by Student-T distribution with degrees of freedom equal to number of cases minus 1.

Hypothesis Testing

1) Set-up the hypothesis:

- H_0 : Position of no change
- H_A : Alternative to H_0 - One-sided and two-sided hypothesis test.

2) Assume threshold values:

- α-significance level - typicaly 0.05

2.1) Check CLT conditions

3) Calculate the Results

p-value - probability of something as different as point estimate existing within H_0 sampling distribution.

4) Draw conclusions

Hypothesis Testing

1) Set-up the hypothesis:

- H_0 : Position of no change
- H_A : Alternative to H_0

One-sided and two-sided hypothesis test.

2) Assume threshold values:

- α-significance level - typicaly 0.05

Hypothesis Testing

3) Calculate the Results

- p-value - probability of something as different as point estimate existing within H_0 sampling distribution.

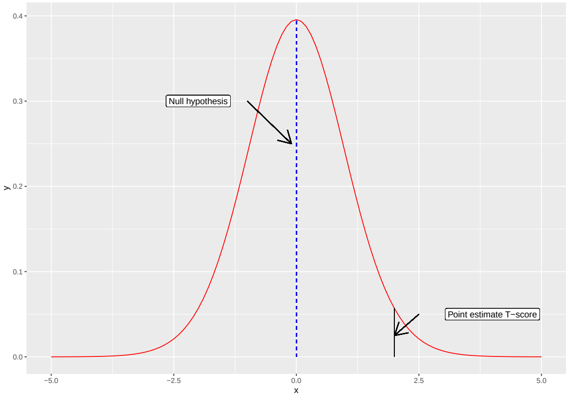
First calculate point estimate and SE, then:

$$T\text{-}score = \frac{\text{point estimate} - \text{null hypothesis}}{SE}$$

Hypothesis Testing

3) Calculate the Results

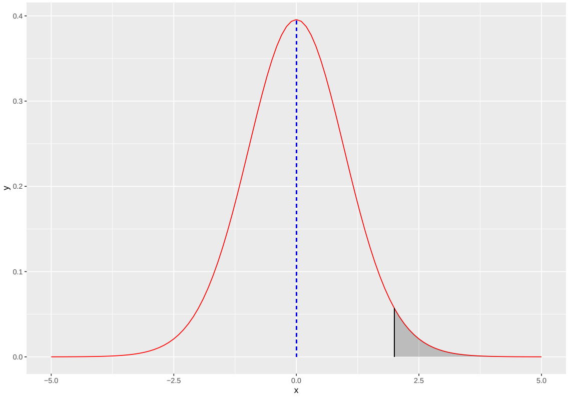
- p-value - probability of something as different as point estimate existing within H_0 sampling distribution.



Hypothesis Testing

3) Calculate the Results

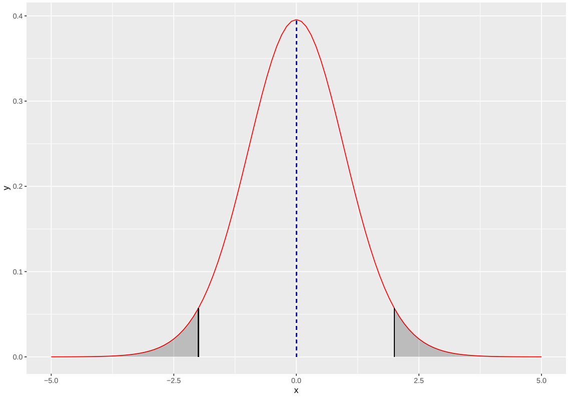
- p-value - probability of something as different as point estimate existing within H_0 sampling distribution.



Hypothesis Testing

3) Calculate the Results

- p-value - probability of something as different as point estimate existing within H_0 sampling distribution.



Hypothesis Testing

4) Draw conclusions

$$p - \text{value} > \alpha - \text{significance level}$$

We keep the Null Hypothesis and reject the Alternative (We failed to reject null hypothesis). The difference is not significant.

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

We reject Null Hypothesis in favour of Alternative. The difference is significant.

Hypothesis Testing

Type 1 and Type 2 hypothesis errors.

	H_0 is True	H_A is True
Reject H_0	Type I Error	correct
Accept H_0	correct	Type II error

One sample t-test

Within Hypothesis we compare a sample point estimate to a **set value**.

The sample must fulfill conditions of **CLT**.

One sample t-test

$$H_0 : \mu = \text{set value}$$

$$H_A : \mu \neq \text{set value}$$

One sample t-test

$$SE = \frac{s}{\sqrt{n}}$$

One sample t-test

Kiwi dataset.

We want to check, if mean male Great Spotted height is 45 cm.

One sample t-test

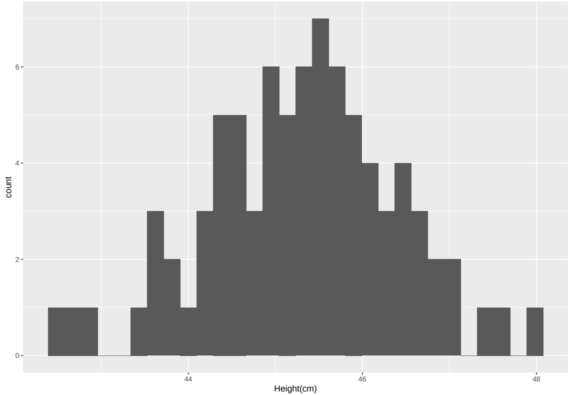
H_0 : population mean height of male kiwis is 45 cm

H_A : population mean height of male kiwis is not 45 cm

One sample t-test

CLT Conditions?

Number of male Great Spotted Kiwis is 82.



One sample t-test

```
(point_estimate <- mean(kiwi_GS_M$`Height(cm)`))
```

```
## [1] 45.30854
```

```
(SE <- sd(kiwi_GS_M$`Height(cm)`)/sqrt(nrow(kiwi_GS_M)))
```

```
## [1] 0.1213097
```

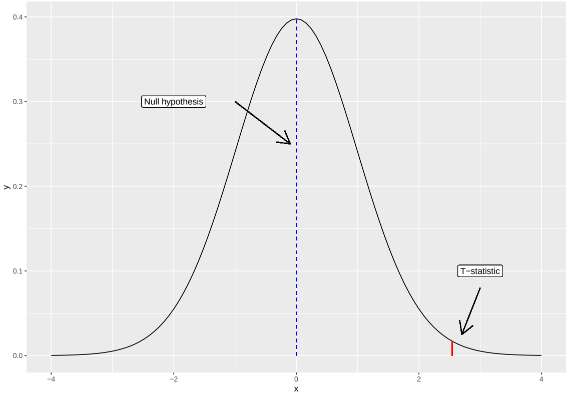
```
(df <- nrow(kiwi_GS_M) - 1)
```

```
## [1] 81
```

One sample t-test

```
(t_statistic <- (point_estimate - 45)/SE)
```

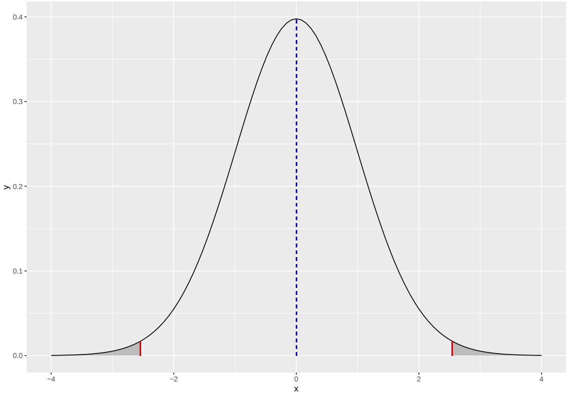
```
## [1] 2.543379
```



One sample t-test

```r
(p_value <- 2*(1- pt(t_statistic, df = df)))
```

```
## [1] 0.01288087
```



One sample t-test

We reject null hypothesis in favour of Alternative.

Mean Male Great Spotted Kiwi is not 45 cm tall.

One sample t-test

Other way of conducting the t test:

```r
t.test(kiwi_GS_M$`Height(cm)`, mu = 45)
```

```
##
##  One Sample t-test
##
## data:  kiwi_GS_M$`Height(cm)`
## t = 2.5434, df = 81, p-value = 0.01288
## alternative hypothesis: true mean is not equal to 45
## 95 percent confidence interval:
##  45.06717 45.54990
## sample estimates:
## mean of x
##  45.30854
```

One sample t-test

Within Hypothesis we compare a sample point estimate to a set value.

The sample must fulfill conditions of CLT.

Difference of means t-test

Used to compare two different groups.

Category or sample.

Both samples must fulfill conditions of **CLT**.

Difference of means t-test

- Male and Females
- Regular American Males and NBA players
- Great Spotted Kiwis and Tokoeka Kiwis
- SDU and UCL students

Difference of means t-test

$$H_0 : \mu_1 = \mu_2 \longrightarrow \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 \neq \mu_2 \longrightarrow \mu_1 - \mu_2 \neq 0$$

Difference of means t-test

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Difference of means t-test

Degrees of freedom

Welch-Satterthwaite equation

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

or. . . smaller of the two.

Difference of means t-test

Kiwi dataset.

We want to check, if mean male Great Spotted are same height as mean male Tokoeka.

Difference of means t-test

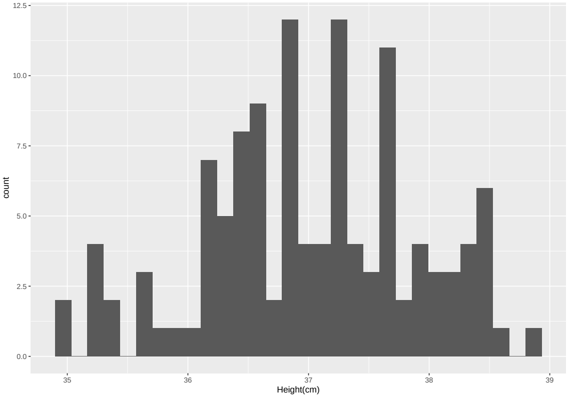
$$H_0 : mean_{GS} - mean_{Tokoeka} = 0$$

$$H_A : mean_{GS} - mean_{Tokoeka} \neq 0$$

Difference of means t-test

CLT Conditions?

Number of male Great Spotted Kiwis is 82, Number of male Tokoekas is 119.



Difference of means t-test

```
(point_estimate <- mean(kiwi_GS_M$`Height(cm)`) -
   mean(kiwi_T_M$`Height(cm)`) )
```

```
## [1] 8.295091
```

```
(SE<-sqrt((sd(kiwi_GS_M$`Height(cm)`)^2/nrow(kiwi_GS_M)) +
           (sd(kiwi_T_M$`Height(cm)`)^2/nrow(kiwi_T_M))))
```

```
## [1] 0.1452033
```

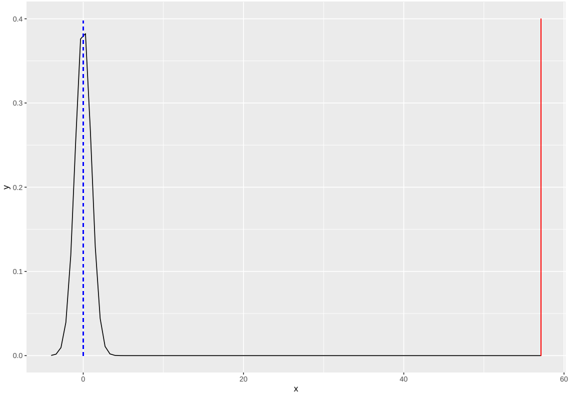
```
(df <- nrow(kiwi_GS_M) - 1)
```

```
## [1] 81
```

Difference of means t-test

```
(t_statistic <- (point_estimate - 0)/SE)
```

```
## [1] 57.12745
```



Difference of means t-test

```
(p_value <-2*(1-pt(t_statistic, df = df)))
```

```
## [1] 0
```

We reject null hypothesis in favour of Alternative. Mean Male Great Spotted Kiwis are of different height than Tokoekas.

Difference of means t-test

Other way of conducting the t test:

```
t.test(kiwi_GS_M$`Height(cm)`, kiwi_T_M$`Height(cm)`)
```

```
##
##  Welch Two Sample t-test
##
## data:  kiwi_GS_M$‘Height(cm)‘ and kiwi_T_M$‘Height(cm)‘
## t = 57.127, df = 147.33, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  8.008141 8.582041
## sample estimates:
## mean of x mean of y
##  45.30854  37.01345
```

Note different df value.

Difference of means t-test

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

```
M.v <- var(kiwi_GS_M$`Height(cm)`)
M.n <- nrow(kiwi_GS_M)
T.v <- var(kiwi_T_M$`Height(cm)`)
T.n <- nrow(kiwi_T_M)
(((M.v/M.n) + (T.v/T.n))^2)/
  ((((M.v/M.n)^2)/(M.n-1)) + (((T.v/T.n)^2)/(T.n-1)))
```

```
## [1] 147.3307
```

Difference of means t-test

Other way of conducting the t test:

```
hm <- kiwi %>% filter(Species_code == 'GS' |
                        Species_code == 'Tok') %>%
  filter(Gender == 'M')
t.test(`Height(cm)`~Species_code, data = hm)
```

```
##
##  Welch Two Sample t-test
##
## data:  Height(cm) by Species_code
## t = 57.127, df = 147.33, p-value < 2.2e-16
## alternative hypothesis: true difference in means betwee
## 95 percent confidence interval:
##  8.008141 8.582041
## sample estimates:
##  mean in group GS mean in group Tok
##          45.30854          37.01345
```

Difference of means t-test

Used to compare two different groups.

Category or sample.

Both samples must fulfill conditions of **CLT**.

Paired data t-test

Used to compare two observations within same cases.

- Consumer grades of different product types,
- Prices of products in between shops,

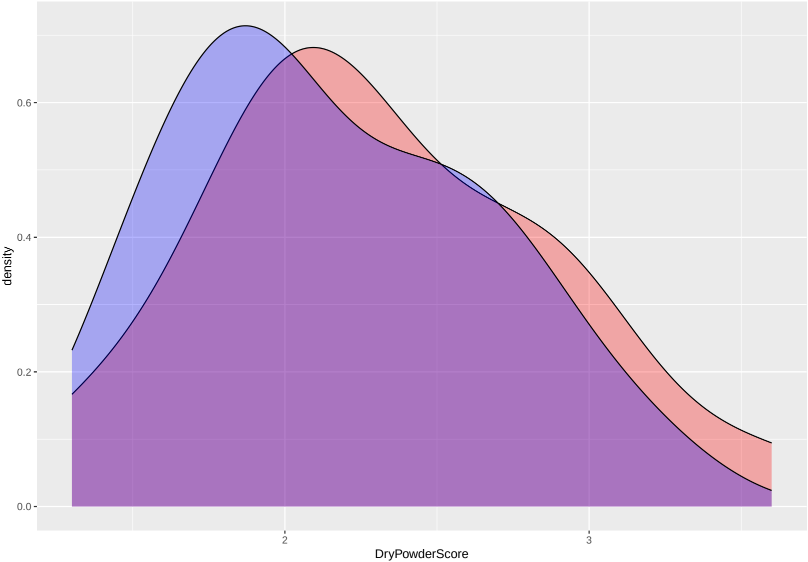
Paired data t-test

H_0: mean DryPowderScore is equal to mean PowderInOilScore

H_A: mean DryPowderScore is not equal to mean PowderInOilScore

ID	DryPowderScore	PowderInOilScore
1	2.0	1.9
2	2.8	2.4
3	1.3	1.5
4	1.8	1.8
5	1.9	1.8

Paired data t-test



Paired data t-test

If we disregard paired observation and use simple difference of means scenario:

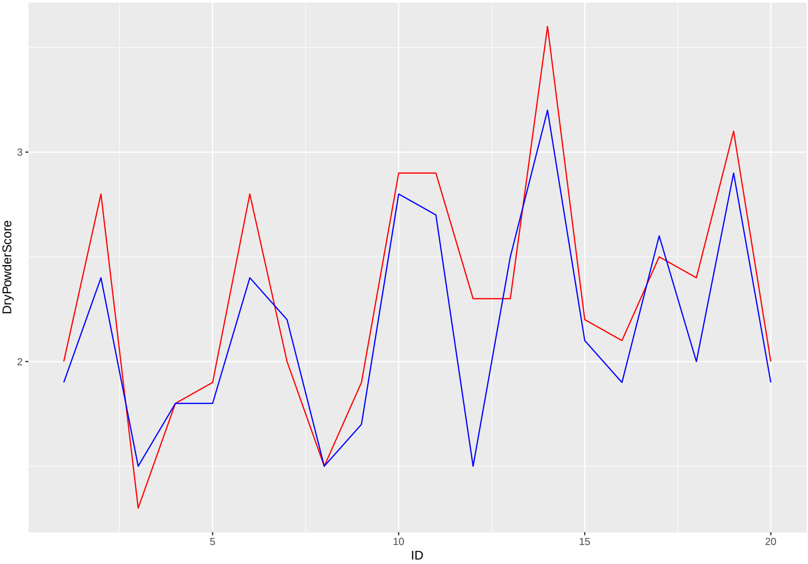
```r
t.test(anti$DryPowderScore, anti$PowderInOilScore)
```

```
##
##  Welch Two Sample t-test
##
## data:  anti$DryPowderScore and anti$PowderInOilScore
## t = 0.88603, df = 37.417, p-value = 0.3813
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  -0.1928938  0.4928938
## sample estimates:
## mean of x mean of y
##     2.315     2.165
```

we would keep null hypothesis.

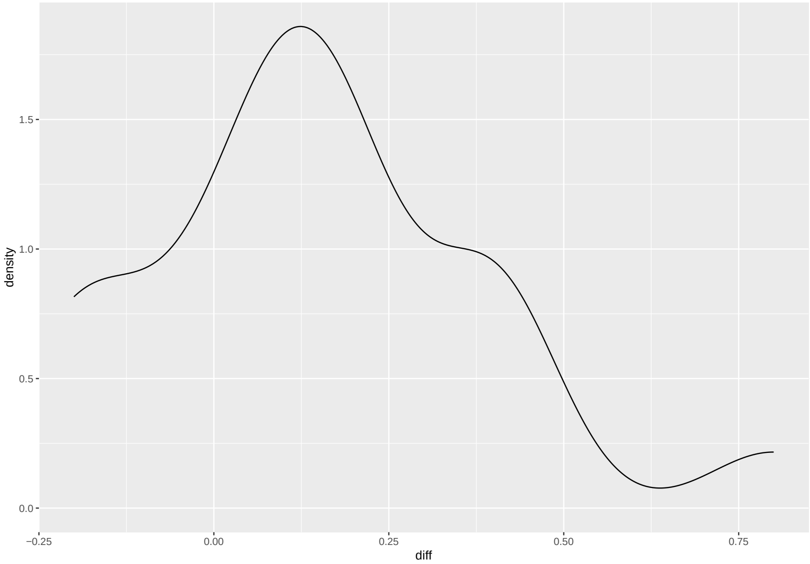
Paired data t-test

But if we look at the data in a different way



Paired data t-test

So we compute difference variable:



Paired data t-test

$$H_0 : \mu_{diff} = 0$$

$$H_A : \mu_{diff} \neq 0$$

Paired data t-test

$$SE = \frac{s_{diff}}{\sqrt{n}}$$

Degrees of freedom is again number of cases minus 1.

Paired data t-test

Back to antierspirant data

```
(point_estimate <- mean(anti$diff))
```

```
## [1] 0.15
```

```
(SE <- sd(anti$diff)/sqrt(nrow(anti)))
```

```
## [1] 0.05548826
```

Paired data t-test

Back to antierspirant data

```
(df <- nrow(anti) - 1)
```

```
## [1] 19
```

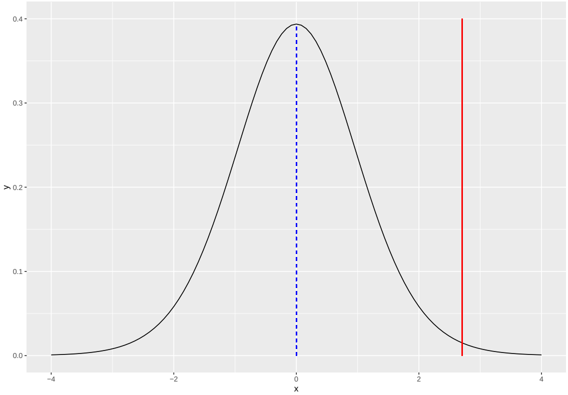
```
(t_statistic <- (point_estimate - 0)/SE)
```

```
## [1] 2.703274
```

Paired data t-test

```r
(p_value <- 2*(1- pt(t_statistic, df = df)))
```

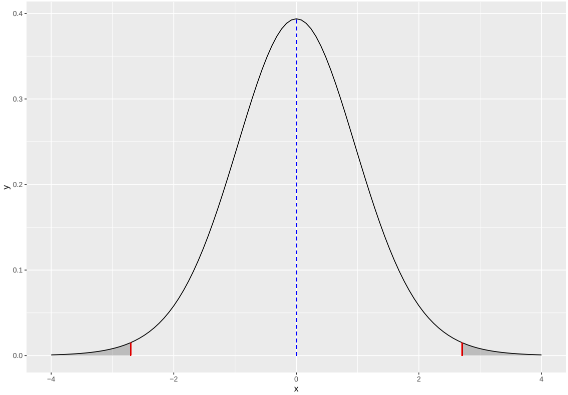
```
## [1] 0.01408939
```



Paired data t-test

```
(p_value <- 2*(1- pt(t_statistic, df = df)))
```

```
## [1] 0.01408939
```



What would you do?

Lars was 6 years old, when on a fishing trip with his dad he caught 3 Atlantic herrings. All were almost 50 cm long, 1.5 kg heavy. Since then (and Lars is now 50) he always said, that the average Atlantic herring is 50 cm long and weights 1.5 kg. How would you test this hypothesis? How would you gather data, formulate hypothesis and which statistical test would you use?

What would you do?

Who loves licorice more: Danes or Swedes? An online survey was conducted throughout Sweden and Denmark to find licorice king. The participants were asked to grade, on the scale 1 – 10, how much do they love licorice. You are now in possession of the dataset. How do you decide who is the Scandinavian licorice king?

What would you do?

Gammeldags Ice-cream company wants to expand into the Irish market. They want to introduce all of the basic flavors of ice cream and one of the 'exotic' to the market flavors. Based on experience in Danish market they've decided to test two options: licorice and sea buckthorn. They recruited a sample of 100 true Irish inhabitants. Each of them was given sample of both flavors and graded each on the scale 1 – 10. How would you determine which of the flavors is more appropriate for the Irish market? Which test would you use? Are conditions fulfilled?

What would you do?

Nordisk Film Biografer is refurbishing their rooms. They have two chair types to choose from. The chairs have been tested by independent audience (12 hours LOTR extended marathon viewers). First chair received mean score of 7.8 (out of 10) with standard deviation of 1.3, the other 7.6 with standard deviation of 0.9. Both were tested by audiences of 50. Can you say with statistical certainty that one of them is better? Which test would you use to check this? How would you set up hypothesis? How would you calculate it?