

Parametric tests

Basics of Health Intelligent Data Analysis
PhD Programme in Health Data Science

1. One sample t-test
2. Paired samples t-test
3. Independent samples t-test
4. One-way ANOVA
5. Testing the assumptions of parametric tests

Cláudia Camila Dias

Pedro Pereira Rodrigues

Hypothesis test

1. Define the hypothesis

H_0 = null hypothesis - no effect on population

2. Set the significance level (α) - usually 0.05

3. Get the test statistic with the sample data and obtain the p-value: probability of obtaining the result we obtained or even more extreme, assuming H_0 is true.

5. Interpret the p-value:

if $p < \alpha$, we have enough evidence to reject H_0

if $p \geq \alpha$, we do not have enough evidence to reject H_0

One Sample T-test

With a sample of individuals we want to know if the average of their population is a certain value.

Define the hypothesis and significance level (alpha)

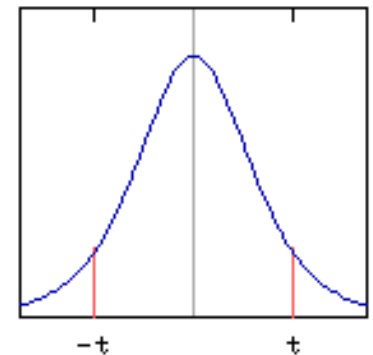
H_0 : Population mean equal to μ_1

H_1 : Population mean is different than μ_1

Get the test statistic with the sample data

$$t = \frac{\bar{X} - \mu_1}{s / \sqrt{n}}$$

which follows a t-distribution with $n-1$ degrees of freedom



One Sample T-test

EXAMPLE:

- A sample of 51 Portuguese pregnant women was taken
- Assuming this is a representative sample of Portuguese pregnant women, we would like to know if the average weight gain during pregnancy in Portuguese pregnant women is 10 kg.
- Among these 51 women, the average weight increase was 12.0549 kg and the standard deviation was 4.61932 kg.

One Sample T-test

We define the hypothesis and significance level:

$$H_0: \mu=10 \text{ kg}$$

$$\alpha = 0.05$$

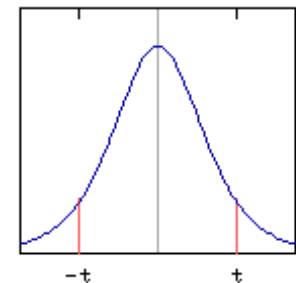
We get the test statistic with the sample data

$$t = \frac{12.0549 - 10}{\frac{4.61931}{\sqrt{51}}} = 3.177 \quad \text{which follows a t-distribution with 50 degrees of freedom}$$

[We obtain the p-value=0.003]

We interpret the value of p:

if $p < 0.05$, we have enough evidence to reject H_0



One Sample T-test

- Let's test if the average birth weight is 3300g

```
> t.test(nutricao$pn, mu=3300)
```

One Sample t-test

```
data:  nutricion$pn
t = 1.9101, df = 57, p-value = 0.06115
alternative hypothesis: true mean is not equal to 3300
95 percent confidence interval:
 3294.616 3528.143
sample estimates:
mean of x
 3411.379
```

Paired Sample T-test

With two paired samples of individuals, we want to know if the means of the two groups in the population are equal.

We define the Hypothesis

$$H_0: \mu_1 = \mu_2 \text{ ou } \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 \neq \mu_2 \text{ ou } \mu_1 - \mu_2 \neq 0$$

We set the significance level (alpha) - usually 0.05

We get the test statistic with the sample data

$$t = \frac{\text{média das diferenças}}{\text{erro padrão das diferenças}} \quad \text{which follows a t-distribution with } n-1 \text{ degrees of freedom}$$

We get the p-value

We interpret the value of p

Paired Sample T-test

Let's test if there is a variation in the global scale of development between 5 and 8 years

```
> t.test(develop$generalinc)
```

One Sample t-test

```
data: develop$generalinc
t = -4.0779, df = 38, p-value = 0.0002239
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -6.637987 -2.233808
sample estimates:
mean of x
-4.435897
```

```
> t.test(develop$general2, develop$general1, paired=T)
```

Paired t-test

```
data: develop$general2 and develop$general1
t = -4.0779, df = 38, p-value = 0.0002239
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.637987 -2.233808
sample estimates:
mean of the differences
-4.435897
```


Independent Sample T-test

With two independent samples of individuals we want to know if the means of the two groups in the population are equal.

We define the hypothesis

$$H_0: \mu_1 = \mu_2 \text{ ou } \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 \neq \mu_2 \text{ ou } \mu_1 - \mu_2 \neq 0$$

We get the test statistic with the sample data

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{sp \sqrt{1/n_1 + 1/n_2}}, \text{ which follows a t-distribution with } (n_1 + n_2 - 2) \text{ degrees of freedom}$$

We set the significance level (alpha) - usually 0.05

We get the p-value

We interpret the value of p

Independent Sample T-test

And if the variances are not equal, how to calculate sp ?

Levene's test tests the hypothesis that the variances are equal in both groups.

If they are not equal we cannot calculate the grouped estimate of standard deviations (sp) and we have to resort to a **modified t-test**.

Levene test

- Suppose we want to compare head circumference at birth in boys and girls

```
> summaryBy(pc ~ sexo, data=nutricao, FUN=var, na.rm=T)
      sexo  pc.var
1 Masculino 1.446640
2  Feminino 1.403448
```

```
> leveneTest(pc ~ sexo, data=nutricao)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.6973 0.4076
      51
```

Independent Sample T-test

- In this case the p value for the Levene test is $p > 0.05$ (*alpha*).
- For a significance level of 0.05, we accept equality of variance between the two groups.

```
> t.test(pc ~ sexo, data=nutricao, var.equal=T)

Two Sample t-test

data:  pc by sexo
t = 3.0652, df = 51, p-value = 0.003474
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.349531 1.676556
sample estimates:
mean in group Masculino  mean in group Feminino
          35.91304          34.90000
```

For $\alpha=0.05$, we have enough evidence to reject H_0 because **$p=0.003474 < 0.05$**

i.e. the differences between groups are statistically significant

Independent Sample T-test

- If the p-value for the Levene test is less than 0.05, we have to reject the hypothesis of equal variance between the two groups.

```
> t.test(pc ~ sexo, data=nutricao, var.equal=F)

Welch Two Sample t-test

data:  pc by sexo
t = 3.0589, df = 47.121, p-value = 0.003658
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3468441 1.6792429
sample estimates:
mean in group Masculino  mean in group Feminino
      35.91304           34.90000
```

For $\alpha=0.05$, we have enough evidence to reject H_0 because $p=0.003658 < 0.05$
i.e. the differences between groups are statistically significant

One-Way ANOVA – more than 2 groups

We define the null hypothesis :

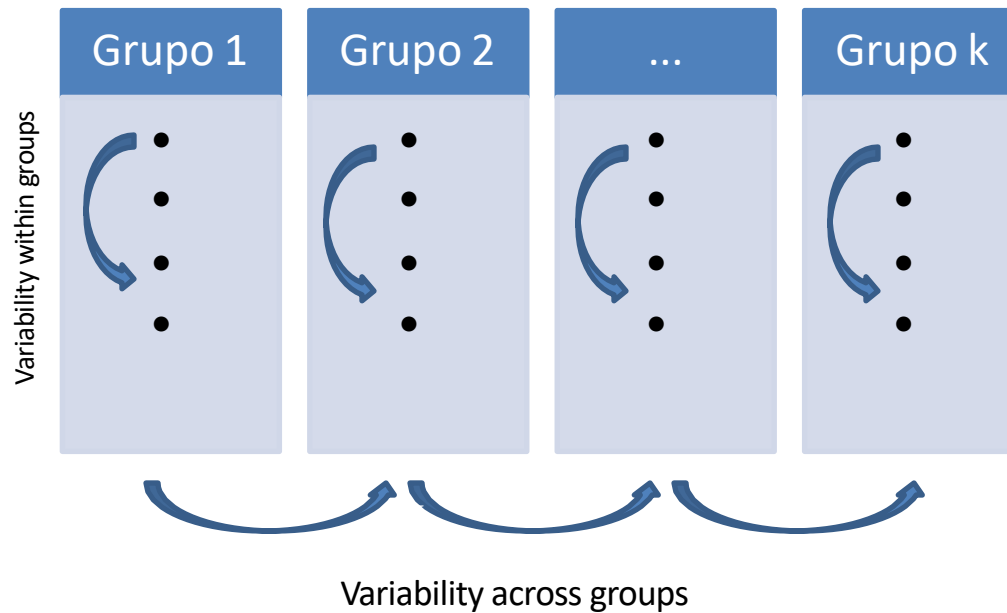
$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

We will have a set of i groups with n_i individuals each, a total of N individuals, an average of each group x_i and a common average X

Example: Weights in Kg of 3 groups of individuals from different ethnic groups.

Group 1: 72; 75; 73; 67; 76; 71; 71; 70; 78; 64	$X = 71,70$ kg
Group 2: 64; 74; 63; 69; 70; 62; 69; 65; 68; 73	$X = 67,70$ kg
Group 3: 58; 59; 61; 63; 66; 53; 68; 69; 61; 57	$X = 61,50$ kg
$X=66,97$ kg $i=3$ $n_1=10$ $n_2=10$ $n_3=10$ $N=30$	

One-Way ANOVA – more than 2 groups



One-Way ANOVA – more than 2 groups

- Between groups variability
 - *Between groups mean sum of squares (Between MS)*

$$\frac{\sum (\bar{X}_i - \bar{X})^2}{k-1}$$

- Within groups variability
 - *Within groups mean sum of squares (Within MS)*

Weighted mean of the groups variances

But beware: the variance has to be equal in all groups - Levene test

One-Way ANOVA – more than 2 groups

It is proved that if $\mu_1 = \mu_2 = \dots = \mu_k$, then, Between MS \sim Within MS

Otherwise, if $\mu_1 \neq \mu_2 \neq \dots \neq \mu_k$, Between MS will be higher than Within MS

So, to test $H_0: \mu_1 = \mu_2 = \dots = \mu_k$, the F statistic is calculated

We set the significance level (alpha) - usually 0.05

We get the test statistic with the sample data

$F = \text{Between MS} / \text{Within MS}$

We get the p-value

We interpret the value of p

One-Way ANOVA – more than 2 groups

Returning to the example:

Example: Weights in Kg of 3 groups of individuals from different ethnic groups.

$$H_0: \mu_1 = \mu_2 = \mu_3$$

Group 1: 72; 75; 73; 67; 76; 71; 71; 70; 78; 64	X = 71,70 kg
Group 2: 64; 74; 63; 69; 70; 62; 69; 65; 68; 73	X = 67,70 kg
Group 3: 58; 59; 61; 63; 66; 53; 68; 69; 61; 57	X = 61,50 kg
X=66,97kg i=3 n ₁ =10 n ₂ =10 n ₃ =10 N=30	

One-Way ANOVA – more than 2 groups

```
> summaryBy(peso ~ grupo, data=anova.example, FUN=c(mean,var))
      grupo peso.mean peso.var
1 Grupo1      71.7 17.34444
2 Grupo2      67.7 16.90000
3 Grupo3      61.5 25.83333
> leveneTest(peso ~ grupo, data=anova.example)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.2426 0.7863
      27
> oneway.test(peso ~ grupo, data=anova.example, var.equal=T)

One-way analysis of means

data:  peso and grupo
F = 13.19, num df = 2, denom df = 27, p-value = 0.0001009
```

One-Way ANOVA – more than 2 groups

Another example: comparing the weight of newborns in three different hospitals.

```
> summaryBy(birthwt ~ hospital, data=alcohol, FUN=c(mean,sd))
      hospital birthwt.mean birthwt.sd
1 Hospital A      3292.976    538.4036
2 Hospital B      3243.599    521.9887
3 Hospital C      3311.685    521.4014
> leveneTest(birthwt ~ hospital, data=alcohol)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2    0.172 0.8421
      451
> oneway.test(birthwt ~ hospital, data=alcohol, var.equal=T)

      One-way analysis of means

data:  birthwt and hospital
F = 0.73738, num df = 2, denom df = 451, p-value = 0.4789
```

One-Way ANOVA – more than 2 groups

- Another example: comparison of newborn weight in four gestational age categories.

```
> summaryBy(birthwt ~ gest4, data=alcohol, FUN=c(mean,sd), na.rm=T)
      gest4 birthwt.mean birthwt.sd
1    <38      2646.545    628.3576
2  38-39      3202.609    494.5665
3     40      3415.675    426.7055
4    >40      3360.366    418.9211
5    <NA>      3205.000   1025.3048
> leveneTest(birthwt ~ gest4, data=alcohol)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  3  5.1147 0.00173 **
      448
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> oneway.test(birthwt ~ gest4, data=alcohol, var.equal=F)

      One-way analysis of means (not assuming equal variances)

data:  birthwt and gest4
F = 26.412, num df = 3.00, denom df = 139.29, p-value = 1.38e-13
.
```

One-Way ANOVA – more than 2 groups

When differences are found, it is sometimes important to know which groups are different.

One solution is to make multiple comparisons (two by two). In the previous case, we need 6 comparisons:

$$H_0: \mu_1 = \mu_2$$

$$H_0: \mu_2 = \mu_3$$

$$H_0: \mu_1 = \mu_3$$

$$H_0: \mu_2 = \mu_4$$

$$H_0: \mu_1 = \mu_4$$

$$H_0: \mu_3 = \mu_4$$

So... why didn't we do this from the beginning?

One-Way ANOVA – more than 2 groups

Suppose H_0 true

We define the significance level $P(\text{type I error}) = 0.05$ for each test

We will have 6 tests to do

How likely are you to make at least one type I mistake?

Is it 0.05?

One-Way ANOVA – more than 2 groups

- The probability of making a type I error in at least one test is greater than 0.05!
- So after doing one way ANOVA and if we want to identify which groups are different we can make multiple comparisons but we have to correct the level of significance.
- There are several fixes implemented in R:
Bonferroni, Sheffe, Tukey, etc

One-Way ANOVA – more than 2 groups

- We can make multiple comparisons, in this case with Bonferroni correction

```
> with(alcohol, pairwise.t.test(birthwt, gest4, p.adjust.method="bonferroni"))
```

Pairwise comparisons using t tests with pooled SD

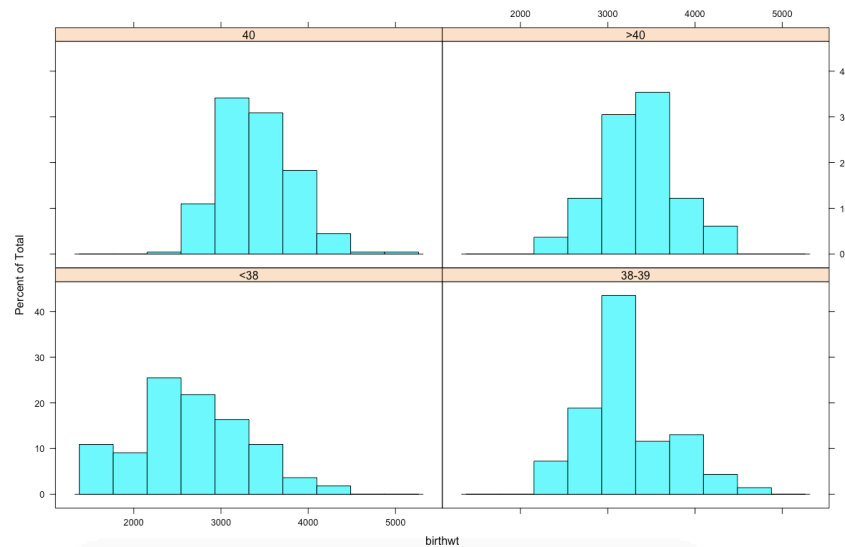
data: birthwt and gest4

	<38	38-39	40
38-39	6.2e-10	-	-
40	< 2e-16	0.005	-
>40	< 2e-16	0.230	1.000

P value adjustment method: bonferroni

Assumption of normality verification

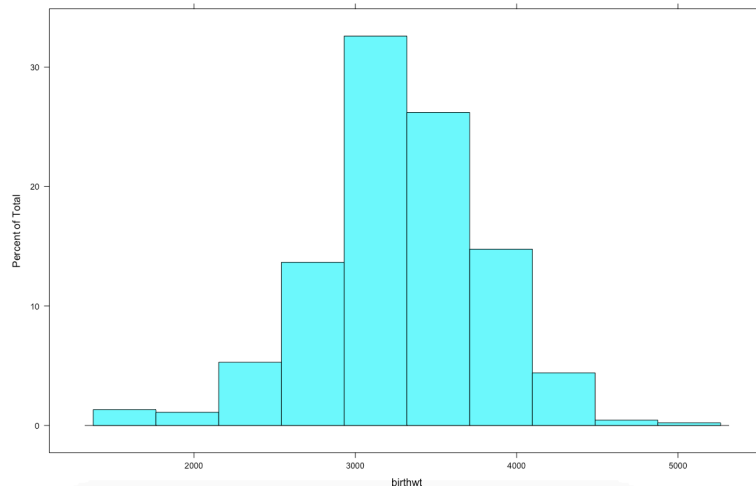
To verify the assumption of normality we can use the visual analysis of the histogram of the variable in the sample in each group.



Kolmogorov-Smirnov test

Normality test

Is the histogram obtained with this sample compatible with a population with normal distribution?



```
> with(alcohol, ks.test(birthwt, "pnorm", mean(birthwt), sd(birthwt)))
```

One-sample Kolmogorov-Smirnov test

```
data: birthwt
D = 0.049885, p-value = 0.2086
alternative hypothesis: two-sided
```

Kolmogorov-Smirnov test

The usefulness of the K-S test is limited.

- It has **very little power** with a **small sample**, i.e. it easily accepts normality for lack of contrary information.
- It is also **very sensitive** when the **sample is large**, i.e. a significant p value is obtained with small deviations from the normal distribution in the sample.

Kolmogorov-Smirnov test

Normality test (sample of 1000 cases with replacement)

```
> ks.test(sam, "pnorm", mean(sam), sd(sam))
```

One-sample Kolmogorov-Smirnov test

data: sam

D = 0.066893, p-value = 0.0002596

alternative hypothesis: two-sided

