

*Curso de Formación de Personal Investigador  
Usuario de Animales para Experimentación*

# **INFERENCE ON K POPULATIONS. ANOVA**

**Jesús Piedrafita Arilla**

jesus.piedrafita@uab.cat

*Departament de Ciència Animal i dels Aliments*

**UAB**

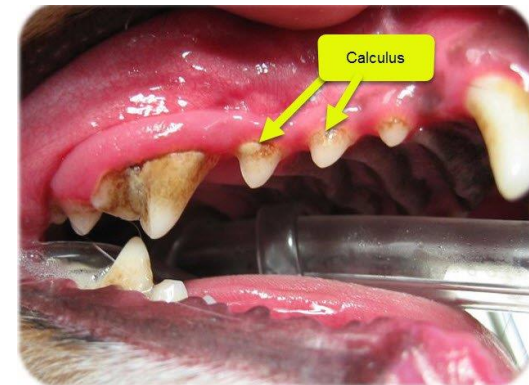
# Learning objectives

- ❑ Define the Completely Random Design, CRD (or One-way ANOVA).
- ❑ Distinguish the between group variability from the within group variability.
- ❑ Define a mathematical model describing CRD and distinguish the dependent and the independent variables.
- ❑ Develop a procedure (ANOVA) to achieve an  $F$  test statistic: calculations with R Commander.
- ❑ Explain the conditions of applicability of ANOVA (normality and within group homogeneous variances) and diagnostic procedures to test them.
- ❑ Determine  $R^2$ , the proportion in which the independent variable explains the dependent variable.
- ❑ Define the idea of Multiple comparisons of means: Tukey test.
- ❑ Power and sample size in ANOVA.

# One-way ANOVA (1) – Completely Random Design

Suppose we have 26 Beagle dogs assigned randomly to three treatments and we measure Index of Calculus Formation (ICF) (Petrie and Watson, 2013).

$\tau_i$	Control	P207	HMP
	0.49	0.34	0.34
	1.05	0.76	0.05
	0.79	0.45	0.53
	1.35	0.69	0.19
	0.55	0.87	0.28
	1.36	0.94	0.45
$y_{ij}$	1.55	0.22	0.71
	1.66	1.07	0.95
	1.00	1.38	



<https://www.dogtrainingpsychology.com/how-to-brush-your-dogs-teeth>

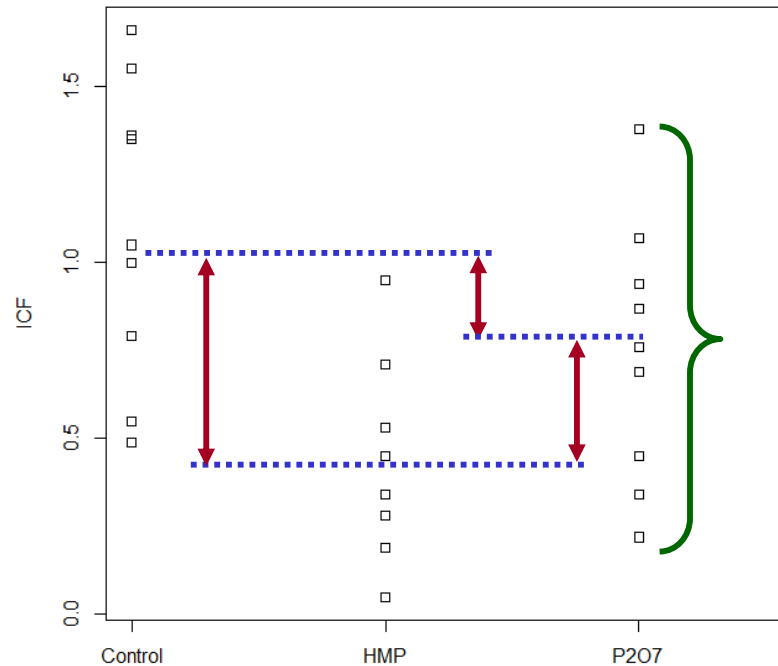
Are significantly different the means of the three treatments,  $H_0 : \mu_1 = \mu_2 = \mu_3$  ?

# One-way ANOVA (2)

Data: dogsICF

Graphs > Strip chart

Factors: Treatment; Response variable: ICF; Options: Stack



**Between treatments variability**

**Within treatments variability**

**Total variability**

# One-way ANOVA (3)

## Model:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where

- i)  $1 \rightarrow t$
- j)  $1 \rightarrow n$ , balanced design
- $1 \rightarrow n_i$ , unbalanced design

Level of a factor (fixed effect)

Mean of level  $i$       Overall mean

$$y_{ij} - \bar{y}_{..} = (y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..})$$

## Sums of squares (SS)

By squaring and summing these quantities, we arrive after some algebra at the computing formulas of the next slide


Total SS,  $SS_T$

Within groups (residual),  $SS_W$

Between groups,  $SS_B$

**! We make a partition of the variation!**

# One-way ANOVA (4) (for a balanced design)

Source of variation	Degrees of freedom	Sum of squares	Mean squares	$F$
Between groups	$t-1$	$SS_B$	$SS_B / (t-1)$	$MS_B / MS_W$
Within groups (error)	$t(n-1)$	$SS_W$	$SS_W / (tn-t)$	
Total	$tn-1$	$SS_T$	$SS_T / (tn-1)$	$F = \frac{(SS_B / \sigma^2) / (t-1)}{(SS_W / \sigma^2) / (tn-t)}$

$$SS_B = \frac{\sum_i y_{i.}^2}{n} - \frac{y_{..}^2}{tn}$$

$$SS_W = \sum_{ij} y_{ij}^2 - \frac{\sum_i y_{i.}^2}{n}$$

$$SS_T = SS_B + SS_W \quad \rightarrow \quad SS_T = \sum_{ij} y_{ij}^2 - \frac{y_{..}^2}{tn}$$

# One-way ANOVA (5), contrast of hypothesis

The **null hypothesis** and the **alternative hypothesis** can be stated as:

$H_0: \mu_1 = \mu_2 = \mu_3$ , the population means are equal


$H_1: \mu_i \neq \mu_{i'}$ , for at least one pair  $(i, i')$ , the means are not equal

An **equivalent formulation** of the hypothesis is:

$H_0: \tau_1 = \tau_2 = \tau_3$ , there is no difference among treatments

$H_1: \tau_i \neq \tau_{i'}$ , for at least one pair  $(i, i')$ , a difference among treatments exist

It can be shown that the **expectations of the mean squares** are:

$E(MS_W) = \sigma^2$    $MS_W$  is an unbiased estimator of  $\sigma^2$

$$E(MS_B) \begin{cases} = \sigma^2 & \text{if } H_0 \\ > \sigma^2 & \text{if not } H_0 \end{cases} \quad \text{red arrow} \quad E(MS_B) = \sigma^2 + \frac{n \sum_i \tau_i^2}{t-1}, \quad \text{for } \sum_i \tau_i = 0$$

$$F = \frac{E(MS_B)}{E(MS_W)} = \frac{\sigma^2 + \frac{n \sum_i \tau_i^2}{t-1}}{\sigma^2} > 1 \quad \text{if not } H_0$$

# One-way ANOVA (6), our example

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F
Between groups	$3 - 1 = 2$	$17.2200 - 15.4154 = 1.8046$	$1.8046 / 2 = 0.9023$	$0.9023 / 0.1353 = 6.668$
Within groups (error)	$26 - 3 = 23$	$20.3320 - 17.2200 = 3.112$	$3.112 / 23 = 0.1353$	

$$\sum_{ij} y_{ij}^2 \Rightarrow 0.49^2 + 1.05^2 + 0.79^2 + \dots + 0.45^2 + 0.71^2 + 0.95^2 = 20.332$$

$$\sum_i \frac{y_i^2}{n_i} \Rightarrow \frac{9.80^2}{9} + \frac{6.72^2}{9} + \frac{3.50^2}{8} = 17.22$$

$$\frac{y_{..}^2}{N} \Rightarrow \frac{(0.49 + 1.05 + 0.79 + \dots + 0.45 + 0.71 + 0.95)^2}{26} = 15.4154$$

**p-value of F**

**> 1-pf(6.668, 2, 23)**  
**[1] 0.005199879**



**Reject  $H_0$ ,  $p < 0.05$**

**This belongs to the era of calculators!!!**



# Conditions of applicability of ANOVA

The conditions of applicability (**additivity and normal errors identically distributed with common variance  $\sigma^2$** ) must be met when we want to make some inference, such as the estimation of the confidence interval or some hypothesis testing. There are **two main conditions** to be checked:

1. **Normality of errors.** We check this both in boxplots for each group and in the **residuals**, our best estimate of the errors.

ANOVA is **little sensitive (robust) to the non normality**. In practice, it is enough to avoid the use of ANOVA when the samples deviate heavily from the Normal distribution.

2. **Homogeneity of within group variances.** It can be tested both in raw data and in the residuals.

If some of the conditions of applicability are not met, we can use some **transformations**

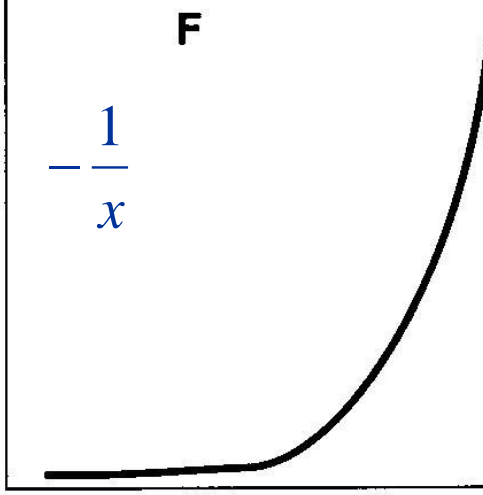
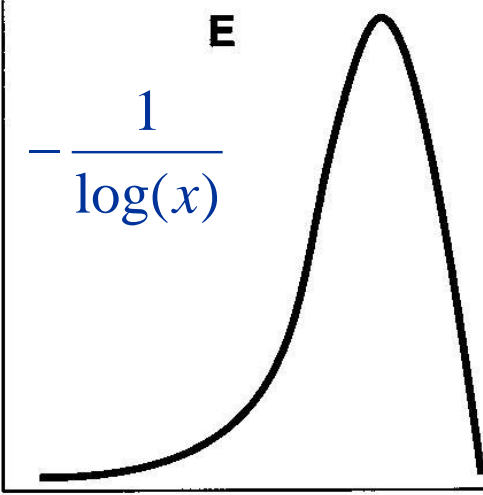
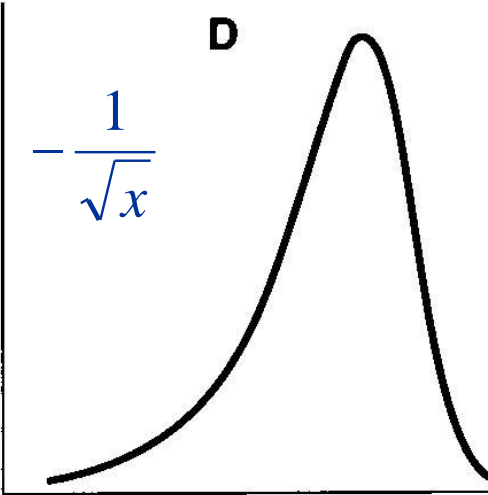
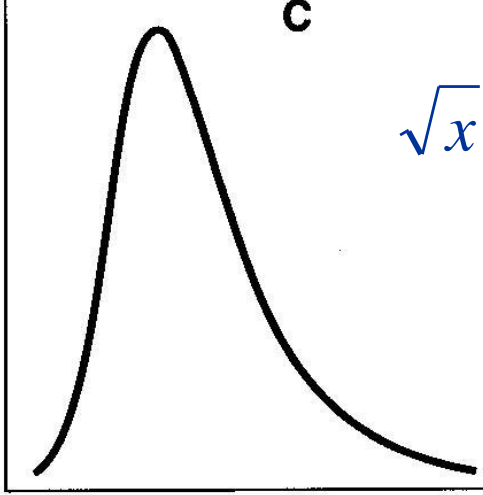
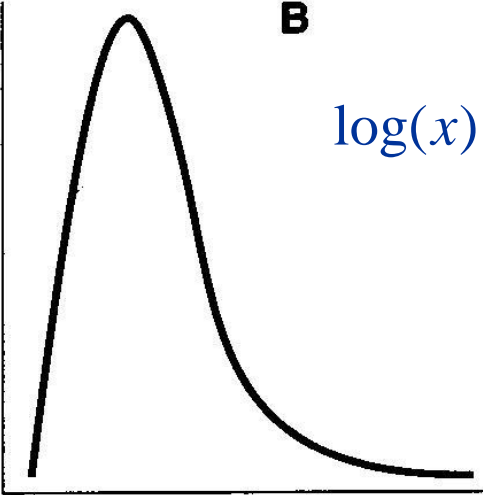
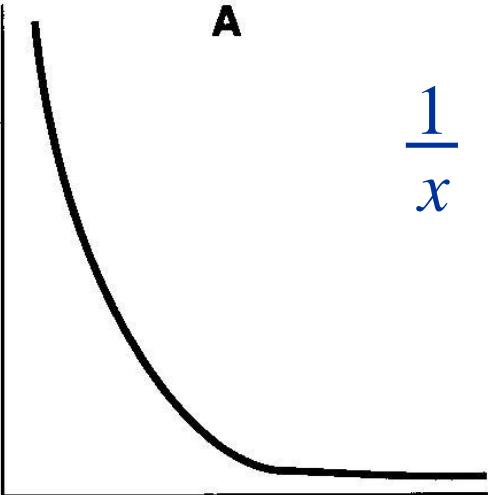
# Transformations (1)

The objective is to normalise the distribution and/or to stabilise the variances. If transformations does not give satisfactory results, Non Parametric Tests can be used, such as Kruskal-Wallis test.

1. **Logarithmic.** When the treatments have a multiplicative effect, i.e., when they increment or decrement the measurements in a percent and not in a fixed quantity.
2. **Root square.** For data consisting of integers coming from counting (ticks on a cow). It tends to equalise  $\sigma^2$ .
3. **Angular (arcsine).** Data are the number of individuals with some particular characteristic (percentages and proportions). Equalises  $\sigma^2$ .
4. **Probit.** For percentage data, like mortality. It is used in pharmacology.
5. **Box-Cox.** A general methodology to transform data.

**To present a true mean value of data in the linear scale it is necessary to reconvert the transformed mean. The standard deviation in this case is of no value and you should compute confidence limits of the transformed data and then convert these to the linear scale.**

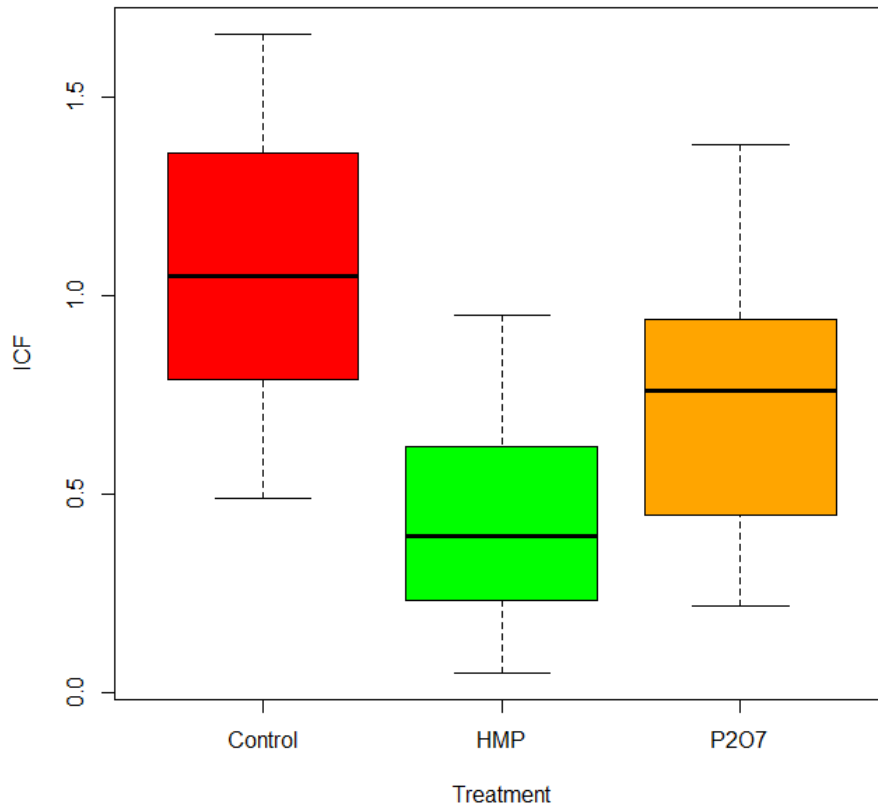
# Transformations (2)



# Working ANOVA – Boxplots (normality)

## Graphs > Boxplot

Variable: ICF; Plot by: Treatment



`col=c("red", "green", "orange")`

No obvious violations of normality and homogeneity of variance: boxplots not asymmetrical, do not vary greatly in size, no outliers

In addition to that we can suspect differences among groups when the boxes of some of the groups do not overlap

# Working ANOVA – Homogeneity of variances

## Statistics > Variances

Factors: Treatment; Response Variable: ICF

Levene's Test for Homogeneity of Variance

(center = "median")

	Df	F value	Pr(>F)
group	2	0.7437	0.4864
	23		

Levene's test confirms the homogeneity of variances of the three treatments ( $\text{Pr}(>F) = 0.4864$ )

This test is only needed if we suspect in the boxplot a clear heterogeneity of the within group variances

# Working ANOVA – Anova table

## Statistics > Means > One-way Anova

Groups: Treatment; Response Variable: ICF

✓ Pairwise comparison of means

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.805	0.9023	6.668	0.0052 **
Residuals	23	3.112	0.1353		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The *p-value* of Treatment effect ( $p = 0.0052$ ) allows us to **reject the overall null hypothesis  $H_0$**  of equality among the means of all treatments

*Residual standard error (RSE)* is an estimate of the within group standard deviation (“noise”). It is computed as the square root of the *Residuals Mean Square*:

$$RSE = \sqrt{0.1353} = 0.3678$$

# Multiple comparisons

If  $H_0$  is not rejected, it is not necessary or appropriate to further analyse the problem, although the researcher must be aware of the possibility of a Type II error.

If  $H_0$  is rejected, then we must question which treatment or treatments caused a differential effect, that is, between which groups is the significant difference found.

For  $t$  treatments, there is a total of  $\binom{t}{2}$  pair-wise comparisons of means. For each comparison there is the possibility of making Type I or Type II errors.

Looking at the experiment as a whole, the probability of making an error in conclusion is defined as the Experiment-wise Error Rate (EER).

There are many procedures of pair-wise comparisons of means: Bonferroni, Duncan, **Dunnet**, LSD, Scheffé, Student-Newman-Keuls, **Tukey**, among others.

# Working ANOVA – Comparison of means, Tukey (1)

	mean	sd	data:n
Control	1.0888889	0.4225353	9
HMP	0.4375000	0.2906520	8
P207	0.7466667	0.3695267	9

## Tukey's Honestly Significant Difference

Multiple Comparisons of Means: Tukey Contrasts

Linear Hypotheses:

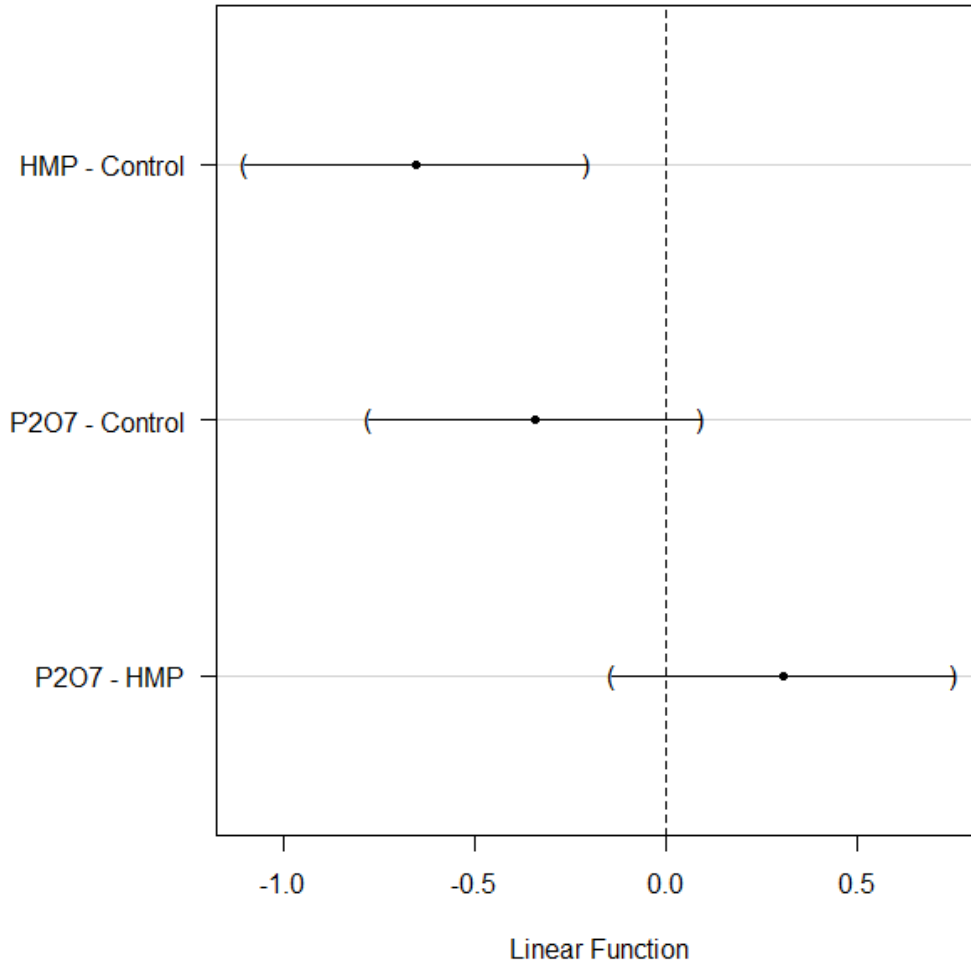
	Estimate	Std. Error	t value	Pr(> t )	
HMP - Control == 0	-0.6514	0.1787	-3.644	0.00368	**
P207 - Control == 0	-0.3422	0.1734	-1.974	0.14128	
P207 - HMP == 0	0.3092	0.1787	1.730	0.21572	

Treatment HMP differs significantly from Control, but P207 does not differ both from Control and HMP



# Working ANOVA – Comparison of means, Tukey (2)

95% family-wise confidence level



Observe how the two lower intervals in the figure contain 0, but in both extremes of the interval

## Working ANOVA – Comparison of means, Tukey (3)

Linear Hypotheses:

	Estimate	lwr	upr
HMP - Control == 0	-0.65139	-1.09911	-0.20367
P207 - Control == 0	-0.34222	-0.77657	0.09213
P207 - HMP == 0	0.30917	-0.13855	0.75689

Control	HMP	P207
"b"	"a"	"ab"

We can observe the numerical values used to construct the previous figure

The last two lines give us the **grouping of means**. HMP differs from Control, but not from P207. Control does not differ from P207.

# Working ANOVA – Comparison of means, Dunnett

```
> library(DescTools)
> DunnettTest(ICF ~ Treatment, data = DogsICF)
```

Dunnett's test for comparing several treatments with a control:

95% family-wise confidence level

\$Control

	diff	lwr.ci	upr.ci	pval	
HMP-Control	-0.6513889	-1.0728610	-0.22991675	0.0026	**
P207-Control	-0.3422222	-0.7511103	0.06666581	0.1081	

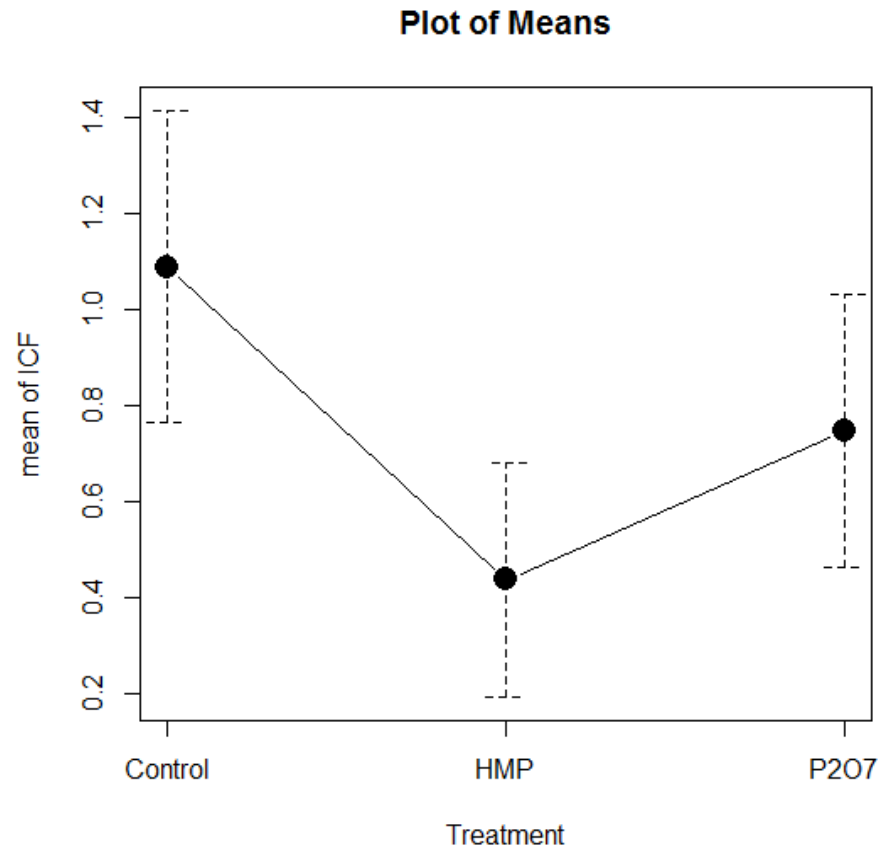
In this test we have only contrasted the two treatments with the control. Only HMP presented a significant effect.

# Working ANOVA – Plotting the means

## Graphs > Plot of means

Factors: Treatment; Response Variable: ICF

Options: Confidence intervals, Top center

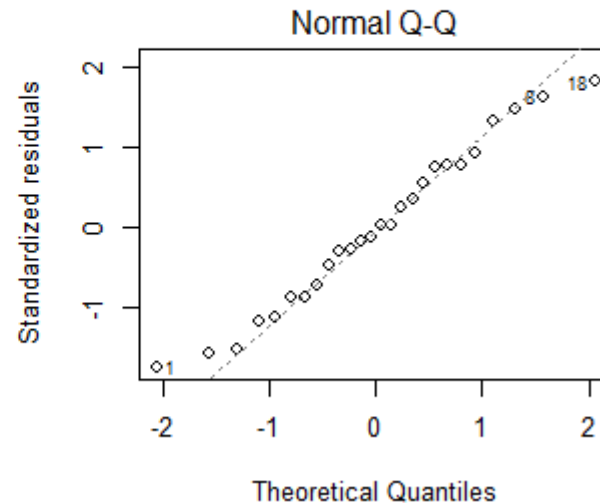
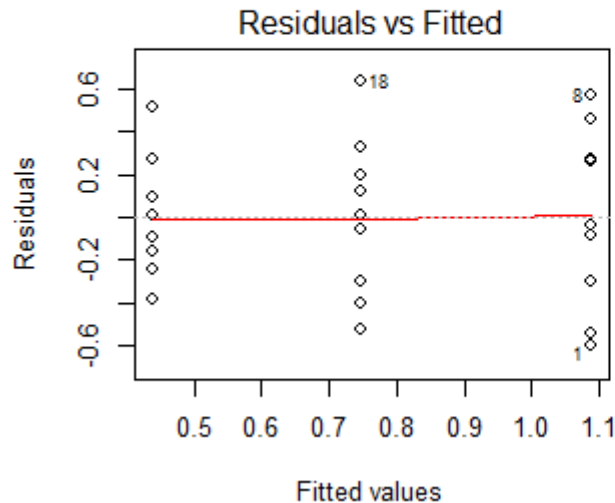


# Working ANOVA – Analysis of residuals

Models > Graphs > Basic diagnostic plots

$$\hat{e}_{ij} = y_{ij} - \hat{\mu} - \hat{\tau}_i$$

aov(ICF ~ Treatment)



No obvious violation of homogeneity of variance: residuals distributed randomly to both sides of the red line (0); no wedge shape in residuals

No obvious violation of normality: Q-Q plot of residuals is linear

# Working ANOVA – Power analysis

```
> MEANS <- tapply(dogsICF$ICF, dogsICF$Treatment, na.rm=T, mean)
> power.anova.test(groups=3, n=9, between.var=var(MEANS),
within.var=anova(AnovaModel.1)[["Mean Sq"]][2], sig.level=0.05)
```

Balanced one-way analysis of variance power calculation

```
groups = 3
n = 9
between.var = 0.1061679
within.var = 0.135306
sig.level = 0.05
power = 0.8942738
```

We have a probability of almost 90% of rejecting  $H_0$  when false

NOTE: n is number in each group

The test has been developed for balanced ANOVA designs. In our case we assume 9 observations per group.

The students could resort to GRANMO: <http://www.imim.cat/ofertadeserveis/granmo.html>

# Working ANOVA – Sample size

```
> power.anova.test(groups=3, power=0.8, between.var=var(MEANS),  
within.var=anova(AnovaModel.1)[["Mean Sq"]][2], sig.level=0.05)
```

Balanced one-way analysis of variance power calculation

```
groups = 3  
n = 7.237864  
between.var = 0.1061679  
within.var = 0.135306  
sig.level = 0.05  
power = 0.8
```



We will have needed more than 7 animals (8) to detect as significant the differences observed between the means

NOTE: n is number in each group

As before, the students could resort also to GRANMO:

<http://www.imim.cat/ofertadeserveis/granmo.html>