

# Analysis of Variance

## 1 Introduction

## 2 ANOVA

- Understanding ANOVA
- ANOVA In Practice
- The Nonparametric Equivalence of ANOVA
- Multiple Comparison Tests
- ANOVA Assumptions Checking

## 1 Introduction

## 2 ANOVA

- Understanding ANOVA
- ANOVA In Practice
- The Nonparametric Equivalence of ANOVA
- Multiple Comparison Tests
- ANOVA Assumptions Checking

# Motivating Scenario

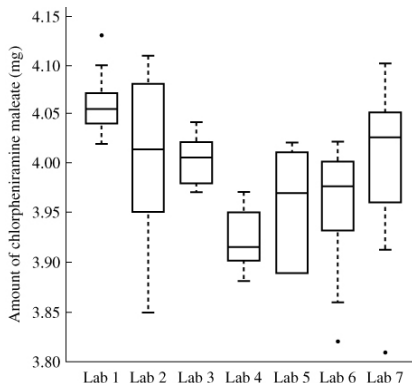
## Example (Growth Upon Different Concentrations of Fertilizer)

- 3 different concentrations (low, med, high) of a fertilizer are used on a particular plant species.
- Each concentration is applied to 5 randomly selected plants, and their growth is measured after 2 weeks.
- In addition, a control group is assigned, where no fertilizer is used.
- Assess if there is a significant difference between the mean growth for each group at  $\alpha = 0.05$ .

# Example

## Example (Chlorpheniramine Maleate in Tablets)

- Kirchhoefer (1979) studied the measurement of chlorpheniramine maleate in tablets.
- Measurements of composites that had nominal dosages equal to 4 mg were made by seven laboratories, each making 10 measurements.

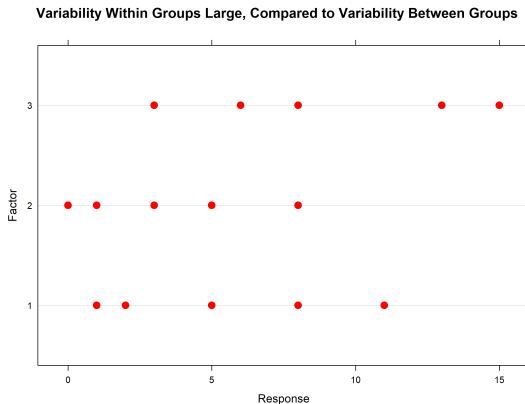


# What We Concern Is

- From the boxplots, there may be some systematic differences between the labs and, there is less variability in some labs than in others.
- We concern: are the differences in the means of the measurements from the various labs significant, or might they be due to chance?

# Variability Is The Key

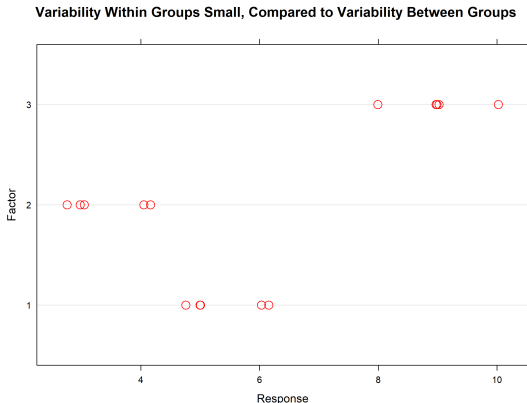
Consider the following data from three groups:



The group means are given by 5.4, 3.4 and 9.

# Variability Is The Key

Consider the following data from three groups, that have the same sample means:



Which data (this slide or the previous slide) provides more evidence that the group means are different?

In order to figure out, we would need to make comparison between groups. The comparisons can be made using parametric (ANOVA) or nonparametric (Kruskal-Wallis) test.



## 1 Introduction

## 2 ANOVA

- Understanding ANOVA
- ANOVA In Practice
- The Nonparametric Equivalence of ANOVA
- Multiple Comparison Tests
- ANOVA Assumptions Checking

## 1 Introduction

## 2 ANOVA

- Understanding ANOVA
- ANOVA In Practice
- The Nonparametric Equivalence of ANOVA
- Multiple Comparison Tests
- ANOVA Assumptions Checking

# The $F$ Test

- A one-way layout is an experimental design in which independent measurements are made under each of several treatments.
- The technique is hence the generalizations of the technique for comparing 2 independent samples.
- We first discuss the analysis of variance and the  $F$  test in the case of  $I$  **treatments/groups** and  $J$  **measurements in each group**. In the Example 1,  $I = 7$  and  $J = 10$ .
- We will discuss the case of unequal sample sizes later.

# Some Notations

- $Y_{ij}$ : the  $j$ th observation of the  $i$ th treatment/group.
- We consider the observations are corrupted by random errors and that the error in one observation is independent of the errors in the other observations. Hence, the **one-way ANOVA model** is:

$$Y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \dots, I; j = 1, \dots, J,$$

where  $\mu$  is the overall mean,  $\alpha_i$  is the differential effect of the  $i$ th treatment, and  $e_{ij}$  are IID  $\sim N(0, \sigma^2)$ .

The  $\alpha_i$  are assumed to have

$$\sum_{i=1}^I \alpha_i = 0$$

# The Null Hypothesis

- The expected response to the  $i$ th treatment is

$$E(Y_{ij}) = \mu + \alpha_i$$

- Hence, if  $\alpha_i = 0$  for all  $i$ , then all treatments have the same expected response, and in general  $(\alpha_i - \alpha_k)$  is the difference between the expected values under treatment  $i$  and treatment  $k$ .
- We'll derive a test for the **null hypothesis that all the means are equal**.

# Sums of Squares

- Let  $\bar{Y}_i = \frac{1}{J} \sum_{j=1}^J Y_{ij}$  be the average of the observations under the  $i$ th treatment.
- Denote  $\bar{\bar{Y}} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J Y_{ij}$  be the overall average.
- The analysis of variance is based on the square of deviation of each observation from its overall average.

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - \bar{\bar{Y}})^2 &= \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - \bar{Y}_i)^2 & + J \sum_{i=1}^I (\bar{Y}_i - \bar{\bar{Y}})^2 \\ SS_{TOT} &= SS_W & + SS_B \end{aligned}$$

- **Sum of squares = sum of squares within groups + sum of squares between groups.**

# Sums of Squares

- $SS_W$  is a measure of the variation of the data within the groups.
- $SS_B$  is a measure of variation of the means between groups.
- The basic idea underlying Anova is the comparison of the sizes of sums of squares.
- Hence, calculating the expected values of the sums of squares is needed.

## Lemma (A)

Let  $X_i$ , where  $i = 1, \dots, n$  be independent random variables with  $E(X_i) = \mu_i$  and  $\text{Var}(X_i) = \sigma^2$ . Then

$$E(X_i - \bar{X})^2 = (\mu_i - \bar{\mu})^2 + \frac{n-1}{n}\sigma^2$$

where

$$\bar{\mu} = \frac{1}{n} \sum_{i=1}^n \mu_i$$

# Expectation of Sums of Squares

## Theorem (A)

*Under the assumptions for the model stated before,*

$$\begin{aligned}E(SS_W) &= \sum_{i=1}^I \sum_{j=1}^J E(Y_{ij} - \bar{Y}_i)^2 \\&= \sum_{i=1}^I \sum_{j=1}^J \frac{J-1}{J} \sigma^2 \\&= I(J-1)\sigma^2 \\E(SS_B) &= J \sum_{i=1}^I E(\bar{Y}_i - \bar{\bar{Y}})^2 \\&= J \sum_{i=1}^I \left[ \alpha_i^2 + \frac{(I-1)\sigma^2}{IJ} \right] \\&= J \sum_{i=1}^I \alpha_i^2 + (I-1)\sigma^2\end{aligned}$$



- For  $E(SS_W)$ , Lemma A is applied with the role of  $X_i$  is played by  $Y_{ij}$  and  $\bar{X}$  is played by  $\bar{Y}_i$ .
- Note that  $E(Y_{ij}) = E(\bar{Y}_i) = \mu + \alpha_i$ .
- For  $E(SS_B)$ , Lemma A is applied again with  $\bar{Y}_i$  and  $\bar{\bar{Y}}$ .

## Estimate $\sigma^2$

- $SS_W$  can be used to estimate  $\sigma^2$ , the unbiased estimate is

$$s_p^2 = \frac{SS_W}{I(J-1)}$$

- Denote  $s_i^2$  be the sample variance in the  $i$ th group,  $SS_W$  can be written as

$$SS_W = \sum_{i=1}^I (J-1)s_i^2$$

hence, estimates of  $\sigma^2$  from the  $I$  groups are pooled together.

- If all the  $\alpha_i$  are equal to zero, then

$$E[SS_B/(I-1)] = \sigma^2$$

Thus,  $SS_W/[I(J-1)]$  and  $SS_B/(I-1)$  should be about equal. If some of the  $\alpha_i$  are nonzero,  $SS_B$  will be inflated.

# Test Statistics for ANOVA

## Theorem (B)

*If errors are independent and normally distributed with mean 0 and variances  $\sigma^2$ , then  $SS_W/\sigma^2$  follows a chi-square distribution with  $I(J - 1)$  degrees of freedom. If, additionally, the  $\alpha_i$  are all equal to zero, then  $SS_B/\sigma^2$  follows a chi-square distribution with  $I - 1$  degrees of freedom and is independent of  $SS_W$ .*

- For the null hypothesis of ANOVA:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_I$$

the test statistic is

$$F = \frac{SS_B/(I - 1)}{SS_W/[I(J - 1)]}.$$

# Distribution of Test Statistic F

- By Theorem A,  $E\{SS_W/[I(J-1)]\} = \sigma^2$ , and

$$E[SS_B/(I-1)] = J(I-1)^{-1} \sum_{i=1}^I \alpha_i^2 + \sigma^2.$$

- Thus, under  $H_0$ ,  $F$  should be close to 1.
- If  $H_0$  is false, the numerator of  $F$  reflects the variation between the different groups as well as the variation within groups, whereas the denominator reflects only the variation within groups.  
Hence,  $H_0$  is rejected for large values of  $F$ .

## Theorem (C)

*Under the assumption that the errors are normally distributed, the null distribution of  $F$  is the  $F$  distribution with  $(I-1)$  and  $I(J-1)$  degrees of freedom.*

- In summary, if  $F > F_{I-1, I(J-1)}(\alpha)$ , reject  $H_0$ ; and do not reject  $H_0$  otherwise.
- **The exact p-value is given by**

$$p\text{-value} = P[F_{I-1, I(J-1)} > F]$$

## Example 1: Chlorpheniramine Maleate in Tablets

- $SS_W$ : sum of squares due to errors: 0.231
- $SS_B$ : sum of squares due to labs: 0.125
- Test statistic:  $F = 5.66$ , follows  $F$  distribution with  $df_1 = 6$  and  $df_2 = 63$ .
- From the table for  $F$  distribution,  $F_{6,60}(0.01) = 3.12$  and  $F_{6,120}(0.01) = 2.96$ , the p-value corresponds to 5.66 from  $F_{6,63}$  is smaller than 0.01.
- We may thus can conclude that the means of the measurements from the various labs are significantly different.

# Popular Seen Output

Source	Sum of Squares	df	Mean Square	F
Labs	0.1247	6	0.0208	5.66
Error	0.2314	63	0.003673	
Total	0.3561	69		

## Groups with Different Size

- **Assume  $I$  groups with the size of  $J_1, \dots, J_I$ .** Total sample size  $n = \sum_{i=1}^I J_i$ .
- Conceptually, the analysis is the same as for the case when  $J_1 = J_2 = \dots = J_I = J$ .
- The difficulties with this case are the algebraic (which is skipped here).
- It can be shown that

$$E(SS_W) = \sigma^2 \sum_{i=1}^I (J_i - 1)$$

$$E(SS_B) = (I - 1)\sigma^2 + \sum_{i=1}^I J_i \alpha_i^2$$

- The test statistic  $F$  is then follows  $F$  distribution with degrees of freedom  $I - 1$  and  $\sum_{i=1}^I J_i - I = n - I$ .

# Assumptions of ANOVA

- Random samples
- Equal variance
- Independence of errors
- Normal distribution of errors
- Additivity of treatment effects



## 1 Introduction

## 2 ANOVA

- Understanding ANOVA
- **ANOVA In Practice**
- The Nonparametric Equivalence of ANOVA
- Multiple Comparison Tests
- ANOVA Assumptions Checking

# ANOVA in R (1)

```
> data<-read.csv("C:/Data/tablets1.txt", sep = ",", quote = ' ', head = 1)
> names(data) = c("lab1", 'lab2', 'lab3', 'lab4', 'lab5', 'lab6', 'lab7')
> attach(data)
> data
```

	lab1	lab2	lab3	lab4	lab5	lab6	lab7
1	4.13	3.86	4.00	3.88	4.02	4.02	4.00
2	4.07	3.85	4.02	3.88	3.95	3.86	4.02
3	4.04	4.08	4.01	3.91	4.02	3.96	4.03
4	4.07	4.11	4.01	3.95	3.89	3.97	4.04
5	4.05	4.08	4.04	3.92	3.91	4.00	4.10
6	4.04	4.01	3.99	3.97	4.01	3.82	3.81
7	4.02	4.02	4.03	3.92	3.89	3.98	3.91
8	4.06	4.04	3.97	3.90	3.89	3.99	3.96
9	4.10	3.97	3.98	3.97	3.99	4.02	4.05
10	4.04	3.95	3.98	3.90	4.00	3.93	4.06

- This is the original dataset, however it is not of the form of a dataframe that we want.

## ANOVA in R (2)

```
> amount <- c(lab1, lab2, lab3, lab4, lab5, lab6, lab7)
> lab<-c(rep(1,10),rep(2,10),rep(3,10),rep(4,10),rep(5,10), rep(6,10), rep(7,10))
> newdata<-data.frame(amount = amount,lab = lab)
> attach(newdata)
> newdata$lab = as.factor(newdata$lab)
```

- “newdata” is the correct dataframe (two variables = two columns: measurements and the labs; and 70 rows.)

## ANOVA in R (3)

```
> anova<-aov(amount~lab, data = newdata)
> summary(anova)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
lab	6	0.1247	0.020790	5.66	9.45e-05 ***
Residuals	63	0.2314	0.003673		

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

- Can you write R code to calculate the F value in the output above?

```
> tapply(amount, lab, mean)#to get mean of amount for different lab
```

1	2	3	4	5	6	7
4.062	3.997	4.003	3.920	3.957	3.955	3.998

```
> #tapply(amount, lab, mean) - mean(amount)
```

# ANOVA in Python (1)

```
data.columns = ['lab1', 'lab2', 'lab3', 'lab4', 'lab5', 'lab6', 'lab7']  
print (data)
```

	lab1	lab2	lab3	lab4	lab5	lab6	lab7
0	4.13	3.86	4.00	3.88	4.02	4.02	4.00
1	4.07	3.85	4.02	3.88	3.95	3.86	4.02
2	4.04	4.08	4.01	3.91	4.02	3.96	4.03
3	4.07	4.11	4.01	3.95	3.89	3.97	4.04
4	4.05	4.08	4.04	3.92	3.91	4.00	4.10
5	4.04	4.01	3.99	3.97	4.01	3.82	3.81
6	4.02	4.02	4.03	3.92	3.89	3.98	3.91
7	4.06	4.04	3.97	3.90	3.89	3.99	3.96
8	4.10	3.97	3.98	3.97	3.99	4.02	4.05
9	4.04	3.95	3.98	3.90	4.00	3.93	4.06

```
lab1= data['lab1']  
lab2 = data['lab2']  
lab3 = data['lab3']  
lab4 = data['lab4']  
lab5 = data['lab5']  
lab6 = data['lab6']  
lab7 = data['lab7']
```

## ANOVA in Python (2)

```
##### ANOVA
import statsmodels.api as sm
from statsmodels.formula.api import ols

mod = ols('amount ~ lab', data=newdata).fit()
anova = sm.stats.anova_lm(mod, typ=2)
print(anova)
```

	sum_sq	df	F	PR(>F)
lab	0.124737	6.0	5.660069	0.000095
Residual	0.231400	63.0	NaN	NaN

```
##### ANOVA
anova2 = scipy.stats.f_oneway(lab1, lab2, lab3, lab4, lab5, lab6, lab7)
print(anova2)
# this is equivalent to the above
```

```
F_onewayResult(statistic=5.66006914433881, pvalue=9.452826448501141e-05)
```

# ANOVA in SAS (1)

## Reading data file

```
* IMPORTING DATA FROM A TEXT/CSV FILE;;  
FILENAME REFFILE '/folders/myfolders/tablets.csv';  
  
.....  
PROC IMPORT DATAFILE=REFFILE  
    DBMS=DLM  
    OUT=WORK.newdata;  
    DELIMITER=",";  
    GETNAMES=YES;  
    DATAROW=2;  
RUN;  
  
.....  
proc print data=newdata;  
run;
```

# ANOVA in SAS (2)

## Data

Obs	VAR1	amount	lab
1	1	4.13	1
2	2	4.07	1
3	3	4.04	1
4	4	4.07	1
5	5	4.05	1
6	6	4.04	1
7	7	4.02	1
8	8	4.06	1
9	9	4.1	1
10	10	4.04	1
11	11	3.86	2
12	12	3.85	2
13	13	4.08	2
14	14	4.11	2
15	15	4.08	2
16	16	4.01	2
17	17	4.02	2



# ANOVA in SAS (3)

## Anova

```
/*                ANOVA                */  
PROC ANOVA data = newdata;  
class lab;  
model amount = lab;  
means;  
run;
```

The ANOVA Procedure

Dependent Variable: amount

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	0.12473714	0.02078952	5.66	<.0001
Error	63	0.23140000	0.00367302		
Corrected Total	69	0.35613714			

R-Square	Coeff Var	Root MSE	amount Mean
0.350250	1.521002	0.060605	3.984571

Source	DF	Anova SS	Mean Square	F Value	Pr > F
lab	6	0.12473714	0.02078952	5.66	<.0001

## 1 Introduction

## 2 ANOVA

- Understanding ANOVA
- ANOVA In Practice
- **The Nonparametric Equivalence of ANOVA**
- Multiple Comparison Tests
- ANOVA Assumptions Checking

# The Kruskal-Wallis Test

- When the assumptions of Anova are violated, such as: there are  $I$  groups but the observations are categorical or non-normal; Or the observations are continuous but with un-equal variances, then conducting Anova to compare groups is not a good idea.
- The **Kruskal-Wallis test is equivalent to ANOVA**, used for the situation of non-normal data. It's a generalization of the Mann-Whitney test.
- Observations are assumed to be independent, but no particular form of distribution is assumed.
- The observation are pooled together and ranked. The test will work with the rank rather than with the observations.
- The Kruskal-Wallis test makes no assumption of normality and thus has a wider range of applicability than  $F$  test.
- It is especially useful in small-sample situation.
- Its null hypothesis: the groups are from the same distribution.

# Kruskal-Wallis Test in R

- Even though we can use Anova for the data of Chlorpheniramine, but we can try to use Kruskal-Wallis test on this.

```
> kruskal.test(amount~lab)  
Kruskal-Wallis rank sum test
```

```
data: amount by lab
```

```
Kruskal-Wallis chi-squared = 29.606, df = 6, p-value = 4.67e-05
```

# Kruskal-Wallis Test in Python

```
#### KRUSKAL-WALLIS TEST
from scipy.stats import kruskal
kruskal = scipy.stats.kruskal(lab1, lab2, lab3, lab4, lab5, lab6, lab7)

print(kruskal)
```

```
KruskalResult(statistic=29.605984376371413, pvalue=4.6701021928163535e-05)
```

# Kruskal-Wallis Test in SAS

```
*      Kruskal-Wallis Test;
proc npar1way data= newdata wilcoxon dscf;
class lab;
var amount;
run;
```

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable amount  
Classified by Variable lab

lab	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	10	598.50	355.0	59.482760	59.850
2	10	396.00	355.0	59.482760	39.600
3	10	390.00	355.0	59.482760	39.000
4	10	155.00	355.0	59.482760	15.500
5	10	264.00	355.0	59.482760	26.400
6	10	265.50	355.0	59.482760	26.550
7	10	416.00	355.0	59.482760	41.600

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	DF	Pr > ChiSq
29.6060	6	<.0001

## 1 Introduction

## 2 ANOVA

- Understanding ANOVA
- ANOVA In Practice
- The Nonparametric Equivalence of ANOVA
- **Multiple Comparison Tests**
- ANOVA Assumptions Checking

## If ANOVA or Kruskal-Wallis Test is Significant...

- Imagine if  $H_0$  of ANOVA F test is rejected, meaning the means are not all equal, but the test does not tell us how they differ, in particular about which pair are significantly different.
- The real interest may be focused on comparing pairs or groups and estimating the group means and their differences.
- A naive approach would be to **compare all pairs of group**. In general, methods used to find group differences after the null hypothesis of Anova has been rejected are called **post hoc**, or **multiple comparison tests**.
- Imagine, if there are 7 groups, each individual test for a pairwise comparison has significance level  $\alpha = 0.05$ , then what is the probability of making **at least one type 1 error** among all the 21 pairwise comparisons? About 66%!
- Hence, when multiple comparison tests are derived, correction (in order to reduce the probability of making type 1 error or all the pairwise comparisons) is needed.



# Multiple Comparison Tests

Some multiple comparison tests:

- **Bonferroni**: this multiple comparison test not only can be applied to the data where Anova is applicable but also can be applied to the data where Anova is not applicable (data do not satisfy the normality assumptions).
- **Tukey**: is used in Anova
- **Least significant difference (LSD)** test: is used in Anova.
- Others: Duncan's multiple-range test; Student-Newman-Keul's multiple-range test; Scheffe's multiple-comparison procedure.
- In multiple comparison tests, a very simple idea is, instead of controlling "locally" the probability of type 1 error for each individual test, we can control "globally" the probability of at least a type 1 error among all these  $k$  test, by a value that is called **family error rate**  $\alpha$  or experimentwise error rate.
- The role of this family significance level  $\alpha$  to the  $k$  tests is the same as the role of individual significance level  $\alpha$  to each test.

# Bonferroni's Correction

- Bonferroni's method is very simple: if  $k$  null hypotheses are to be tested, a desired family error rate of  $\alpha$  can be guaranteed by testing each individual null hypothesis at level  $\alpha = \alpha/k$ .
- The idea of this correction is simple but give pretty good results if  $k$  is not too large.

# Bonferroni Test in R

```
> pairwise.t.test(amount, lab, p.adj = "bonf")
```

Pairwise comparisons using t tests with pooled SD

data: amount and lab

	1	2	3	4	5	6
2	0.4084	-	-	-	-	-
3	0.6982	1.0000	-	-	-	-
4	4.2e-05	0.1271	0.0678	-	-	-
5	0.0054	1.0000	1.0000	1.0000	-	-
6	0.0042	1.0000	1.0000	1.0000	1.0000	-
7	0.4476	1.0000	1.0000	0.1147	1.0000	1.0000

P value adjustment method: bonferroni

# Bonferroni Test in Python

```
##### BONFERRONI TEST:
import statsmodels.stats.multicomp as mc

comparison = mc.MultiComparison(newdata['amount'], newdata['lab'])
result, tbl1, tbl2 = comparison.allpairtest(stats.ttest_ind, method= "bonf")

print(result)
#Bonf = comparison.allpairtest(stats.ttest_ind, method= "bonf")|
#Bonf
```

```
Test Multiple Comparison ttest_ind
FWER=0.05 method=bonf
alphacSidak=0.00, alphacBonf=0.002
```

```
=====
```

group1	group2	stat	pval	pval_corr	reject
1.0	2.0	2.1538	0.0451	0.9462	False
1.0	3.0	4.6692	0.0002	0.004	True
1.0	4.0	9.6322	0.0	0.0	True
1.0	5.0	5.046	0.0001	0.0018	True
1.0	6.0	4.5392	0.0003	0.0053	True
1.0	7.0	2.2271	0.0389	0.8177	False
2.0	3.0	-0.2048	0.84	1.0	False
2.0	4.0	2.5446	0.0203	0.4269	False
2.0	5.0	1.1892	0.2498	1.0	False
2.0	6.0	1.186	0.251	1.0	False
2.0	7.0	-0.0256	0.9798	1.0	False

```
-----
```

# Bonferroni Test in SAS (1)

```
/* Multiple Comparisons: Bonferroni */  
  
PROC ANOVA data =newdata;  
class lab;  
model amount = lab;  
means lab / Bon cldiff alpha=0.05;  
run;
```

# Bonferroni Test in SAS (2)

## Output

### The ANOVA Procedure

#### Bonferroni (Dunn) t Tests for amount

**Note:** This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha	0.05
Error Degrees of Freedom	63
Error Mean Square	0.003673
Critical Value of t	3.16614
Minimum Significant Difference	0.0858

Comparisons significant at the 0.05 level are indicated by \*\*\*.

lab Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 3	0.05900	-0.02681	0.14481	
1 - 7	0.06400	-0.02181	0.14981	
1 - 2	0.06500	-0.02081	0.15081	
1 - 5	0.10500	0.01919	0.19081	***
1 - 6	0.10700	0.02119	0.19281	***
1 - 4	0.14200	0.05619	0.22781	***

Output provides the adjusted confidence intervals for each comparison.

# Tukey Test

- The idea of Tukey's correction is more complicated than Bonferroni's, however generally it is better than Bonferroni's correction in the situation where data are applicable for Anova.
- Set  $\alpha$  to be the family type 1 error rate for  $k$  tests.
- Tukey's method is used to construct CIs for the differences of all pairs of means, in such a way that the CIs simultaneously have coverage probability of  $\alpha$ .
- Assume the sample sizes are all equal, and the errors are normally distributed with mean 0 and variance  $\sigma^2$ .
- We have:

$$\bar{Y}_i = \mu + \alpha_i + \frac{1}{J} \sum_{j=1}^J e_{ij}$$

Denote  $\mu_i = \mu + \alpha_i$ , then  $\bar{Y}_i - \mu_i = 1/J \sum_{j=1}^J e_{ij}$  for  $i = 1, \dots, I$  are

independent and normally distributed with mean 0 and variance  $\sigma^2/J$ .

## Tukey Test (2)

- The term  $\sigma^2/J$  can be estimated by  $s_p^2/J$ .
- Consider a random pairwise test between group  $i_1$  and group  $i_2$  where  $i_1, i_2 \in \{1, \dots, I\}$ . The probability distribution of the random variable

$$\max_{\{i_1, i_2\}} \frac{|(\bar{Y}_{i_1} - \mu_{i_1}) - (\bar{Y}_{i_2} - \mu_{i_2})|}{s_p/\sqrt{J}}$$

is the **studentized range distribution** with parameter  $I$  and  $I(J-1)$ .

- Denote the upper  $100a$  percentage point of the distribution by  $q(a)$ , then:

$$\begin{aligned} &P\left[|(\bar{Y}_{i_1} - \mu_{i_1}) - (\bar{Y}_{i_2} - \mu_{i_2})| \leq q(a) \frac{s_p}{\sqrt{J}}, \text{ for all } i_1 \text{ and } i_2\right] \\ &= P\left[\max_{\{i_1, i_2\}} |(\bar{Y}_{i_1} - \mu_{i_1}) - (\bar{Y}_{i_2} - \mu_{i_2})| \leq q(a) \frac{s_p}{\sqrt{J}}\right] = 1 - a, \quad \text{by definition.} \end{aligned}$$



## Tukey Test (3)

- Since

$$P\left[|(\bar{Y}_{i_1} - \mu_{i_1}) - (\bar{Y}_{i_2} - \mu_{i_2})| \leq q(a) \frac{s_p}{\sqrt{J}}, \text{ for all } i_1 \text{ and } i_2\right] = 1 - a$$

the  $100(1 - a)\%$  CIs for  $\mu_{i_1} - \mu_{i_2}$  then

$$(\bar{Y}_{i_1} - \bar{Y}_{i_2}) \pm q(a) \frac{s_p}{\sqrt{J}}.$$

- By the duality of CIs and hypothesis tests, if the CI does not include zero then the null hypothesis of equal means (of group  $i_1$  and group  $i_2$ ) is rejected at level  $a$ .

# Tukey Test in R

```
> TukeyHSD(anova)
```

```
Tukey multiple comparisons of means  
95% family-wise confidence level
```

```
Fit: aov(formula = amount ~ lab, data = newdata)
```

```
$lab
```

	diff	lwr	upr	p adj
2-1	-0.065	-0.147546752	0.017546752	0.2165897
3-1	-0.059	-0.141546752	0.023546752	0.3226101
4-1	-0.142	-0.224546752	-0.059453248	0.0000396
5-1	-0.105	-0.187546752	-0.022453248	0.0045796
6-1	-0.107	-0.189546752	-0.024453248	0.0036211
7-1	-0.064	-0.146546752	0.018546752	0.2323813
3-2	0.006	-0.076546752	0.088546752	0.9999894
4-2	-0.077	-0.159546752	0.005546752	0.0830664
5-2	-0.040	-0.122546752	0.042546752	0.7578129
6-2	-0.042	-0.124546752	0.040546752	0.7140108
7-2	0.001	-0.081546752	0.083546752	1.0000000
4-3	-0.083	-0.165546752	-0.000453248	0.0478900

# Tukey Test in Python

```
##### TUKEY TEST:

import statsmodels.stats.multicomp as mc

comparison = mc.MultiComparison(newdata['amount'], newdata['lab'])
tukey = comparison.tukeyhsd()
print(tukey.summary() )
```

Multiple Comparison of Means - Tukey HSD,FWER=0.05

```
=====
group1 group2 meandiff lower upper reject
-----
1.0     2.0    -0.065  -0.1475  0.0175 False
1.0     3.0    -0.059  -0.1415  0.0235 False
1.0     4.0    -0.142  -0.2245 -0.0595  True
1.0     5.0    -0.105  -0.1875 -0.0225  True
1.0     6.0    -0.107  -0.1895 -0.0245  True
1.0     7.0    -0.064  -0.1465  0.0185 False
2.0     3.0     0.006   -0.0765  0.0885 False
2.0     4.0    -0.077  -0.1595  0.0055 False
2.0     5.0    -0.04   -0.1225  0.0425 False
2.0     6.0    -0.042  -0.1245  0.0405 False
2.0     7.0     0.001   -0.0815  0.0835 False
```

# Tukey Test in SAS

```
/* Multiple Comparisons: Tukey */  
  
PROC ANOVA data =newdata;  
class lab;  
model amount = lab;  
means lab / tukey cldiff alpha=0.05;  
run;
```

## The ANOVA Procedure

### Tukey's Studentized Range (HSD) Test for amount

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	63
Error Mean Square	0.003673
Critical Value of Studentized Range	4.30711
Minimum Significant Difference	0.0825

Comparisons significant at the 0.05 level are indicated by \*\*\*.

lab Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 3	0.05900	-0.02355	0.14155	
1 - 7	0.06400	-0.01855	0.14655	
1 - 2	0.06500	-0.01755	0.14755	
1 - 5	0.10500	0.02245	0.18755	***
1 - 6	0.10700	0.02445	0.18955	***
1 - 4	0.14200	0.05945	0.22455	***
3 - 1	-0.05900	-0.14155	0.02355	

# LSD Test

- For group  $i_1$  and group  $i_2$ , where  $i_2, i_1 \in \{1, \dots, I\}$ , which have the size  $J_1$  and  $J_2$  respectively, consider the difference  $\mu_{i_1} - \mu_{i_2}$ .
- The least significance difference (LSD) for  $\mu_{i_1} - \mu_{i_2}$  is given by

$$t_{n-I}(\alpha/2) \sqrt{MS_W \left( \frac{1}{J_1} + \frac{1}{J_2} \right)}$$

where  $n$  is the total observations,  $MS_W = SS_W / (n - I)$ .

- We conclude  $\mu_{i_1}$  is different from  $\mu_{i_2}$  if

$$\bar{Y}_{i_1} - \bar{Y}_{i_2} > LSD.$$

# LSD in R

```
> MSW <- sum(anova$res^2)/63 #  $I(J-1) = 63$   
> lsd <- qt(0.975,63)*sqrt(MSW*2/7) # using  $\alpha = 0.05$ 
```

- Any two groups that have the difference in the sample means that is larger than “lsd” computed above then we can conclude that the two means are different.

## 1 Introduction

## 2 ANOVA

- Understanding ANOVA
- ANOVA In Practice
- The Nonparametric Equivalence of ANOVA
- Multiple Comparison Tests
- ANOVA Assumptions Checking

# Assumptions Checking

We should always check the assumptions made for the F-test.

- Equal variances for all groups: use the Bartlett test or Levene test.
- Errors are independent normally distributed:
  - Do a Normality test on residuals.
  - Look at Residual plots (QQ plot on residuals; Plot residuals against the groups).



# Assumptions Checking in R: Normality Test

A normality test has  $H_0$  that the sample (that is tested) follows normal distribution.

- Shapiro Wilk test:

```
> shapiro.test(amount) # for the variable amount  
Shapiro-Wilk normality test
```

```
data: amount
```

```
W = 0.9745, p-value = 0.1634
```

```
> shapiro.test(anova$res) # for the residuals  
Shapiro-Wilk normality test
```

```
data: anova$res
```

```
W = 0.95254, p-value = 0.009873
```

- Kolmogorov-Smirnov test:

```
> anova = aov(amount~lab, data = newdata)  
> resid=anova$res  
> ks.test(resid,"pnorm",mean(resid),sd(resid))  
One-sample Kolmogorov-Smirnov test
```

```
data: resid
```

# Assumptions Checking in R: Equal Variances Test

- Bartlett test (when samples are assumed normality):

```
> res <- bartlett.test(amount ~ lab, data = newdata)
> res
```

Bartlett test of homogeneity of variances

data: amount by lab

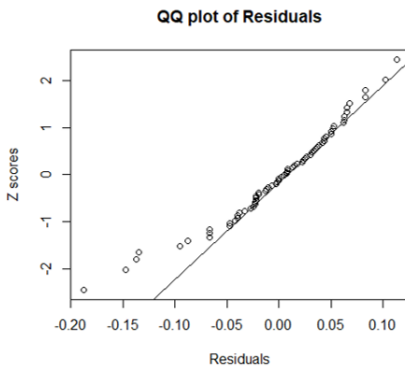
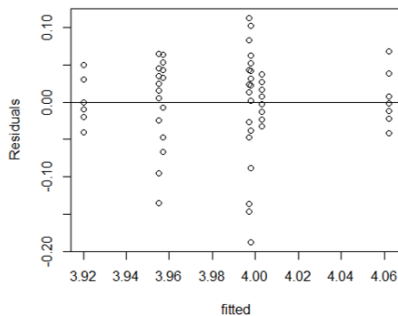
Bartlett's K-squared = 24.37, df = 6, p-value = 0.0004465

- Levene test (suitable if distribution of sample is unknown)

```
> leveneTest(amount ~ lab, data = newdata)
Levene's Test for Homogeneity of Variance (center = median)
  Df F value  Pr(>F)
group 6  2.2445 0.05018 .
    63

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

# Assumptions Checking in R: Plots



# Assumptions Checking in Python (1)

```
##### Equal Variances TEST:
```

```
#Bartlett test (when samples are assumed normality):
```

```
Bar = scipy.stats.bartlett(lab1, lab2, lab3, lab4, lab5, lab6, lab7)
print(Bar)
```

```
BartlettResult(statistic=24.369716004698912, pvalue=0.0004464732077665512)
```

```
##### Equal Variances TEST:
```

```
#Levene test:
```

```
Levene = scipy.stats.levene(lab1, lab2, lab3, lab4, lab5, lab6, lab7)
print(Levene)
```

```
LeveneResult(statistic=2.2444617178390893, pvalue=0.050181442540930994)
```

## Assumptions Checking in Python (2)

### Shapiro Wilk Test for normality checking

```
##### normality checking :  
  
# Shappiro Wilk Test:  
from scipy.stats import shapiro  
Shapiro1 = scipy.stats.shapiro(amount) # for the variable amount  
print(Shapiro1)  
Shapiro2 = scipy.stats.shapiro(mod.resid) # for the residuals  
print(Shapiro2)  
  
(0.9744956493377686, 0.16336943209171295)  
(0.952541708946228, 0.0098727997392416)
```

# Assumptions Checking in Python (3)

## Kolmogorov Smirnov Test of normality

```
##### normality checking :  
# Kolmogorov Smirnov Test for the variable amount  
test = np.random.normal(mean(amount), np.std(amount), 70 )  
KS = scipy.stats.ks_2samp(amount,test)  
print(KS)
```

```
Ks_2sampResult(statistic=0.12857142857142856, pvalue=0.5795548581933454)
```

```
##### normality checking :  
# Kolmogorov Smirnov Test for the residuals  
res = np.random.normal(mean(mod.resid), np.std(mod.resid), 70 ) #(***)  
KS_res = scipy.stats.ks_2samp(mod.resid,res)  
print(KS_res) # p-value can be different due to the randomness in (***)
```

```
Ks_2sampResult(statistic=0.08571428571428574, pvalue=0.9504984566743319)
```

# Assumptions Checking in SAS (1)

## The SAS code

```
* Test for equal variances: Levene test ;
PROC ANOVA data =newdata;
class lab;
model amount = lab;
means lab / hovtest = levene alpha=0.05;
run;

* Test for equal variances: Bartlett test ;
PROC ANOVA data =newdata;
class lab;
model amount = lab;
means lab / HOVTEST=BARTLETT alpha=0.05;
run;

*Test for normality ;
proc univariate data=newdata normal ;
var amount;
histogram amount /normal;
qqplot /normal (mu=est sigma=est);
run;
```

## Assumptions Checking in SAS (2)

The output for Levene and Bartlett tests of equal variances.

### The ANOVA Procedure

Levene's Test for Homogeneity of amount Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
lab	6	0.000450	0.000075	2.45	0.0344
Error	63	0.00193	0.000031		

### The ANOVA Procedure

Bartlett's Test for Homogeneity of amount Variance			
Source	DF	Chi-Square	Pr > ChiSq
lab	6	24.3697	0.0004



## Assumptions Checking in SAS (3)

The output for normality checking (of variable amount) by tests.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.974495	Pr < W	0.1634
Kolmogorov-Smirnov	D	0.099306	Pr > D	0.0862
Cramer-von Mises	W-Sq	0.133501	Pr > W-Sq	0.0404
Anderson-Darling	A-Sq	0.708729	Pr > A-Sq	0.0646