

University of California, Los Angeles
Department of Statistics

Statistics 100B

Instructor: Nicolas Christou

Analysis of variance (ANOVA)

We wish to test the equality of k population means. We have k normal populations:
 $Y_1 \sim N(\mu_1, \sigma)$, $Y_2 \sim N(\mu_2, \sigma)$, \dots , $Y_k \sim N(\mu_k, \sigma)$. The null and alternative hypotheses are:

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

H_a : At least 2 means are not equal

To test the above hypothesis we select:

n_1	observations from population	1
n_2	observations from population	2
\vdots	\vdots	\vdots
n_k	observations from population	k

Total number of observations: $N = n_1 + n_2 + \dots + n_k$.

Set-up:

	Sample from the i th population	
Overall Mean \bar{Y}	y_{11}	\bar{Y}_1
	y_{12}	
	\vdots	
	y_{1n_1}	
	y_{21}	\bar{Y}_2
	y_{22}	
	\vdots	
	y_{2n_2}	
	\vdots	\vdots
	\vdots	
	\vdots	
	\vdots	\vdots
	\vdots	
\vdots		
\vdots		
	y_{k1}	\bar{Y}_k
	y_{k2}	
	\vdots	
	y_{knk}	

Define:

$$\text{Total sum of squares } SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2$$

$$\text{Within sum of squares } WSS = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

$$\text{Between sum of squares } BSS = \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2$$

It is true that $SST = WSS + BSS$.

ANOVA table:

Source	d.f.	SS	MS	F
Between	$k - 1$	BSS	$MBSS = \frac{BSS}{k-1}$	$\frac{MBSS}{MWSS}$
Within	$N - k$	WSS	$MWSS = \frac{WSS}{N-k}$	
Total	$N - 1$	SST		

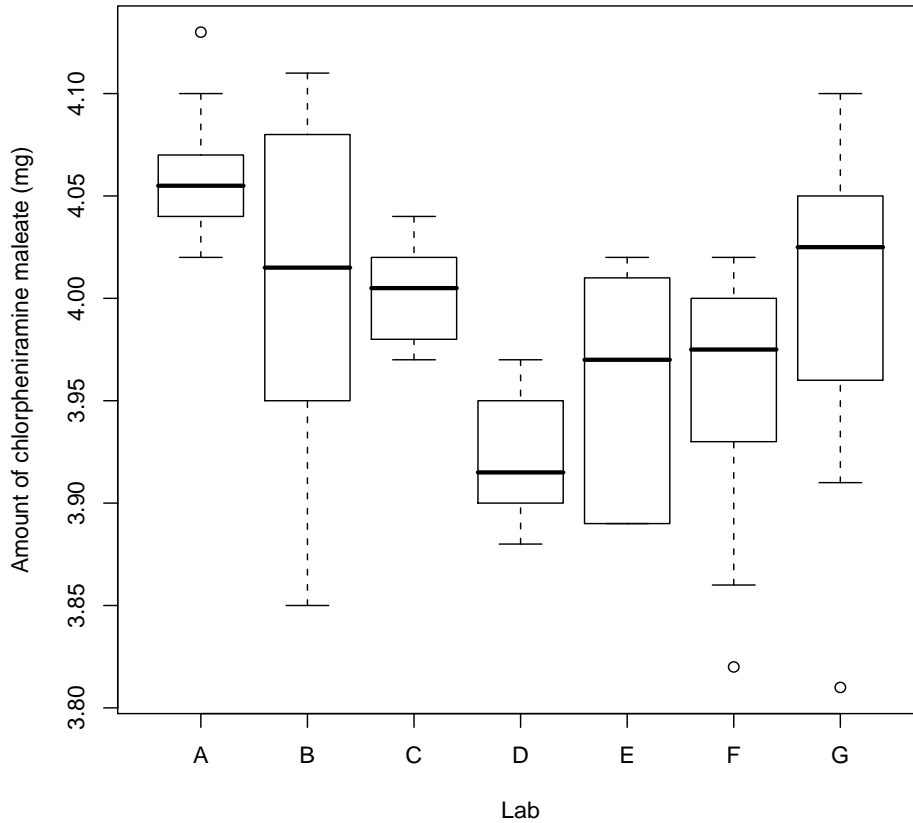
Example: From John Rice, *Mathematical Statistics and Data Analysis*, Second Edition, Duxbury Press (1995), pp. 443-444.

For each of four manufacturers, composites were prepared by grinding and mixing together tablets in order to measure the amount of chlorpheniramine maleate. Seven labs were asked to make 10 determinations on each composite. The purpose of the experiment was to study the consistency between labs. The data for the 7 labs are shown on the table below. The data set as it was read in R appears on the next page.

Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
4.13	3.86	4.00	3.88	4.02	4.02	4.00
4.07	3.85	4.02	3.88	3.95	3.86	4.02
4.04	4.08	4.01	3.91	4.02	3.96	4.03
4.07	4.11	4.01	3.95	3.89	3.97	4.04
4.05	4.08	4.04	3.92	3.91	4.00	4.10
4.04	4.01	3.99	3.97	4.01	3.82	3.81
4.02	4.02	4.03	3.92	3.89	3.98	3.91
4.06	4.04	3.97	3.90	3.89	3.99	3.96
4.10	3.97	3.98	3.97	3.99	4.02	4.05
4.04	3.95	3.98	3.90	4.00	3.93	4.06

The boxplots below show some differences in the medians, the quartile range, and the variation among the seven laboratories. Are these differences significant?

Boxplots of determination of amounts of chlorpheniramine maleate in tablets by seven laboratories



Summary statistics:

Lab1	Lab2	Lab3	Lab4	Lab5
Min. :4.020	Min. :3.850	Min. :3.970	Min. :3.880	Min. :3.890
1st Qu.:4.040	1st Qu.:3.955	1st Qu.:3.982	1st Qu.:3.900	1st Qu.:3.895
Median :4.055	Median :4.015	Median :4.005	Median :3.915	Median :3.970
Mean :4.062	Mean :3.997	Mean :4.003	Mean :3.920	Mean :3.957
3rd Qu.:4.070	3rd Qu.:4.070	3rd Qu.:4.018	3rd Qu.:3.942	3rd Qu.:4.008
Max. :4.130	Max. :4.110	Max. :4.040	Max. :3.970	Max. :4.020

Lab6	Lab7
Min. :3.820	Min. :3.810
1st Qu.:3.938	1st Qu.:3.970
Median :3.975	Median :4.025
Mean :3.955	Mean :3.998
3rd Qu.:3.998	3rd Qu.:4.048
Max. :4.020	Max. :4.100

```

> a <- read.table("anova_example.txt", header=TRUE)
> a
  amount lab
1    4.13  A
2    4.07  A
3    4.04  A
4    4.07  A
5    4.05  A
6    4.04  A
7    4.02  A
8    4.06  A
9    4.10  A
10   4.04  A
11   3.86  B
12   3.85  B
13   4.08  B
14   4.11  B
15   4.08  B
16   4.01  B
17   4.02  B
18   4.04  B
19   3.97  B
20   3.95  B
21   4.00  C
22   4.02  C
23   4.01  C
24   4.01  C
25   4.04  C
26   3.99  C
27   4.03  C
28   3.97  C
29   3.98  C
30   3.98  C
31   3.88  D
32   3.88  D
33   3.91  D
34   3.95  D
35   3.92  D
36   3.97  D
37   3.92  D
38   3.90  D
39   3.97  D
40   3.90  D
41   4.02  E
42   3.95  E
43   4.02  E
44   3.89  E
45   3.91  E
46   4.01  E
47   3.89  E
48   3.89  E
49   3.99  E
50   4.00  E
51   4.02  F
52   3.86  F
53   3.96  F
54   3.97  F
55   4.00  F
56   3.82  F
57   3.98  F
58   3.99  F
59   4.02  F
60   3.93  F
61   4.00  G
62   4.02  G
63   4.03  G
64   4.04  G
65   4.10  G
66   3.81  G
67   3.91  G
68   3.96  G
69   4.05  G
70   4.06  G

```

ANOVA in R:

```

> g<-lm(a$amount ~ a$lab)
> anova(g)
Analysis of Variance Table

Response: a$amount
      Df  Sum Sq Mean Sq F value    Pr(>F)
a$lab   6  0.124737  0.020790  5.6601 9.453e-05 ***
Residuals 63  0.231400  0.003673
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```