بسم الله الرحمن الرحيم

# L XVII

# ANALYSIS OF VARIANCE

# ANOVA

• . Prof. Dr. Waqar AL-Kubaisy

Example

A new analgesic drug has been proposed by a pharmaceutical house. It is desired to compare the effect of drug with aspirin and placebo for use in treatment of simple headache. The variable measure is the number of hours a patient is free from pain following administration of the drug. A small pilot study,2 patients given placebo, 4 patients were given the new drug and 3 patients were given aspirin the number of hours patient free of pain as follows

| | | | | |
|---|---|---|---|---|
| Placebo | 00 | 1.0 | | |
| New drug | 2.3 | 3.5 | 2.8 | 2.5 |
| Aspirin | 3.1 | 2.7 | 3.8 | |

Are the group means significantly differ at level of significance 95%

- **In t test we were testing hypothesis concerning the means of two populations or samples**
- In many experiments situations samples are selected from **several different populations,**
- the problems in such situations is to determine whether there are any deference among the population means.
- More precisely we frequently want to test the
- Ho that there is no difference in the means of the populations from which the separated samples have been drown.

- For example
- **analysis of H2O samples drown from different sampling points in a city to see if there is a significant variation in the mean H2Oquality of sampling points**. Or

- serum cholesterol levels determine from serum pool of serum by several different among the examiners means.

- **So the technique used here to test the hypothesis is the Analysis of Variance**

- In ANOVA
- **we are testing the hypothesis concerning more than two population means.**
- So ANOVA is used to test the significance of the difference between more than two population means.
- In ANOVA
- we want to decide whether observed **difference among more than** two samples means can be attributed to chance or
- whether there are real different among the picked up samples
- **It is based on the following**
- **is there significantly more variation among the group means than there is within the groups**

- The term analysis of variance is used because the total variability in the set of data can be broken up into :

❖ Some of variability Among the sample means and

❖ the variability Within sample

- The collected or pooled variation within groups is used as a standard to comparison, because it measuring the inherent observational variability in the data .

- A differences in means should be large relative to the inherent variability

- Total variability in compete set of data $\Sigma\,\Sigma(X - \bar{X})$ _
- **The total variance** :
  Deviation of **observation X** in all group from the **over all mean $\bar{X}$**

- **Between variance Between** $\frac{SQ}{K-1}$ K = No. of groups d.f.= K-1
- Deviation of **group means** from the over **all mean ( $\bar{X}$)** Measures
- for variability Among(Between) the samples means is
- $\Sigma(\bar{X}_i - \bar{X})^2$

- **Within Variation : Within SQ**
- **N---K**
- Deviation of observation in each group from its mean
- $\Sigma(X_i - \bar{X}_i)^2$ **in each group**
- **d.f. N-K N= No. of observations**

_ 2

## Between (Among) variance

$$\frac{SQ}{K-1}$$

K = No. of groups

Deviation of group means ($\bar{X}i$) from the over all mean ( $\bar{X}$)

$$\Sigma ( \bar{X}i - \bar{X})^2$$

d.f.= K-1

**Within Variation** : $\frac{\text{Within SQ}}{\text{N---K}}$

Deviation of observation Xi in each group from its mean ($\bar{Xi}$)

$\Sigma ( Xi - \bar{Xi})^2$ in each group

d.f. N-K N= No. of observations

The total variability **within** the samples would be obtained by

summing the $\Sigma \Sigma ( X_i - \bar{X_i} )^2$

Total variability in full set of data should be the **Sum of the** variability **within samples** **plus** the variability **Among (Between)** sample

**We compare** the Among (Between) samples mean square to **within** samples mean square.

- If the ratio is larger than could reasonably attributed to chance factor we reject the hypothesis .
- If this ratio is not too large we accept the Ho
- Thus we have the ratio of the Among (between) sample mean square to the within samples mean square is our test statistics

**the distribution of F test is likely to be one tailed test**

**calculated F > tabulated → Reject Ho → signif. Differences**

- **Calculated F < tabulated → Accept Ho → no signifi.**

- **Although it is called Analysis of variance ,**
- **it is basically a method for study variation among means.**
- **This variation being measured as by a variance**

- **Analysis of Variance table**

| Source of Variation | SS | df | MS | F Statistic | p Value |
|---|---|---|---|---|---|
| Between samples | SSB | K−1 | MSB | MSB/MSW | p |
| Within samples | SSW | N− K | MSW | | |
| Total | SST | N−1 | | | |

| Source of Variation | Sum of square | d.f | mean of sum of square |
|---|---|---|---|
| Between samples | SSB<br>$\sum\sum(\bar{X}_i - \bar{\bar{X}})^2$<br>$\sum \frac{t_i^2}{n_i} - \frac{T^2}{N}$ | k−1 | SSB/ K−1<br>$\sum\sum(\bar{X}_i - \bar{\bar{X}})^2 / K-1$<br>$\sum \frac{t_i^2}{n_i} - \frac{T^2}{N} / K-1$ |
| Within samples | SSW<br>$\sum\sum(X - \bar{X}_i)^2$<br>$\sum\sum X^2 - \frac{\sum t_i^2}{n_i}$ | N−k | SSW/N− K<br>$\sum\sum X^2 - \frac{\sum t_i^2}{n_i} / N-K$ |

$$F = \frac{\left(\frac{\Sigma t_i^2}{n_i} - \frac{T^2}{N}\right) / K-1}{\left(\Sigma\Sigma X^2 - \frac{\Sigma t_i^2}{n_i}\right) / N-K}$$

$$\frac{SSB/K-1}{SSW/N-K}$$

$$F = \frac{\left(\frac{\Sigma t_i^2}{n_i} - \frac{T^2}{N}\right) / K-1}{\left(\Sigma\Sigma X^2 - \frac{\Sigma t_i^2}{n_i}\right) / N-K}$$

Example

A new analgesic drug has been proposed by a pharmaceutical house, it is desired to compare the effect of drug with aspirin and placebo for use in treatment of simple headache. The variable measure is the **number of hours a patient is free from pain** following administration of the drug in a small pilot study. 2 patients given placebo, 4 patients were given the new drug and 3 patients were given aspirin. The **number of hours** patient free of pain as follows

| | | | | |
|---|---|---|---|---|
| Placebo | 00 | 1.0 | | |
| New drug | 2.3 | 3.5 | 2.8 | 2.5 |
| Aspirin | 3.1 | 2.7 | 3.8 | |

Are the group means significantly differ at level of significance 95%

| | | | | |
|---|---|---|---|---|
| Placebo | 00 | 1.0 | | |
| New drug | 2.3 | 3.5 | 2.8 | 2.5 |
| Aspirin | 3.1 | 2.7 | 3.8 | |

$n_1 = 2$ $n_2 = 4$ $n_3 = 3$

$t_1 = 1$ $t_2 = 11.1$ $t_3 = 9.6$

$N = 9$

$K = 3$

$$F = \frac{\left(\frac{\Sigma t_i^2}{n_i} - \frac{T^2}{N}\right) / K - 1}{\left(\Sigma\Sigma X^2 - \frac{\Sigma t_i^2}{n_i}\right) / N - K}$$

**ANOVA**

$$F=\frac{\left(\frac{\Sigma t_i^2}{n_i}-\frac{T^2}{N}\right)/K-1}{\left(\Sigma\Sigma X^2-\frac{\Sigma t_i^2}{n_i}\right)/N-K}$$

$$\frac{\Sigma t_i^2}{n}=\frac{1^2}{2}+\frac{(2.3+3.5+2.8+2.5)^2}{4}+\frac{(3.1+2.7+3.8)^2}{3}$$

$$=\frac{1^2}{2}+\frac{(11.1)^2}{4}+\frac{(9.6)^2}{3}=62.02$$

$$\frac{T^2}{N}=\frac{(21.7)^2}{9}=52.32$$

$$\Sigma\Sigma X_i^2=(00)^2+(1)^2+(2.3)^2+(3.5)^2+(2.8)^2+(2.5)^2+(3.1)^2+(2.7)^2+(3.8)^2$$

$$=63.9$$

**Among (between)** sample means

$$\Sigma\Sigma(\bar{X}_i - \bar{X})^2 / K-1$$



$$F = \frac{\left(\frac{\Sigma t_i^2}{n_i} - \frac{T^2}{N}\right) / K - 1}{\left(\Sigma\Sigma X^2 - \frac{\Sigma t_i^2}{n_i}\right) / N - K}$$

$$\left(\Sigma \frac{t_i^2}{n_i} - \frac{T^2}{N}\right) / K-1 = \frac{62.02 - 52.32}{K-1}$$

$$= \frac{62.02 - 52.32}{2} = 4.85$$

**Within samples means square** SSW/N−K

$$\left(\Sigma\Sigma X^2 - \frac{\Sigma t_i^2}{n_i}\right) / N - K =$$

$$\frac{63.9 - 62.02}{N - K} = \frac{63.9 - 62.02}{9 - 3} = \frac{63.9 - 62.02}{6}$$

$$= 0.325$$

$$F = \frac{4.85}{0.325} = 14.92$$

$$F = \frac{\left(\sum \frac{t_i^2}{n_i} - \frac{T^2}{N}\right) / K-1}{\left(\sum\sum X^2 - \sum \frac{t_i^2}{n_i}\right) / N-K}$$

K-1 numerator

N-K denominator d.f. = 2,6

Tabulated F = 5.14

Calculated F > tabulated

Reject Ho Taking the HA

There is a significant difference between the means of the samples

$P < 0.05$

# F Distribution critical values

# F Distribution critical values for P=0.10



| Denominator | Numerator DF | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DF | 1 | 2 | 3 | 4 | 5 | 7 | 10 | 15 | 20 | 30 | 60 | 120 | 500 | 1000 |
| 1 | 39.864 | 49.500 | 53.593 | 55.833 | 57.240 | 58.906 | 60.195 | 61.220 | 61.740 | 62.265 | 62.794 | 63.061 | 63.264 | 63.296 |
| 2 | 8.5264 | 8.9999 | 9.1618 | 9.2434 | 9.2926 | 9.3491 | 9.3915 | 9.4248 | 9.4413 | 9.4580 | 9.4745 | 9.4829 | 9.4893 | 9.4902 |
| 3 | 5.5384 | 5.4624 | 5.3907 | 5.3426 | 5.3092 | 5.2661 | 5.2304 | 5.2003 | 5.1845 | 5.1681 | 5.1513 | 5.1425 | 5.1358 | 5.1347 |
| 4 | 4.5448 | 4.3245 | 4.1909 | 4.1073 | 4.0505 | 3.9790 | 3.9198 | 3.8704 | 3.8443 | 3.8175 | 3.7896 | 3.7753 | 3.7643 | 3.7625 |
| 5 | 4.0605 | 3.7798 | 3.6194 | 3.5202 | 3.4530 | 3.3679 | 3.2974 | 3.2379 | 3.2067 | 3.1740 | 3.1402 | 3.1228 | 3.1094 | 3.1071 |
| 7 | 3.5895 | 3.2575 | 3.0740 | 2.9605 | 2.8833 | 2.7850 | 2.7025 | 2.6322 | 2.5947 | 2.5555 | 2.5142 | 2.4927 | 2.4761 | 2.4735 |
| 10 | 3.2850 | 2.9244 | 2.7277 | 2.6054 | 2.5216 | 2.4139 | 2.3226 | 2.2434 | 2.2007 | 2.1554 | 2.1071 | 2.0818 | 2.0618 | 2.0587 |
| 15 | 3.0731 | 2.6951 | 2.4898 | 2.3615 | 2.2729 | 2.1582 | 2.0593 | 1.9722 | 1.9243 | 1.8727 | 1.8168 | 1.7867 | 1.7629 | 1.7590 |
| 20 | 2.9746 | 2.5893 | 2.3801 | 2.2490 | 2.1582 | 2.0397 | 1.9368 | 1.8450 | 1.7939 | 1.7383 | 1.6768 | 1.6432 | 1.6163 | 1.6118 |
| 30 | 2.8808 | 2.4887 | 2.2761 | 2.1423 | 2.0493 | 1.9269 | 1.8195 | 1.7222 | 1.6674 | 1.6064 | 1.5376 | 1.4990 | 1.4669 | 1.4617 |
| 60 | 2.7911 | 2.3932 | 2.1774 | 2.0409 | 1.9457 | 1.8194 | 1.7070 | 1.6034 | 1.5435 | 1.4756 | 1.3953 | 1.3476 | 1.3060 | 1.2989 |
| 120 | 2.7478 | 2.3473 | 2.1300 | 1.9924 | 1.8959 | 1.7675 | 1.6523 | 1.5450 | 1.4821 | 1.4094 | 1.3203 | 1.2646 | 1.2123 | 1.2026 |
| 500 | 2.7157 | 2.3132 | 2.0947 | 1.9561 | 1.8588 | 1.7288 | 1.6115 | 1.5009 | 1.4354 | 1.3583 | 1.2600 | 1.1937 | 1.1215 | 1.1057 |
| 1000 | 2.7106 | 2.3080 | 2.0892 | 1.9505 | 1.8530 | 1.7228 | 1.6051 | 1.4941 | 1.4281 | 1.3501 | 1.2500 | 1.1813 | 1.1031 | 1.0844 |

# F Distribution critical values for P=0.05

ANOVA

**Denominator** N–K

**Numerator DF** K–1

| DF | 1 | 2 | 3 | 4 | 5 | 7 | 10 | 15 | 20 | 30 | 60 | 120 | 500 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161.45 | 199.50 | 215.71 | 224.58 | 230.16 | 236.77 | 241.88 | 245.95 | 248.01 | 250.10 | 252.20 | 253.25 | 254.06 | 254.19 |
| 2 | 18.513 | 19.000 | 19.164 | 19.247 | 19.296 | 19.353 | 19.396 | 19.429 | 19.446 | 19.462 | 19.479 | 19.487 | 19.494 | 19.495 |
| 3 | 10.128 | 9.5522 | 9.2766 | 9.1172 | 9.0135 | 8.8867 | 8.7855 | 8.7028 | 8.6602 | 8.6165 | 8.5720 | 8.5493 | 8.5320 | 8.5292 |
| 4 | 7.7086 | 6.9443 | 6.5915 | 6.3882 | 6.2560 | 6.0942 | 5.9644 | 5.8579 | 5.8026 | 5.7458 | 5.6877 | 5.6580 | 5.6352 | 5.6317 |
| 5 | 6.6078 | 5.7862 | 5.4095 | 5.1922 | 5.0504 | 4.8759 | 4.7351 | 4.6187 | 4.5582 | 4.4958 | 4.4314 | 4.3985 | 4.3731 | 4.3691 |
| 7 | 5.5914 | 4.7375 | 4.3469 | 4.1202 | 3.9715 | 3.7871 | 3.6366 | 3.5108 | 3.4445 | 3.3758 | 3.3043 | 3.2675 | 3.2388 | 3.2344 |
| 10 | 4.9645 | 4.1028 | 3.7082 | 3.4780 | 3.3259 | 3.1354 | 2.9782 | 2.8450 | 2.7741 | 2.6996 | 2.6210 | 2.5801 | 2.5482 | 2.5430 |

# F Distribution critical values for P=0.05

ANOVA

**Denominator** N □K

| | Numerator DF K □1 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DF** | **1** | **2** | **3** | **4** | **5** | **7** | **10** | **15** | **20** | **30** | **60** | **120** | **500** | **1000** |
| **15** | 4.5431 | 3.6823 | 3.2874 | 3.0556 | 2.9013 | 2.7066 | 2.5437 | 2.4035 | 2.3275 | 2.2467 | 2.1601 | 2.1141 | 2.0776 | 2.0718 |
| **20** | 4.3512 | 3.4928 | 3.0983 | 2.8660 | 2.7109 | 2.5140 | 2.3479 | 2.2032 | 2.1241 | 2.0391 | 1.9463 | 1.8962 | 1.8563 | 1.8498 |
| **30** | 4.1709 | 3.3159 | 2.9223 | 2.6896 | 2.5336 | 2.3343 | 2.1646 | 2.0149 | 1.9317 | 1.8408 | 1.7396 | 1.6835 | 1.6376 | 1.6300 |
| **60** | 4.0012 | 3.1505 | 2.7581 | 2.5252 | 2.3683 | 2.1666 | 1.9927 | 1.8365 | 1.7480 | 1.6492 | 1.5343 | 1.4672 | 1.4093 | 1.3994 |
| **120** | 3.9201 | 3.0718 | 2.6802 | 2.4473 | 2.2898 | 2.0868 | 1.9104 | 1.7505 | 1.6587 | 1.5544 | 1.4289 | 1.3519 | 1.2804 | 1.2674 |
| **500** | 3.8601 | 3.0137 | 2.6227 | 2.3898 | 2.2320 | 2.0278 | 1.8496 | 1.6864 | 1.5917 | 1.4820 | 1.3455 | 1.2552 | 1.1586 | 1.1378 |
| **1000** | 3.8508 | 3.0047 | 2.6137 | 2.3808 | 2.2230 | 2.0187 | 1.8402 | 1.6765 | 1.5811 | 1.4705 | 1.3318 | 1.2385 | 1.1342 | 1.1096 |

$$F = \frac{\left(\frac{\Sigma t_i^2}{n_i} - \frac{T^2}{N}\right) / (K-1)}{\left(\Sigma\Sigma X^2 - \frac{\Sigma t_i^2}{n_i}\right) / (N-K)}$$

$$F = \frac{4.85}{0.325} = 14.92$$

K-1 numerator

N-K denominator d.f. = 2,6

Tabulated F = 5.14

Calculated F > tabulated

Reject Ho Taking the HA

There is a significant difference between the means of the samples

$P < 0.05$

**QIII (10 marks)**

**Twelve** obese patients were selected randomly and assigned into **three groups** of treatments (A, B, &C group). Their weight were determined six months later and weight loss in Kg for those patients were determined as follows

| **Group A** | **Group B** | **Group C** |
|---|---|---|
| 4 | 3 | 12 |
| 7 | 5 | 8 |
| 6 | 2 | 9 |
| 3 | | 11 |
| 2 | | |

At alpha 0.05 can we conclude that there is a significant difference in the mean weight loss among these groups

# Post Hoc Analysis

To find which means are significantly different,

we might be tempted to perform a number of multiple t test between the various pairs of means.

Multiple t tests are inappropriate, however, because the probability of incorrectly rejecting the hypothesis increases with the number of t tests performed

A significant F ratio tell us that there are differences between at least one pair of means.

- **The purpose of post hoc** analysis **is to**
- **find out exactly where those differences are .**
- **Variety of different types of post hoc analysis allows**
- **to make multiple pair wise comparisons and**
- **determine which pairs are significantly different and which are not.**

**The interpretation is similar to the two-sample t test.**

- **The more popular post hoc procedures include**
- Tukey, Tukey Kramer, Scheffe, Bonferroni, Dunnett and Games-Howell

# Thank You