



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 <u>among</u> 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

<u>Total variability in compete</u> set of data $\Sigma \Sigma(X - X)$ _ **The total variance** :

Deviation of **observation X** in all group from the **over all mean X**

- Between variance Between SQ K = No. of groups d.f.= K-1
 K-1 _____
- Deviation of **group means** from the over **all mean (X)** Measures
- for variability Among(Between) the samples means is

 $\Sigma (Xi - X)^2$

- <u>Within Variation</u>: <u>Within SQ</u>
 N---K
- Deviation of observation in each group from its mean
- Σ (Xi Xi⁻)² in each group
- d.f. N-K N= No. of observations



Between (Among) variance

<u>SQ</u> K-1

K = No. of groups

Deviation of group means (Xi) from the over all mean (X) $\Sigma (Xi - X)$

d.f.= K-1

Within Variation :Within SQN---K

Deviation of observation Xi in each group from its mean (Xi)

Σ (Xi - Xi) 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 (Xi - Xi)^2
```

Total variability in full set of data should bethe Sum of the variability within samplesplusthe 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

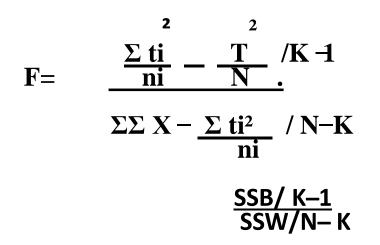
the distribution of F test is likely to be one tailed test calculated F > tabulated <u>Reject Ho</u>signif. Differences

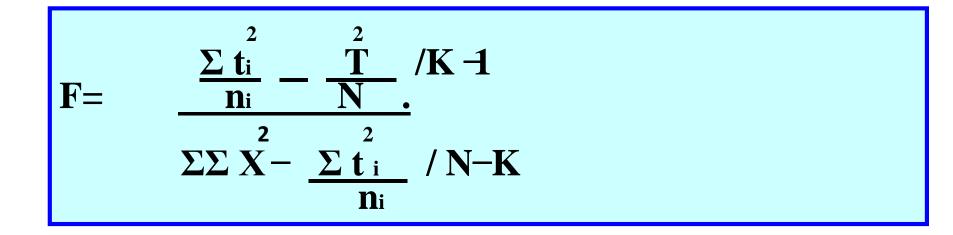
- Calculated F < tabulated Accept Ho no signifi.
- Although it is called Analysis of variance ,
- it is basically a method for study variation among means.
- This variation being measured as by a variance

• Analysis of Variance table

| Source of Variation | SS | df | MS | F Statistic | p Value |
|------------------------|-----|------------------|-----|-------------|---------|
| Between samples | SSB | K - 1 | MSB | MSB/MSW | р |
| 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 $\Sigma \Sigma (\overline{Xi} - \overline{X})^{2}$ $\Sigma \frac{1}{\Sigma t_{i}}^{2} - \frac{T}{N}^{2}$ | k–1 | SSB/ K-1 $\Sigma\Sigma \overline{(Xi-\overline{X})^2} / K-1$ $\Sigma \underline{ti^2} - \underline{T^2} / K-1$ $\overline{n_i} N$ |
| Within samples | $SSW = \overline{\Sigma}\Sigma (X - X\overline{i})^{2}$ $\Sigma\Sigma X^{2} - \sum_{\substack{i \\ n_{i}}}^{2}$ | N-k | $\begin{array}{c c} SSW/N-K \\ 2 & 2 \\ \Sigma\Sigma X- \underline{\Sigma t_i} \\ n_i \end{array} /N-K \\ 15 \end{array}$ |





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

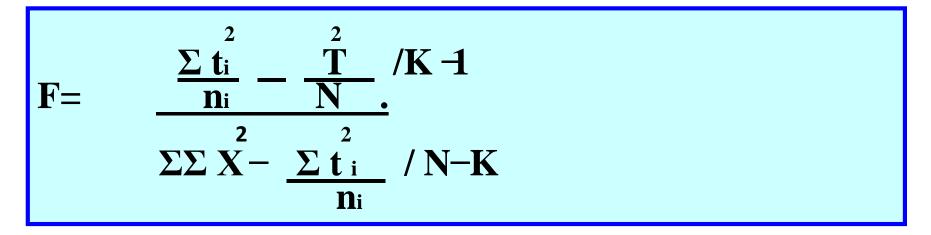
Placebo001.0New drug2.33.52.82.5Aspirin3.12.73.8

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

Placebo New drug Aspirin

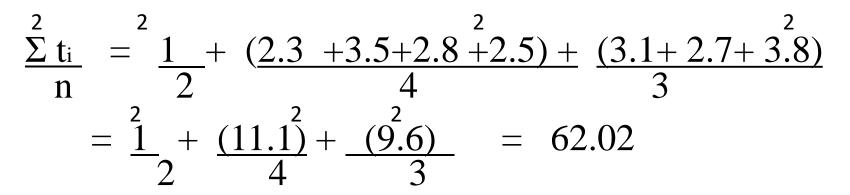
| 00 | 1.0 | | |
|-----------|-----|--------------|-------------|
| 2.3 | 3.5 | 2.8 | 2.5 |
| 3.1 | 2.7 | 3.8 | |
| n 1 =2 | | n2 =4 | n3 =3 |
| $t_1 = 1$ | | $t_2 = 11.1$ | $t_3 = 9.6$ |
| | | | |

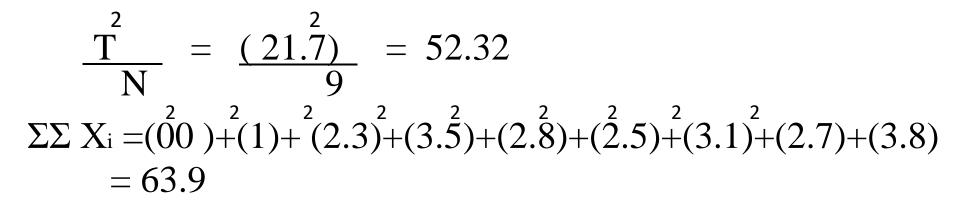
N= 9 K= 3

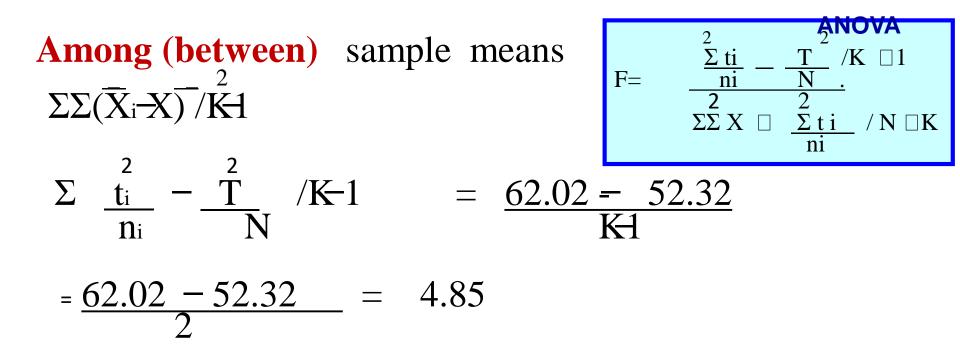


$$F = \frac{2}{\frac{\Sigma ti}{ni}} - \frac{2}{\frac{T}{N}} / K \square 1$$

$$\frac{2}{\Sigma\Sigma X} \square \frac{2}{\frac{\Sigma ti}{ni}} / N \square K$$







Within samples means square SSW/N \Box K $\Sigma\Sigma \stackrel{2}{X} - \stackrel{2}{\Sigma \stackrel{1}{t_i}}_{n_i}$ /N-K =

 $\frac{63.9 - 62.02}{N - K} = \frac{63.9 - 62.02}{9 - 3} = \frac{63.9 - 62.02}{6}$ = 0.325



K-1 numerator N-K denominator d.f. = 2,6Tabulated 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

Т

¥

| Numerator DI |
|---------------------|
|---------------------|

| | | | | | | | | | - | | | | | |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 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

Denominator N-K

| | Num | erato | r DF | K -1 | | | | | | | | | | |
|----|-------|-------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| DF | 1 | 2 | 3 | 4 | 5 | 7 | 10 | 15 | 20 | 30 | 60 | 120 | 500 | 1000 |
| ¥ | 161.4 | 199.5 | 215.7 | 224.5 | 230.1 | 236.7 | 241.8 | 245.9 | 248.0 | 250.1 | 252.2 | 253.2 | 254.0 | 254.1 |
| 1 | 5 | 0 | 1 | 8 | 6 | 7 | 8 | 5 | 1 | 0 | 0 | 5 | 6 | 9 |
| 2 | 18.51 | 19.00 | 19.16 | 19.24 | 19.29 | 19.35 | 19.39 | 19.42 | 19.44 | 19.46 | 19.47 | 19.48 | 19.49 | 19.49 |
| | 3 | 0 | 4 | 7 | 6 | 3 | 6 | 9 | 6 | 2 | 9 | 7 | 4 | 5 |
| 3 | 10.12 | 9.552 | 9.276 | 9.117 | 9.013 | 8.886 | 8.785 | 8.702 | 8.660 | 8.616 | 8.572 | 8.549 | 8.532 | 8.529 |
| | 8 | 2 | 6 | 2 | 5 | 7 | 5 | 8 | 2 | 5 | 0 | 3 | 0 | 2 |
| 4 | 7.708 | 6.944 | 6.591 | 6.388 | 6.256 | 6.094 | 5.964 | 5.857 | 5.802 | 5.745 | 5.687 | 5.658 | 5.635 | 5.631 |
| | 6 | 3 | 5 | 2 | 0 | 2 | 4 | 9 | 6 | 8 | 7 | 0 | 2 | 7 |
| 5 | 6.607 | 5.786 | 5.409 | 5.192 | 5.050 | 4.875 | 4.735 | 4.618 | 4.558 | 4.495 | 4.431 | 4.398 | 4.373 | 4.369 |
| | 8 | 2 | 5 | 2 | 4 | 9 | 1 | 7 | 2 | 8 | 4 | 5 | 1 | 1 |
| 7 | 5.591 | 4.737 | 4.346 | 4.120 | 3.971 | 3.787 | 3.636 | 3.510 | 3.444 | 3.375 | 3.304 | 3.267 | 3.238 | 3.234 |
| | 4 | 5 | 9 | 2 | 5 | 1 | 6 | 8 | 5 | 8 | 3 | 5 | 8 | 4 |
| 10 | 4.964 | 4.102 | 3.708 | 3.478 | 3.325 | 3.135 | 2.978 | 2.845 | 2.774 | 2.699 | 2.621 | 2.580 | 2.548 | 2.543 |
| | 5 | 8 | 2 | 0 | 9 | 4 | 2 | 0 | 1 | 6 | 0 | 1 | 2 | 0 |

F Distribution critical values for P=0.05

Denominator N \square K

| | Num | erator | DF | K □1 | | | | | | | | | | |
|-----|-------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| DF | 1 | 2 | 3 | 4 | 5 | 7 | 10 | 15 | 20 | 30 | 60 | 120 | 500 | 1000 |
| ▼ | 4.543 | 3.682 | 3.287 | 3.055 | 2.901 | 2.706 | 2.543 | 2.403 | 2.327 | 2.246 | 2.160 | 2.114 | 2.077 | 2.071 |
| 15 | 1 | 3 | 4 | 6 | 3 | 6 | 7 | 5 | 5 | 7 | 1 | 1 | 6 | 8 |
| 20 | 4.351 | 3.492 | 3.098 | 2.866 | 2.710 | 2.514 | 2.347 | 2.203 | 2.124 | 2.039 | 1.946 | 1.896 | 1.856 | 1.849 |
| | 2 | 8 | 3 | 0 | 9 | 0 | 9 | 2 | 1 | 1 | 3 | 2 | 3 | 8 |
| 30 | 4.170 | 3.315 | 2.922 | 2.689 | 2.533 | 2.334 | 2.164 | 2.014 | 1.931 | 1.840 | 1.739 | 1.683 | 1.637 | 1.630 |
| | 9 | 9 | 3 | 6 | 6 | 3 | 6 | 9 | 7 | 8 | 6 | 5 | 6 | 0 |
| 60 | 4.001 | 3.150 | 2.758 | 2.525 | 2.368 | 2.166 | 1.992 | 1.836 | 1.748 | 1.649 | 1.534 | 1.467 | 1.409 | 1.399 |
| | 2 | 5 | 1 | 2 | 3 | 6 | 7 | 5 | 0 | 2 | 3 | 2 | 3 | 4 |
| 120 | 3.920 | 3.071 | 2.680 | 2.447 | 2.289 | 2.086 | 1.910 | 1.750 | 1.658 | 1.554 | 1.428 | 1.351 | 1.280 | 1.267 |
| | 1 | 8 | 2 | 3 | 8 | 8 | 4 | 5 | 7 | 4 | 9 | 9 | 4 | 4 |
| 500 | 3.860 | 3.013 | 2.622 | 2.389 | 2.232 | 2.027 | 1.849 | 1.686 | 1.591 | 1.482 | 1.345 | 1.255 | 1.158 | 1.137 |
| | 1 | 7 | 7 | 8 | 0 | 8 | 6 | 4 | 7 | 0 | 5 | 2 | 6 | 8 |
| 100 | 3.850 | 3.004 | 2.613 | 2.380 | 2.223 | 2.018 | 1.840 | 1.676 | 1.581 | 1.470 | 1.331 | 1.238 | 1.134 | 1.109 |
| 0 | 8 | 7 | 7 | 8 | 0 | 7 | 2 | 5 | 1 | 5 | 8 | 5 | 2 | 6 |

$$F = \frac{\frac{2}{\Sigma ti}}{ni} - \frac{\frac{2}{T}}{N} / K \Box 1$$

$$\frac{\frac{2}{\Sigma ti}}{\Sigma \Sigma X} \Box - \frac{2}{\Sigma ti} / N \Box K$$

$$\frac{1}{ni}$$

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

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 | <u>Group C</u> |
|---------|---------|----------------|
| 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

