ONE-WAY ANALYSIS OF VARIANCE

This methodology provides an extension of the comparisons of the means of two populations.

The name (ANALYSIS OF VARIANCE) is misleading — it's purpose is to provide a comparison of a number of means.
This analysis (usually abbreviated to ANOVA) was developed by R.A. FISHER, and is usually carried out on data that have arisen from a formal data-collection design (termed an EXPERIMENTAL DESIGN).
The simplest type of experimental design is the COMPLETELY RANDOMIZED design where the experimental units are fairly uniform and there are k treatments whose effects on the experimental units are to be compared.
One should randomly allocate the k treatments to the n experimental units, and suppose that there are n_i experimental units that receive treatment i, $i = 1, 2, \ldots k$.

THE DATA STRUCTURE (AND NOTATION)

TREATMENTS	OBSERVATIONS	TOTALS	AVERAGES
1	$y_{11}\ y_{12}\ \cdots\ y_{1n_1}$	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	$y_{21}\ y_{22}\ \cdots\ y_{2n_2}$	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
3	$y_{31}\ y_{32}\ \cdots\ y_{3n_3}$	$y_{3\cdot}$	$\bar{y}_{3\cdot}$
$\vdots$	$- - - - - - -$	$- -$	$- -$
k	$y_{k1}\ y_{k2}\ \cdots\ y_{kn_k}$	$y_{k\cdot}$	$\bar{y}_{k\cdot}$

$\bar{y}_{\cdot\cdot}$ OVERALL AVE.

THE MODEL

$$y_{ij} = \mu + \tau_i + e_{ij} \qquad \begin{array}{l} i = 1, 2, \ldots, k \\ j = 1, 2, \ldots, n_i \end{array}$$

μ is the 'overall' population mean
τ_i is the deviation of the i^{th} treatment mean from μ
— usually called the i^{th} treatment effect.
e_{ij} is the RANDOM ERROR TERM. We often assume $e_{ij} \sim N(0, \sigma^2)$

We also assume that all the e_{ij} are independent of each other. And we take $\sum_{i=1}^{k} n_i \tau_i = 0$.

Estimation of the Model Parameters

It is possible to derive estimators for the parameters in the one-way analysis of variance model

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

An appropriate estimation criterion is to estimate μ and τ_i such that the sum of the squares of the errors or deviations e_{ij} is a minimum. This method of parameter estimation is called the method of *least squares*. In estimating μ and τ_i by least squares, the normality assumption on the errors e_{ij} is not needed. To find the least squares estimators of μ and τ_i, we form the sum of squares of the errors

$$L = \sum_{i=1}^{k} \sum_{j=1}^{n_i} e_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \mu - \tau_i)^2 \tag{12-12}$$

and find values of μ and τ_i, say $\hat{\mu}$ and $\hat{\tau}_i$, that minimize L. The values $\hat{\mu}$ and $\hat{\tau}_i$ are the solutions to the $k + 1$ simultaneous equations

$$\frac{\partial L}{\partial \mu} = 0$$

$$\frac{\partial L}{\partial \tau_i} = 0 \qquad i = 1, 2, \ldots, k$$

Differentiating equation 12-12 with respect to μ and τ_i and equating to zero, we obtain

$$-2 \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu} - \hat{\tau}_i) = 0$$

and

$$-2 \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu} - \hat{\tau}_i) = 0 \qquad i = 1, 2, \ldots, k$$

After simplification these equations become

$$\begin{array}{llllll} n\hat{\mu} + n_1\hat{\tau}_1 + n_2\hat{\tau}_2 + & \cdots & + n_k\hat{\tau}_k & = y_{..} \\ n_1\hat{\mu} + n_1\hat{\tau}_1 & & & = y_{1.} \\ n_2\hat{\mu} & + n_2\hat{\tau}_2 & & = y_{2.} \\ \vdots & \vdots & & \vdots \\ n_k\hat{\mu} & & + n_k\hat{\tau}_k & = y_{k.} \end{array} \tag{12-13}$$

Equations 12-13 are called the *least squares normal equations*. Notice that if we add the last k normal equations we obtain the first normal equation. Therefore, the normal equations are not linearly independent, and there are no unique estimates for μ and $\tau_1, \ldots, \tau_k$. To deal with this, it is conventional to impose

$$\sum_{i=1}^{k} n_i \hat{\tau}_i = 0$$

This leads to $\hat{\mu} = \bar{y}_{..}$ $\left(\bar{y}_{..} = \dfrac{\sum\sum y_{ij}}{n} \right.$

and $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$ $\left. \equiv \dfrac{\sum_i y_{i.}}{\sum_i n_i} \right)$

ESTIMATION OF THE 'COMMON σ^2'

For the observations receiving the i^{th} treatment, we denote

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$$

and $E(S_i^2) = \sigma^2$; and $\dfrac{(n_i - 1) S_i^2}{\sigma^2}$ is $\sim \chi^2_{n_i - 1}$

We have an S_i^2 for each treatment, each providing information about σ^2 — and we can combine all this information:

$$E\left[\sum_{i=1}^{k} (n_i - 1) S_i^2 \right] = \sum_{i=1}^{k} (n_i - 1) \sigma^2$$

$$= (n-k)\sigma^2$$

So that we could estimate σ^2 using $\dfrac{1}{(n-k)} \sum_{i=1}^{k} (n_i - 1) S_i^2$

Also, since $\dfrac{(n_i - 1) S_i^2}{\sigma^2}$ is $\chi^2_{n_i - 1}$

then $\sum_{i=1}^{k} \dfrac{(n_i - 1) S_i^2}{\sigma^2}$ is $\chi^2_{\sum_i (n_i - 1)} \longrightarrow \chi^2_{n-k}$

$$\frac{1}{\sigma^2} \underbrace{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2} \text{ is } \sim \chi^2_{n-k}$$

THIS IS TERMED THE <u>ERROR SUM OF SQUARES</u> (OR <u>WITHIN-SAMPLE SUM OF SQUARES</u>, OR <u>RESIDUAL SUM OF SQUARES</u>)

We will denote it by SSE

Before we consider other SUMS OF SQUARES, we mention the following NULL HYPOTHESIS:

$$H_0 : \tau_i = 0 \quad \text{for all } i = 1, 2, \ldots, k$$

I.E. NO DIFFERENCE IN TREATMENT EFFECTS

and $H_A : \tau_i \neq 0$ for at least one i.

PARTITIONING THE TOTAL SUM OF SQUARES

A measure of the total variation in the data-set $\{y_{ij}\}$ is provided by

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$$

TOTAL SUM OF SQUARES

This can be split:

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{y}_{i.} - \bar{y}_{..})^2$$

[TO SEE THIS: Consider $(y_{ij} - \bar{y}_{..}) = (y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..})$

Square each side, then apply $\sum\sum$
— the cross-product term is zero.]

Thus

$$SST = SSE + SSB$$

('BETWEEN GROUPS' SUM OF SQUARES
OR TREATMENT SUM OF SQUARES)

DEGREES OF FREEDOM

Associated with each sum of squares, there is a number of degrees of freedom. We take this to be as follows:

THE NUMBER OF DEGREES OF FREEDOM FOR A SUM OF SQUARES IS EQUAL TO: (THE NUMBER OF TERMS THAT ARE BEING SQUARED) MINUS (THE NUMBER OF LINEAR CONSTRAINTS THAT APPLY TO THEM)

e.g. for SST there are n terms being squared, but there is only one constraint: $\sum\sum(y_{ij} - \bar{y}_{..})$ must be zero

$\Longrightarrow$ $(n-1)$ degrees of freedom for SST

For SSE, there are n terms being squared, but for each i, we must have $\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.}) = 0$, for $i = 1, 2, \ldots, k$

$\Longrightarrow$ a total of k constraints.

$\Longrightarrow$ $(n-k)$ degrees of freedom for SSE

SSB has $(k-1)$ degrees of freedom.

Independence of SSE and SSB

Remember we proved that for a sample from $N(\mu, \sigma^2)$ then $\bar{X}$ and S^2 were independent r. vars.

Thus, for each group i, we have a sample from $N(\mu+\tau_i, \sigma^2)$ and $\bar{y}_i$ and S_i^2 are independent random vars.

A consequence of this is that

$$\sum_i \sum_j (\bar{y}_i - \bar{y}_{..})^2 \quad \text{and} \quad \sum_{i=1}^{k} (n_i - 1) S_i^2 \quad \text{are independent}$$

I.E. SSB and SSE are independent.

Sampling Distribution of SSB and SSE

We already know that SSE/σ^2 is χ^2_{n-k}

But, 'Under H_0', $\frac{SST}{\sigma^2}$ is χ^2_{n-1}

THEORETICAL RESULT	IF X_1, X_2 ARE INDEPENDENT, AND IF $X_3 = X_1 + X_2$, WITH $X_1 \sim \chi^2_{n_1}$ AND $X_3 \sim \chi^2_{n_3}$, THEN X_2 IS $\chi^2_{n_3 - n_1}$

As a consequence, we conclude that, 'UNDER H_0'

$$\frac{SSB}{\sigma^2} \text{ is } \chi^2_{k-1}$$

But we already know that $\frac{SSE}{\sigma^2}$ is χ^2_{n-k} (and independent of SSB)

Thus we have the following IMPORTANT distribution result:

$\boxed{\text{UNDER } H_0:}$

$$\frac{SSB/k-1}{SSE/n-k} \text{ is distributed as } F_{k-1,\, n-k}$$

This ratio is often written as:

$$\frac{\text{Mean of Squares Between Treatments}}{\text{Mean of Squares for Error}} \text{ is distributed as } F_{k-1,\, n-k}$$

The calculations needed to arrive at this 'F ratio' are usually presented in an ANOVA table:

Source of Variation	Degrees of Freedom	Sums of Squares	Mean of Squares	F
Between Treatments	$k-1$	SSB	$SSB/_{k-1}$	$\frac{SSB/k-1}{SSE/n-k}$
Error (or RESIDUAL)	$n-k$	SSE	$SSE/_{n-k}$	—
Total	$n-1$	SST	—	—

NOTE: $\frac{SSE}{n-k}$ is usually termed MEAN SQUARED ERROR (MSE), and is of course an estimate of σ^2

EXAMPLE: An experiment was conducted to compare different methods of teaching arithmetic. 45 students were randomly allocated into five groups (of 9). Groups A, B were taught by the current method, the other 3 by three new methods. Then, each student took a standard test, with the following results:

WE WISH TO TEST
$H_0: \tau_i = 0 \ \forall i$
$H_A: \tau_i \neq 0$ FOR SOME i

Group										MEAN	$y_{i\cdot}$
Group A	17	14	24	20	24	23	16	15	14	$18.5\dot{5}$	167
" B	21	23	13	19	13	19	20	21	16	$18.3\dot{3}$	165
" C	28	30	29	24	27	30	28	28	23	$27.4\dot{4}$	247
" D	19	28	26	26	19	24	24	23	22	$23.4\dot{4}$	211
" E	21	14	13	19	15	15	10	18	20	16.11	145
											$\Sigma = 935$

The Group means ($\bar{y}_{i\cdot}$) certainly indicate differences in the effect of the teaching methods.

$$\Sigma\Sigma\, y_{ij}^2 = 20661 \Rightarrow SST = 20661 - \frac{(935)^2}{45} = 1233.7\dot{7}$$

Computational form for SSB $= \sum_i \frac{(y_{i\cdot})^2}{n_i} - \frac{(y_{\cdot\cdot})^2}{n} = \left\{ \frac{167^2}{9} + \frac{165^2}{9} + \frac{247^2}{9} + \frac{211^2}{9} + \frac{145^2}{9} - \frac{(935)^2}{45} \right.$

$$= 758.2\dot{2}$$

(SSE can be found from SST − SSB.)

ANOVA TABLE:

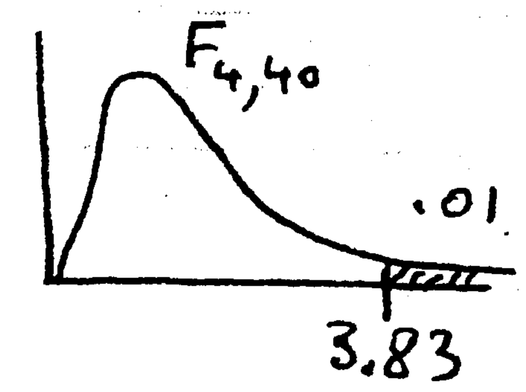
Source	d.f.	S.S.	M.S.	F
Between Treatments	4	$758.2\dot{2}$	$189.5\dot{5}$	15.944
Error	40	$475.5\dot{5}$	$11.8\dot{8}8$	
Total	44	$1233.7\dot{7}$	—	

THE 'F RATIO'

UNDER H_0, The F ratio should be distributed as $F_{4,40}$,

Using R, The p-value for 15.944 is $< 10^{-7}$

So we strongly reject H_0



CHECKING THE MODEL — USING THE RESIDUALS

The use of residuals to check on the assumptions of a model is an important technique in Statistical methods — as you know from the Regression Analysis course.

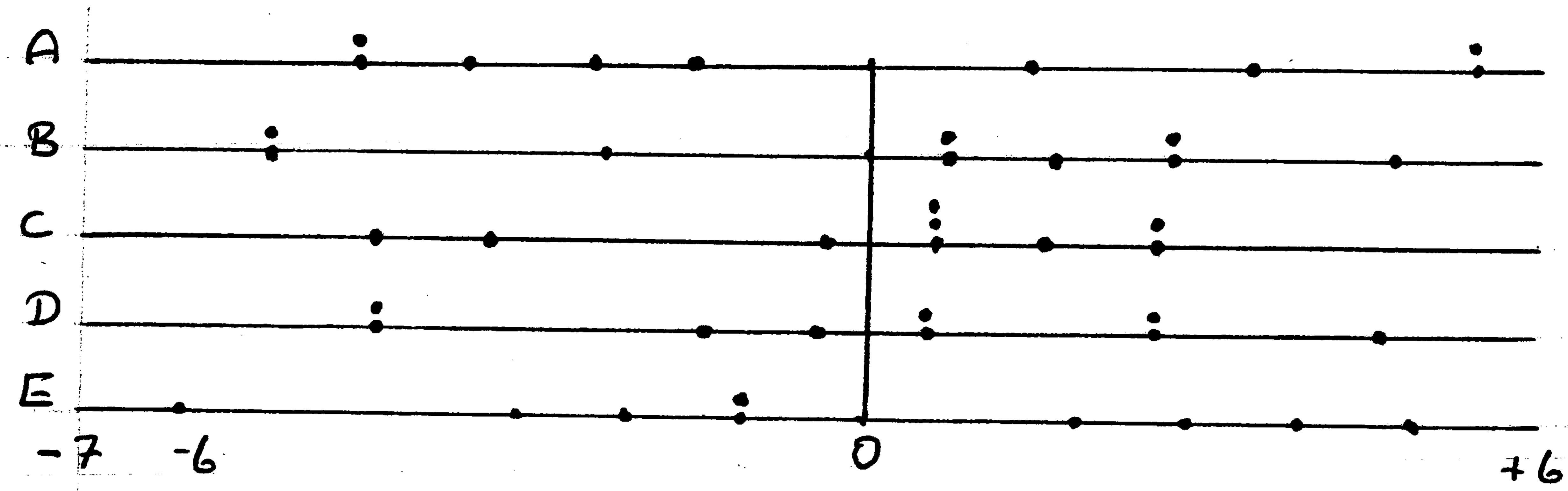
The residuals r_{ij} are the estimated 'errors':

$$r_{ij} = \hat{e}_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.}$$

Our assumption in the model specification is that

$$e_{ij} \sim N(0, \sigma^2) \qquad \text{(Same } \sigma^2 \text{ for each Treatment level.)}$$

— so we can look at some residual plots to see whether $\{\hat{e}_{ij}\}$ contradict this.



	$\bar{y}_i$
A	$18.5\dot{5}$
B	$18.3\dot{3}$
C	$27.4\dot{4}$
D	$23.4\dot{4}$
E	$16.1\dot{1}$

These plots (known as DOTPLOTS) seem to support the Common σ^2 assumption, although the NORMALITY looks doubtful.

COMPARING INDIVIDUAL MEANS

Next, we put the group means in rank order:

$$\bar{y}_{5.} < \bar{y}_{2.} < \bar{y}_{1.} < \bar{y}_{4.} < \bar{y}_{3.}$$
$$16.11 \quad 18.33 \quad 18.55 \quad 23.44 \quad 27.44$$

We can test to see if $H_0: \tau_2 = \tau_5$ (against $H_A: \tau_2 \neq \tau_5$)

Often, the LEAST SIGNIFICANT DIFFERENCE is computed — which is convenient when all group sizes are the same (as here)

This is the difference of (2) group means which is just barely (statistically) significant:

$$t\,\hat{\sigma}\left(\frac{1}{n_2} + \frac{1}{n_5}\right)^{\frac{1}{2}}$$

and t is the t-table entry for (0.975) and $(n-k)$ degrees of freedom.

For our example: $\hat{\sigma} = \sqrt{MSE} = \sqrt{11.8\dot{8}}$

and $n_2 = n_5 = 9$; $t = 2.021$ $(n-k = 40)$

So that $LSD = (2.021)\sqrt{11.88\dot{8}}\sqrt{\frac{2}{9}}$

$= 3.285$

The 'LSD' procedure is as follows: we can use this computed least significant difference (here 3.285) as a criterion for declaring differences in group means (statistically) significant.

Since $|\bar{y}_{5.} - \bar{y}_{2.}| < 3.285$, $\bar{y}_{5.}$ and $\bar{y}_{2.}$ are <u>NOT</u> significantly different.

Similarly, $\bar{y}_{2.}$ and $\bar{y}_{1.}$ are not significantly different but $\bar{y}_{1.}$ and $\bar{y}_{4.}$ <u>are</u> SIGNIFICANTLY different. (and also $\bar{y}_{4.}$ and $\bar{y}_{3.}$)

This conclusion is often presented as follows:

$$\underline{\bar{y}_{5.} < \bar{y}_{2.}} \underline{\quad < \bar{y}_{1.}} < \bar{y}_{4.} < \bar{y}_{3.}$$

Note that (here) $\bar{y}_{5.}$ and $\bar{y}_{1.}$ are <u>not</u> significantly different, and this would be conveyed as follows:

$$\underline{\bar{y}_{5.} < \bar{y}_{2.} < \bar{y}_{1.}} < \bar{y}_{4.} < \bar{y}_{3.}$$

Thus the performances of teaching methods A, B, E are not significantly different, but methods C and D differ from A, B, E and from each other. (The best method is method C).

<u>CRITICISM</u> (THE MULTIPLE COMPARISONS PROBLEM.)
The LSD method has been heavily criticised, <u>because the significance level</u> <u>becomes inflated</u> when we do a series of tests (with the same <u>data</u>.)
There are a number of methods to deal with this. (Tukey's method, Duncan's method for <u>multiple comparisons</u>.)