Analysis of Variance

Math 36b

May 7, 2009

Contents

2 ANOVA: Analysis of Variance

The basic idea is to determine if different "treatments" have different effects by comparing two different measures of variability of data to determine how much of the variation of data is random and how much is due to the treatment.

2.1 Basic ANOVA

We start with the basic definitions and theorems.

$k =$ number of *treatments*. We have data divided into categories called "treatments." The word refers to application of chemicals or other methods to improve crop yield on pieces of land. But now we generalize this to other things like the effectiveness of k drugs.

$n_j =$ size of sample from the jth treatment. We get numbers

$$Y_{1j}, Y_{2j}, Y_{3j}, \cdots, Y_{n_j j}$$

which indicate how well the jth treatment worked to improve crop output, of how well the jth drug worked for the patients who took it.

$Y_{ij} =$ ith sample point from jth treatment.

$\overline{Y}_{\bullet j} =$ average of sample from jth treatment.

To make it easier to talk about, I used an example with really simple numbers to continue:

	A	B	C			
data:	1	6	9			
	3	5	8			
		1	7			
$T_{\bullet j}$	4	12	24	40	$T_{\bullet\bullet}$	totals
n_j	2	3	3	8	n	sample sizes
$\overline{Y}_{\bullet j}$	2	4	8	5	$Y_{\bullet\bullet}$	averages
S_j^2	2	7	1			sample variances
$(n_j - 1)S_j^2$	2	14	2	18	SSE	error
$(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2$	9	1	9			
$n_j(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2$	18	3	27	48	$SSTr$	treatment

Here i is the row number and j is the column number. So: $Y_{11} = 1, Y_{21} = 3, Y_{12} = 6$, etc.

$T_{\bullet j} =$ $\sum_i Y_{ij}$ is the sum of the numbers in the j-th column.

$T_{\bullet\bullet} =$ $\sum_j T_{\bullet j} = 40$ is the sum of all the numbers Y_{ij}.

$\overline{Y}_{\bullet j} =$ $T_{\bullet j}/n_j$ is the sample mean from the jth treatment.

$\overline{Y}_{\bullet\bullet} =$ 5 is the average of all 8 sample points. It is also the weighted average of the sample means $\overline{Y}_{\bullet j}$:

$$\overline{Y}_{\bullet\bullet} = \frac{\sum n_j \overline{Y}_{\bullet j}}{n} = \frac{4+12+24}{8} = \frac{40}{8} = 5$$

2.1.1 the model

The theory of ANOVA is based on the following model.

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

where μ_j is the average effect of treatment j and ϵ_{ij} are independent normal variables $\epsilon_{ij} \sim N(0, \sigma_j^2)$, or:

$$Y_{ij} \sim N(\mu_j, \sigma_j^2)$$

To test whether the treatments have any effect, we do a hypothesis test with null hypothesis being that there is no difference between the treatments:

$H_0 :$ $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$ and $\sigma_1 = \sigma_2 = \cdots = \sigma_k = \sigma$

$H_a :$ The means μ_j are different.

The model does not consider the possibility that the σ_j are different!

In the general case when the means and variances are different, the average mean and variance is defined by

$$\mu := \frac{\sum n_j \mu_j}{n}, \quad \sigma^2 := \frac{\sum n_j \sigma_j^2}{n}$$

Then:

$$\overline{Y}_{\bullet\bullet} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

This follows from the fact that

$$\overline{Y}_{\bullet j} \sim N\left(\mu_j, \frac{\sigma_j^2}{n_j}\right)$$

So,

$$\mathbb{E}(\overline{Y}_{\bullet\bullet}) = \frac{1}{n}\sum n_j \mathbb{E}(\overline{Y}_{\bullet j}) = \frac{1}{n}\sum n_j \mu_j = \mu$$

$$\mathrm{Var}(\overline{Y}_{\bullet\bullet}) = \frac{1}{n^2}\sum n_j^2 \,\mathrm{Var}(\overline{Y}_{\bullet j}) = \frac{1}{n^2}\sum n_j^2 \frac{\sigma_j^2}{n_j} = \frac{\sigma^2}{n}$$

and the theorem that any sum of independent normal variables is normal.

2.1.2 treatment sum of squares

Definition 2.1. The *treatment sum of squares* is defined to be:

$$SSTr := \sum_{j=1}^{k} n_j(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2$$

When the treatments are different, this number gets larger.

Theorem 2.2.

$$SSTr = \sum n_j(\overline{Y}_{\bullet j} - \mu)^2 - n(\overline{Y}_{\bullet\bullet} - \mu)^2$$

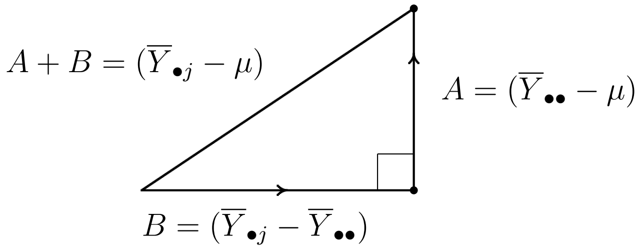
As in all theorems stating that two sums of squares add up to another sum of squares, this is the Pythagorean theorem. We have two vectors in $\mathbb{R}^n = \mathbb{R}^8$: $A = (a_{ij}) = (\overline{Y}_{\bullet\bullet} - \mu)$, the vector all of whose coordinates are equal to the same number $\overline{Y}_{\bullet\bullet} - \mu$ and $B = (b_{ij}) = (\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})$ the vector with numbers $\overline{Y}_{\bullet j} - \mu$ $(= 2, 4, 8$ in the example) repeated n_j $(= 2, 3, 3)$ times. The sum of these vectors is obviously $A + B = C = ((\overline{Y}_{\bullet j} - \mu)$:

$$\begin{array}{rcccccccc} A = & (5-\mu, & 5-\mu; & 5-\mu, & 5-\mu, & 5-\mu; & 5-\mu, & 5-\mu, & 5-\mu) \\ B = & (-3, & -3; & -1, & -1, & -1; & 3, & 3, & 3) \\ A+B = & (2-\mu, & 2-\mu; & 4-\mu, & 4-\mu, & 4-\mu; & 8-\mu, & 8-\mu, & 8-\mu) \end{array}$$

Lemma 2.3. *If the average of the coordinates of a vector V is 0 then V is perpendicular to any vector all or whose coordinates are equal:*

$$V \cdot (x, x, x, \cdots, x) = 0$$

Proof. $V \cdot (x, x, \cdots, x) = \sum x v_i = x \sum v_i = 0$. ☐



This applies to the vector $B = (\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})$ since the average of the coordinates is

$$\frac{1}{n}\sum_j \sum_i (\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet}) = \frac{1}{n}\sum_j n_j(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet}) = \frac{1}{n}\sum_j n_j\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet} = 0$$

Proof of Theorem 2.2. Since $A + B = C$ and $A \perp B = 0$, we have:

$$||A||^2 + ||B||^2 = ||C||^2$$

or:

$$\sum_{i,j}(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2 + \sum_{i,j}(\overline{Y}_{\bullet\bullet} - \mu)^2 = \sum_{i,j}(\overline{Y}_{\bullet j} - \mu)^2$$

$$SSTr + n(\overline{Y}_{\bullet\bullet} - \mu)^2 = \sum n_j(\overline{Y}_{\bullet j} - \mu)^2$$

□

Corollary 2.4. $\mathbb{E}(SSTr) = \sum \sigma_j^2 - \sigma^2 + \sum n_j(\mu_j - \mu)^2$

When $\sigma_1 = \sigma_2 = \cdots = \sigma_k = \sigma$ *this is:*

$$\mathbb{E}(SSTr) = (k-1)\sigma^2 + \sum n_j(\mu_j - \mu)^2$$

If we also assume that $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$ *then we get:*

$$\mathbb{E}(SSTr) = (k-1)\sigma^2$$

Remark 2.5. This shows that, if the means μ_j are different, $SSTr$ will be larger on average. However, if the variances σ_j^2 are different, $SSTr$ will not be affected very much. In fact, if the sample sizes are all equal then $\sum \sigma_j^2 = k\sigma^2$. So, calculating $SSTr$ will not give you any indication of any difference in variances between treatments!

Proof. According to our model, $\overline{Y}_{\bullet j} \sim N(\mu_j, \sigma_j^2/n_j)$ and $\overline{Y}_{\bullet\bullet} \sim N(\mu, \sigma^2/n)$. So,

$$\mathbb{E}\left[(\overline{Y}_{\bullet\bullet} - \mu)^2\right] = \mathrm{Var}(\overline{Y}_{\bullet\bullet}) = \sigma^2/n$$

$$\mathrm{Var}(\overline{Y}_{\bullet j} - \mu) = \mathrm{Var}(\overline{Y}_{\bullet j}) = \sigma_j^2/n_j$$

The variance can also be computed using $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$:

$$\underbrace{\mathrm{Var}(\overline{Y}_{\bullet j} - \mu)}_{\sigma_j^2/n_j} = \mathbb{E}\left[(\overline{Y}_{\bullet j} - \mu)^2\right] - \underbrace{\mathbb{E}(\overline{Y}_{\bullet j} - \mu)^2}_{(\mu_j - \mu)^2}$$

So,

$$\mathbb{E}\left[(\overline{Y}_{\bullet j} - \mu)^2\right] = \sigma_j^2/n_j + (\mu_j - \mu)^2$$

Plugging this into the formula in the theorem we just proved, this gives:

$$\mathbb{E}(SSTr) = \sum n_j \mathbb{E}\left[(\overline{Y}_{\bullet j} - \mu)^2\right] - n\mathbb{E}\left[(\overline{Y}_{\bullet\bullet} - \mu)^2\right]$$

$$= \sum n_j\left(\sigma_j^2/n_j + (\mu_j - \mu)^2\right) - n\left(\sigma^2/n\right)$$

$$= \sum \sigma_j^2 + \sum n_j(\mu_j - \mu)^2 - \sigma^2$$

□

2.1.3 more sums of squares

There are three vectors in $\mathbb{R}^n$ which are perpendicular to each other:

1. the *treatment vector* $(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})$,
2. the *error vector* $(Y_{ij} - \overline{Y}_{\bullet j})$ and
3. any vector all of whose coordinates are equal.

In the example we have:

$$\begin{array}{rrrrcrrrcrrr}
\textit{data vector} & (Y_{ij}) = & (1, & 3 & ; & 6, & 5, & 1 & ; & 9, & 8, & 7) \\
 & (\overline{Y}_{\bullet j}) = & (2, & 2 & ; & 4, & 4, & 4 & ; & 8, & 8, & 8) \\
\textit{error vector} & (Y_{ij} - \overline{Y}_{\bullet j}) = & (-1, & 1 & ; & 2, & 1, & -3 & ; & 1, & 0, & -1) \\
 & (\overline{Y}_{\bullet\bullet}) = & (5, & 5 & ; & 5, & 5, & 5 & ; & 5, & 5, & 5) \\
\textit{treatment vector} & (\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet}) = & (-3, & -3 & ; & -1, & -1, & -1 & ; & 3, & 3, & 3)
\end{array}$$

The error vector has the property that each set of n_j coordinates has sum zero. This means the error vector is perpendicular to the treatment vector and both are perpendicular to the vector all of whose coordinates are $\overline{Y}_{\bullet\bullet}$ ($= 5$ in this case).

Definition 2.6. The *error sum of squares* is given by

$$SSE := \sum_{i,j} (Y_{ij} - \overline{Y}_{\bullet j})^2 = \sum_j (n_j - 1) S_j^2$$

This is the length squared of the error vector $(Y_{ij} - \overline{Y}_{\bullet j}) \in \mathbb{R}^n$. The *total sum of squares* is given by

$$SSTot := \sum_{i,j} (Y_{ij} - \overline{Y}_{\bullet\bullet})^2$$

This is the length squared of the total vector:

$$(Y_{ij} - \overline{Y}_{\bullet\bullet}) = (Y_{ij} - \overline{Y}_{\bullet j}) + (\overline{Y}_{\bullet j} - Y_{\bullet\bullet})$$

Theorem 2.7. *$SSTot = SSE + SSTr$*

Proof. This is the Pythagorean Theorem applied to the right triangle given by the error vector and the treatment vector which are perpendicular with sum equal to the total vector. □

The main theorem is the following.

Theorem 2.8. *Suppose that* $\sigma_1 = \sigma_2 = \cdots = \sigma_k = \sigma$ *and* $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$. *Then we have the following*

$$\frac{SSTr}{\sigma^2} \sim \chi^2_{k-1} \tag{1}$$

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-k} \tag{2}$$

Furthermore, $SSTr$ *and* SSE *are independent.*

Corollary 2.9. *Under the same conditions as above, we get:*

$$\frac{SSTr/(k-1)}{SSE/(n-k)} \sim F_{k-1,n-k}$$

Proof of Theorem 2.8. We know that $\overline{Y}_{ij} \sim N(\mu_j, \sigma_j)$ with mean $\overline{Y}_{\bullet j}$. So,

$$\frac{\overline{Y}_{ij} - \mu_j}{\sigma_j} \sim N(0,1)$$

and we have

$$\sum_{i=1}^{n_j} \frac{(\overline{Y}_{ij} - \mu_j)^2}{\sigma_j^2} \sim \chi^2_{n_j}$$

When we replace μ_j with $\overline{Y}_{\bullet j}$ we lose one degree of freedom and get

$$\sum_{i=1}^{n_j} \frac{(\overline{Y}_{ij} - \overline{Y}_{\bullet j})^2}{\sigma_j^2} = \frac{(n_j-1)S_j^2}{\sigma_j^2} \sim \chi^2_{n_j-1}$$

Assume now that the σ_j are all equal to σ. Then we have the same denominators and we can add to get:

$$\frac{SSE}{\sigma^2} = \frac{\sum_{ij}(\overline{Y}_{ij} - \overline{Y}_{\bullet j})^2}{\sigma^2} = \frac{\sum_j (n_j-1)S_j^2}{\sigma^2} \sim \chi^2_{n-k}$$

For the treatments, we know that $\overline{Y}_{\bullet j} \sim N(\mu_j, \sigma_j^2/n_j)$. Therefore,

$$\frac{\overline{Y}_{\bullet j} - \mu_j}{\sigma_j/\sqrt{n_j}} \sim N(0,1)$$

So, taking squares and adding up we get:

$$\sum_{j=1}^{k} \frac{n_j(\overline{Y}_{\bullet j} - \mu_j)^2}{\sigma_j^2} \sim \chi^2_k$$

Now assume that all $\mu_j = \mu$ are all $\sigma_j = \sigma$. Then this becomes:

$$\sum_{j=1}^{k} \frac{n_j(\overline{Y}_{\bullet j} - \mu)^2}{\sigma^2} \sim \chi^2_k$$

But we also know that $\overline{Y}_{\bullet\bullet} \sim N(\mu, \sigma^2/n)$. So,

$$\frac{n(\overline{Y}_{\bullet\bullet} - \mu)^2}{\sigma^2} \sim \chi^2_1$$

Go back to Theorem 2.2 and divide by σ^2:

$$\frac{\sum n_j(\overline{Y}_{\bullet j} - \mu_j)^2}{\sigma^2} = \frac{SSTr}{\sigma^2} + \frac{n(\overline{Y}_{\bullet\bullet} - \mu)^2}{\sigma^2}$$

$$\chi^2_k = ?? + \chi^2_1$$

Since $SSTr$ and $(\overline{Y}_{\bullet\bullet} - \mu)^2$ are independent (coming from perpendicular normal distributions) we conclude that

$$\frac{SSTr}{\sigma^2} \sim \chi^2_{k-1}$$

The independence of SSE and $SSTr$ come from the fact that they are sums of squares coming from perpendicular normal distributions. (This is one of the "obvious" consequences of the theory of bivariate normal distributions which in turn is another application of the Pythagorean Theorem. I'll try to explain that later.) □

2.1.4 conclusion

We use the F-distribution to test difference in the means μ_j. It does not detect differences in variances.

Definition 2.10. The *mean squares for error* is

$$MSE := \frac{SSE}{n-k}$$

The *mean squares for treatment* is

$$MSTr := \frac{SSTr}{k-1}$$

Theorem 2.11. *The expected values of the mean squares are:*

$$\mathbb{E}(MSE) = \frac{n\sigma^2 - \sum_j \sigma_j^2}{n-k}$$

$$\mathbb{E}(MSTr) = \frac{\sum_j \sigma_j^2 - \sigma^2 + \sum n_j(\mu_j - \mu)^2}{k-1}$$

If either the sample sizes are all the same ($n_1 = n_2 = \cdots = n_k = n/k$) or if the variances are all equal then these simplify to:

$$\mathbb{E}(MSE) = \sigma^2$$

$$\mathbb{E}(MSTr) = \sigma^2 + \frac{1}{k-1}\sum n_j(\mu_j - \mu)^2$$

Proof.

$$\mathbb{E}(SSE) = \sum(n_j - 1)\mathbb{E}(S_j^2) = \sum(n_j - 1)\sigma_j^2 = \sigma^2 - \sum \sigma_j^2$$

since $\sigma^2 := \frac{1}{n}\sum n_j \sigma_j^2$ by definition. When the sample sizes n_j are all equal, σ^2 is the average of the σ_j^2. So, $\sum \sigma_j^2 = k\sigma^2$ giving the simplified equations. ☐

The point is: When the means μ_j are different the mean squares for treatment $MSTr$ gets bigger and MSE stays the same. So the fraction

$$F_{k-1,n-k} = \frac{MSTr}{MSE}$$

gets bigger. If the variances are different then neither of these numbers will change. So the F-statistic will also not change.

Conclusion: The ANOVA F-test is a right tail test which detects differences in means for the treatments but does not detect differences in variances.

For our simple example we have:

$$MSTr = \frac{SSTr}{k-1} = \frac{48}{2} = 24$$

$$MSE = \frac{SSE}{n-k} = \frac{18}{5} = 3.6$$

So, the F-statistic is

$$F_{2,5} = \frac{MSTr}{MSE} = \frac{24}{3.6} = 6.67$$

This is significant since the upper cut-off value for $F_{2,5}$ is

$$F_{2,5,.95} = 5.78615 < 6.67$$

The p-value of our F-statistic is

$$p = 0.03884 < .05$$

The conclusion is that the means μ_1, μ_2, μ_3 are not all equal. These results are usually summarized in the following chart:

Source	*SS*	*df*	*MS*	*F*	*p*
Treatment	48	2	24	6.667	0.03884
Error	18	5	3.6		
Total	66	7			

2.2 Tukey's method

If we reject the null hypothesis and find that there is a difference between treatments, we want to know which ones are different. One method is Tukey's method which uses the Studentized range distribution.

2.2.1 the model

First the model. We are doing ANOVA with k treatments. We need to assume from the beginning that the variances of the treatments are equal:

$$\sigma_1 = \sigma_2 = \cdots = \sigma_k = \sigma$$

We also need to assume that the sample sizes are equal:

$$n_1 = n_2 = \cdots = n_k = r = n/k$$

This implies that the sample mean from the jth treatment is normal:

$$\overline{Y}_{\bullet j} \sim N(\mu_j, \sigma^2/r)$$

And the sample variances all have the same distribution:

$$\frac{(r-1)S_j^2}{\sigma^2} \sim \chi^2_{r-1}$$

Definition 2.12. We say that S^2 is a χ^2 *estimator for* σ^2 *with* ν *degrees of freedom* if

$$\frac{\nu S_j^2}{\sigma^2} \sim \chi^2_\nu$$

I abbreviate this by saying that S^2 is a χ^2_ν estimator for σ^2 or S is a χ^2_ν estimator for σ.

Note that since the expected value a χ^2_ν is ν, we have:

$$\mathbb{E}(S^2) = \sigma^2.$$

In our case, S_j^2 is a χ^2_{r-1} estimator for σ^2.

When we do analysis of variance we look at

$$\frac{SSE}{\sigma^2} = \frac{\sum(r-1)S_j^2}{\sigma^2} \sim \chi^2_{kr-k}$$

If we multiply and divide by $df = kr - k$ this is:

$$\frac{(kr-k)MSE}{\sigma^2} \sim \chi^2_{kr-k}$$

In other words, MSE is a χ^2_{kr-k} estimator for σ^2.

2.2.2 Studentized range distribution

Definition 2.13. Suppose that $Z_1, Z_2, \cdots, Z_k \sim N(0,1)$ are independent standard normal variables. Suppose that $U \sim \chi^2_\nu$ is also independent from the Z_i. Then the *Studentized range* is the random variable

$$Q_{k,\nu} := \frac{\max Z_i - \min Z_i}{\sqrt{U/\nu}}$$

Theorem 2.14 (Tukey's method)**.** *Suppose that $W_1, W_2, \cdots, W_k \sim N(\mu, \sigma^2)$ are i.i.d. normal variables. Suppose that S^2 is a χ^2_ν estimator for σ^2. Then*

$$\frac{\max W_i - \min W_i}{S} \sim Q_{k,\nu}$$

The numerator $R = \max W_i - \min W_i$ is the "range" of W_i.

Remark 2.15. Note that the fraction R/S is typically given by data: W_i are sample points and S^2 is sample variance. We do not need to know the values of μ and σ.

Proof. To get the standard format we need to convert to Z's and U:

$$Z_i := \frac{W_i - \mu}{\sigma} \sim N(0,1)$$

$$U := \frac{\nu S^2}{\sigma^2} \sim \chi^2_\nu$$

We need to show that, in the formula for Q, the unknown population variables μ and σ cancel:

$$Q_{k,\nu} = \frac{\max Z_i - \min Z_i}{\sqrt{U/\nu}}$$

Since $U/\nu = S^2/\sigma^2$, the denominator is

$$\sqrt{U/\nu} = S/\sigma$$

The numerator is:

$$\max\left(\frac{W_i - \mu}{\sigma}\right) - \min\left(\frac{W_i - \mu}{\sigma}\right) = \frac{(\max W_i - \mu) - (\min W_i - \mu)}{\sigma}$$

$$= \frac{\max W_i - \min W_i}{\sigma} = R/\sigma$$

So,

$$Q_{k,\nu} = \frac{R/\sigma}{S/\sigma} = \frac{R}{S}$$

□

Corollary 2.16. $\mathbb{P}(|W_i - W_j| < S \cdot Q_{k,\nu,1-\alpha}$ *for all* $i,j) = 1 - \alpha$

Proof. Since $Q = R/S$, we have $R = SQ$. When we insert the critical value of Q we get the critical value of R: $R_{crit} = SQ_{crit}$.

The statement that every difference $W_i - W_j$ is less than R_{crit} is the same as the statement that the maximum difference R is less than R_{crit}. But, the definition of the critical value is

$$\mathbb{P}(R < R_{1-\alpha}) = 1 - \alpha.$$

$\square$

2.2.3 Tukey's method

There are two slightly different cases.
Case 1: Hypothesis testing Suppose we are testing the hypothesis: $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu$ vs H_a : the μ_j are not all equal. Recall that we are taking samples all of the same size: $n_1 = n_2 = \cdots = n_k = r$. Let

$$W_j = \overline{Y}_{\bullet j} \sim N(\mu, \sigma^2/r)$$

Let

$$S = \sqrt{MSE/r}$$

Lemma 2.17. $S^2 = MSE/r$ *is a* χ^2_{kr-k} *estimator for* σ^2/r.

Proof.

$$\frac{(kr-k)S^2}{\sigma^2/r} = \frac{(kr-k)MSE/r}{\sigma^2/r} = \frac{(kr-k)MSE}{\sigma^2} \sim \chi^2_{kr-k}$$

$\square$

Since S^2 is a χ^2 estimator for the variance of $W_j = \overline{Y}_{\bullet j}$. So, by Tukey's method we have:

$$\mathbb{P}\left(\left|\overline{Y}_{\bullet i} - \overline{Y}_{\bullet j}\right| < \sqrt{\frac{MSE}{r}} Q_{k,kr-k,1-\alpha} \text{ for all } i,j\right) = 1 - \alpha$$

Let $R_{crit} = \sqrt{\frac{MSE}{r}} Q_{k,kr-k,1-\alpha}$. Then we should reject the null hypothesis if

$$\max \overline{Y}_{\bullet j} - \min \overline{Y}_{\bullet j} \geq R_{crit}$$

Case 2: simultaneous confidence intervals for $\mu_i - \mu_j$

If we reject H_0 then we are saying the μ_j are different. But which ones? Here you take

$$W_j = \overline{Y}_{\bullet j} - \mu_j \sim N(0, \sigma^2/r)$$

Then

$$W_i - W_j = (\overline{Y}_{\bullet i} - \mu_i) - (\overline{Y}_{\bullet j} - \mu_j) = (\overline{Y}_{\bullet i} - \overline{Y}_{\bullet j}) - (\mu_i - \mu_j)$$

and Tukey's method says that, with probability $1 - \alpha$ we have

$$\left|(\overline{Y}_{\bullet i} - \overline{Y}_{\bullet j}) - (\mu_i - \mu_j)\right| < R_{crit}$$

So, the 95% confidence interval for $\mu_i - \mu_j$ is given by

$$(\overline{Y}_{\bullet i} - \overline{Y}_{\bullet j} - R_{crit}, \overline{Y}_{\bullet i} - \overline{Y}_{\bullet j} + R_{crit})$$

2.2.4 example: Case 12.3.1.

Here we have $k = 5$ drugs: P, T, S, E, C which we consider to be the treatments. We take a sample of size $r = 4$ for each drug and get:

treatment:	P	T	S	E	C
$\overline{Y}_{\bullet j}$:	28.6	31.375	7.825	19.075	27.8

Standard ANOVA gives:

$Source$	SS	df	MS	F	$F_{4,15,.95}$
$Treatment$	1480.823	4	370.20575	40.8849	3.056
$Error$	135.8225	15	9.05483		
$Total$	1616.6455	19			

Since the F-statistic is much greater than the critical value:

$$F_{4,15} = 40.8849 >> 3.056 = F_{4,15,.95}$$

So, we reject H_0 and conclude that some of the drugs "bind to serum protein" at different rates. But which ones?

Tukey's method says that the critical value for the range at 95% confidence is

$$R_{crit} = \sqrt{\frac{MSE}{r}} Q_{5,15,.95} = \sqrt{\frac{9.05483}{4}} 4.37 = 6.58$$

To (re)do the hypothesis test, we would compare this number with the range

$$R = \max \overline{Y}_{\bullet j} - \min \overline{Y}_{\bullet j} = 31.375 - 7.825 = 23.55$$

Since this is greater than 6.58 we correctly rejected the null hypothesis. Tukey's method now say that the simultaneous 95% confidence interval for $\mu_i - \mu_j$ is

$$\overline{Y}_{\bullet i} - \overline{Y}_{\bullet j} \pm 6.58$$

For example the 95% confidence interval for $\mu_1 - \mu_2$ is $\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2} = -2.77$ plus or minus $R_{crit} = 6.58$:

$$(-2.77 - 6.58, -2.77 + 6.58) = (-9.35, 3.81)$$

this is not significant since 0 is in the confidence interval. The 95% confidence interval for $\mu_1 - \mu_2$ is $\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 3} = 20.77$ plus or minus $R_{crit} = 6.58$:

$$(14.19, 27.35)$$

since 0 is not in this confidence interval we conclude that $\mu_1 \neq \mu_3$. You could conclude that $\mu_1 > \mu_3$ but not everyone would agree with that.

2.3 Comparison with t-test

If $k = 2$ (i.e., there are only two treatments) then we are back to the two sample case. The only difference is that ANOVA assumes from the beginning that the variances of the two populations are equal. (I pointed out that the ANOVA F-test does not detect differences in variance, only differences in means.

Lemma 2.18. *For $k = 2$ treatments, MSE is equal to the pooled variance.*

Proof. The formula for MSE is:

$$MSE = \frac{SSE}{n-k} = \frac{\sum(n_j - 1)S_j^2}{n-k}$$

Since $n = \sum n_j$, when $k = 2$ this is

$$\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

which is the pooled variance. □

But how are the t-test and the F-test related?

Lemma 2.19. *If T has a Student t-distribution with m degrees of freedom then T^2 has an $F_{1,m}$-distribution.*

Proof. By definition of the t-distribution we have:

$$T = \frac{Z}{\sqrt{U/m}}$$

where $Z \sim N(0,1)$ and $U \sim \chi^2_m$. If we square this we get:

$$T^2 = \frac{Z^2}{U/m} = \frac{Z^2/1}{U/m} \sim F_{1,m}$$

which has an F-distribution since $Z^2 \sim \chi^2_1$. □

But what about $MSTr = SSTr/(k-1)$? ($= SSTr$ when $k = 2$)

Lemma 2.20. *When $k = 2$ we have:*

$$SSTr = MSTr = n_1(\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet\bullet})^2 + n_2(\overline{Y}_{\bullet 2} - \overline{Y}_{\bullet\bullet})^2$$

$$= \frac{n_1 n_2(\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2})^2}{n_1 + n_2}$$

Putting these together we get the theorem:

Theorem 2.21. *For $k = 2$ treatments, the two-sided t-test is equivalent to the ANOVA F-test.*

Proof. In the t test we use the statistic:

$$T = \frac{\overline{X} - \overline{Y}}{S_P\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

If we square this we get:

$$T^2 = \frac{(\overline{X} - \overline{Y})^2}{S_P^2(n+m)/nm} \sim F_{1,n+m-2}$$

This is the same as the F-statistic from ANOVA:

$$\frac{MSTr}{MSE} \sim F_{1,n+m-2}$$

since

$$MSTr = SSTr = \frac{nm(\overline{X} - \overline{Y})^2}{n+m}$$

by the previous lemma and $MSE = S_P^2$ by the first lemma. □

For example, let's take problem 9.3.1 comparing short presidents and tall presidents:

n_j	5	26
$\overline{Y}_{\bullet j}$	80.2	69.154
s_j^2	73.7	86.775

ANOVA gives:

Source	*SS*	*df*	*MS*	*F*
Treatment	511.6863524	1	511.6863524	6.02183137
Error	2464.184615	29	84.97188329	
Total	2975.870968	30		

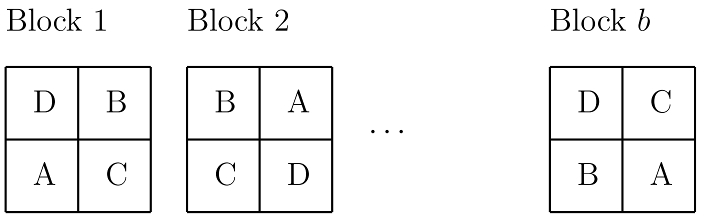
Here you can see that $MSE = 84.97$ is the pooled variance and the two-sided t-test gives

$$\frac{\overline{X} - \overline{Y}}{S_P\sqrt{\frac{1}{5} + \frac{1}{26}}} = 2.453942006 = \sqrt{6.02183137}$$

So, as the theorem says, ANOVA gives the same same result as the 2-sided t-test.

2.4 Randomized block design

Traditionally, treatments are things you do to soil to improve the yield of your crops. The idea is that the soil itself may have a bias. Some plots of soil may be more fertile than others. A *block* is a piece of land in one place which is randomly divided into k plots. We have k treatments which we apply to these plots.



In the figure, there are only $k = 4$ treatments and b blocks.

2.4.1 theory

The assumption is that there is no "interaction" between treatments and blocks. The formula that we are assuming is the following. If Y_{ij} is the "yield" of the ith block and the jth treatment then we assume:

$$\boxed{Y_{ij} = \beta_i + \mu_j + \epsilon_{ij}}$$

Here β_i is the average effect of the i-block. It reflects the fertility of this piece of land. μ_j is the average effect of the jth treatment. It reflects how much more this type of fertilizer or whatever it is will improve the yield. Finally,

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

is the random error which we assume is normally distributed. In other words,

$$Y_{ij} \sim N(\beta_i + \mu_j, \sigma^2).$$

The equation in the box tells us that we are assuming no interaction between treatments and blocks. There is no $c\beta_i\mu_j$ term in this sum. (A terms like this would mean that certain blocks would "magnify" the effects of the treatments.)

We want to know the effects of the treatments. So, we average over the blocks to eliminate the block effects.

$$\begin{array}{r|c|c|}
 & \text{Treat 1} & \text{Treat 2} \\ \cline{2-3}
\text{Block 1} & Y_{11} & Y_{22} \\ \cline{2-3}
\text{Block 2} & Y_{21} & Y_{22} \\ \cline{2-3}
\text{treatment average} & \overline{Y}_{\bullet 1} & \overline{Y}_{\bullet 2} \\
\text{treatment effect} & \overline{Y}_{\bullet 1} - \overline{Y}_{\bullet\bullet} & \overline{Y}_{\bullet 2} - \overline{Y}_{\bullet\bullet}
\end{array}
=
\begin{array}{|c|c|}
\text{Treat 1} & \text{Treat 2} \\ \hline
\beta_1 + \mu_1 + \epsilon & \beta_2 + \mu_2 + \epsilon \\ \hline
\beta_2 + \mu_1 + \epsilon & \beta_2 + \mu_2 + \epsilon \\ \hline
\overline{\beta} + \mu_1 + \epsilon & \overline{\beta} + \mu_2 + \epsilon \\
\mu_1 - \overline{\mu} + \epsilon & \mu_2 - \overline{\mu} + \epsilon
\end{array}$$

2.4.2 block and treatment averages

I took a really simple example with $b = k = 2$ to illustrate the way that we eliminate the block effects from the data.

In this example, I pointed out that the average of the numbers in the first column is

$$\overline{Y}_{\bullet 1} = \frac{1}{b} \sum Y_{i1} = \overline{\beta} + \mu_1 + \epsilon$$

where the error term is actually the average of the error terms in the first column. We still have the average block effect given by

$$\overline{\beta} = \frac{1}{b} \sum_{i=1}^{b} \beta_i$$

For the second column we get

$$\overline{Y}_{\bullet 2} = \frac{1}{b} \sum Y_{i2} = \overline{\beta} + \mu_2 + \epsilon$$

To get rid of the block effect which we don't care about, we subtract the average of these:

$$\overline{Y}_{\bullet\bullet} = \frac{1}{k} \sum \overline{Y}_{\bullet j} = \frac{1}{b} \sum \overline{Y}_{i\bullet}$$

This average is:

$$\overline{Y}_{\bullet\bullet} = \overline{\beta} + \overline{\mu} + \overline{\epsilon}$$

where $\overline{\epsilon}$ is the average of all of the error terms ϵ_{ij}.

The difference is the (net) treatment effect:

$$\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet\bullet} = \mu_1 - \overline{\mu} + \epsilon$$

What we want to do is to add up the squares of these (net) treatment effects. And I drop the word "net."

2.4.3 sums of squares

I now used a bigger example with $b = k = 3$. Here, the numbers are made up so that we can see the block, treatment and error effects.

Data:	treat 1	treat 2	treat 2	mean	block effect
block 1	1.17	1.53	1.65	1.45	-1
block 2	2.15	2.57	2.63	2.45	0
block 3	3.13	3.55	3.67	3.45	1
mean	2.15	2.55	2.65	2.45	
treatment effect	-0.3	0.1	0.2	$\overline{Y}_{\bullet\bullet}$	

In this example, the first digit of the number Y_{ij} comes from the blocks, the second digit (the first decimal) comes from the treatment and the last digit (second decimal) represents error.

From this we construct three perpendicular vectors in $\mathbb{R}^{bk} = \mathbb{R}^9$ by repeating the block and treatment effects over the data points that they come from:

1) We have the treatment vector with coordinates $\overline{Y}_{\cdot j} - \overline{Y}_{\bullet\bullet}$ repeated $b = 3$ times:

$$\begin{array}{ccc} -0.3 & 0.1 & 0.2 \\ -0.3 & 0.1 & 0.2 \\ -0.3 & 0.1 & 0.2 \end{array} \quad \in \mathbb{R}^9$$

The numbers in every row add up to zero and the numbers in the columns are equal.

2) We have the block vector with coordinates $\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}$ repeated over each row:

$$\begin{array}{ccc} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{array} \quad \in \mathbb{R}^9$$

The numbers in every column add up to zero and the numbers in the rows are equal.

3) The error vector has coordinates:

$$\underbrace{(Y_{ij} - \overline{Y}_{\bullet\bullet})}_{total} - \underbrace{(\overline{Y}_{\cdot j} - \overline{Y}_{\bullet\bullet})}_{treatment} - \underbrace{(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})}_{block}$$

$$= Y_{ij} + \overline{Y}_{\bullet\bullet} - \overline{Y}_{\cdot j} - \overline{Y}_{i\bullet}$$

$$\begin{array}{ccc} 0.02 & -0.02 & 0 \\ 0 & 0.02 & -0.02 \\ -0.02 & 0 & 0.02 \end{array}$$

The numbers in every row and every column add up to zero.

These properties imply that the dot products of any two of these three vectors is zero. In other words, they are perpendicular, just like the x, y, z axes. For example, any array of numbers in which the rows add to zero will be perpendicular to the block vector. And any vector where the numbers in the columns add to zero will be perpendicular to the treatment vector. This implies the following theorem by Pythagoras:

Theorem 2.22.

$$SSTot = SSTr + SSB + SSE$$

where the terms are defined below.

Definition 2.23.

1. The *treatment sum of squares* is defined to be:

$$SSTr := \sum_{i=1}^{b}\sum_{j=1}^{k}(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2 = \sum_{j=1}^{k} b(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2$$

   This is the length squared of the treatment vector.

2. The *block sum of squares* is defined to be:

$$SSB := \sum_{i=1}^{b}\sum_{j=1}^{k}(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{b} k(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2$$

   This is the length squared of the block vector.

3. The *error sum of squares* is

$$SSE := \sum_{i=1}^{b}\sum_{j=1}^{k}(Y_{ij} + \overline{Y}_{\bullet\bullet} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j})^2$$

4. The *total sum of squares* is

$$SSTot := \sum_{i=1}^{b}\sum_{j=1}^{k}(Y_{ij} + \overline{Y}_{\bullet\bullet})^2$$

2.4.4 χ^2 estimators

The next step is to produce χ^2 estimators and take their ratios to get F-distributions. For this, again, I took small numbers $b = 2, k = 3$ to illustrate. Here I looked at the $k = 3$ numbers

$$Y_{1j} = \beta_1 + \mu_j + \epsilon_j, \quad \epsilon_j \sim N(0, \sigma^2)$$

Their average is

$$\overline{Y}_{1\bullet} = \beta_1 + \overline{\mu} + \overline{\epsilon} \sim N\left(\beta_1 + \overline{\mu}, \frac{\sigma^2}{k}\right)$$

The reason is that

$$\overline{\epsilon} = \frac{\epsilon_1 + \epsilon_2 + \epsilon_3}{3}$$

which has variance

$$\operatorname{Var}(\overline{\epsilon}) = \frac{1}{9}(\operatorname{Var}(\epsilon_1) + \operatorname{Var}(\epsilon_2) + \operatorname{Var}(\epsilon_3)) = \frac{3\sigma^2}{9} = \frac{\sigma^2}{3}$$

If the expected blocks effects β_i are all equal then these numbers are i.i.d.

$$\overline{Y}_{i\bullet} \sim N\left(\overline{\beta} + \overline{\mu}, \frac{\sigma^2}{k}\right)$$

and the sum of squares of their differences from their mean is a χ^2 estimator for their variances:

$$\frac{\sum_{i=1}^{b}(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2}{\sigma^2/k} \sim \chi^2_{b-1}$$

Multiplying top and bottom by k, we get SSB in the numerator:

$$\frac{\sum_{i=1}^{b} k(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2}{\sigma^2} = \frac{SSB}{\sigma^2} \sim \chi^2_{b-1}$$

In other words:

Theorem 2.24. *If the expected block effects β_i are all equal then MSB is a χ^2 estimator for σ^2 with $b-1$ degrees of freedom.*

If we take the average for the jth treatment we get

$$\overline{Y}_{\bullet j} \sim N\left(\overline{\beta} + \mu_j, \frac{\sigma^2}{b}\right)$$

If we assume that

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \overline{\mu}$$

then the $\overline{Y}_{\bullet j}$ are i.i.d. and we get:

$$\frac{\sum_{j=1}^{k}(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2}{\sigma^2/b} \sim \chi^2_{b-1}$$

But the numerator is $SSTr/b$. So this becomes:

Theorem 2.25. *Assuming H_0 (the μ_j are all equal) we have:*

$$\frac{SSTr}{\sigma^2} \sim \chi^2_{b-1}$$

2.4.5 error sum of squares

The error sum of squares is special.

$$SSE := \sum\sum(\underbrace{Y_{ij} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j} + \overline{Y}_{\bullet\bullet}}_{E_{ij}})^2$$

The error vector (E_{ij}) has coordinates:

$$E_{ij} = Y_{ij} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j} + \overline{Y}_{\bullet\bullet}$$

Let's look at each term:

$$\begin{aligned} Y_{ij} &= \beta_i + \mu_j + \epsilon_{ij} \\ \overline{Y}_{i\bullet} &= \beta_i + \overline{\mu} + \overline{\epsilon}_{i\bullet} \\ \overline{Y}_{\bullet j} &= \overline{\beta} + \mu_j + \overline{\epsilon}_{\bullet j} \\ \overline{Y}_{\bullet\bullet} &= \overline{\beta} + \overline{\mu} + \overline{\epsilon}_{\bullet\bullet} \end{aligned}$$

So, all the terms cancel and we get just a sum of error terms:

$$E_{ij} = \epsilon_{ij} - \overline{\epsilon}_{i\bullet} - \overline{\epsilon}_{\bullet j} + \overline{\epsilon}_{\bullet\bullet}$$

Since ϵ_{ij} are i.i.d. $N(0, \sigma^2)$, $\epsilon_{ij}/\sigma \sim N(0, 1)$. So,

$$\frac{1}{\sigma^2}\sum\sum \epsilon_{ij}^2 \sim \chi^2_{bk}$$

Since $\overline{\epsilon}_{i\bullet}$ are independent $N(0, \sigma^2/k)$,

$$\frac{1}{\sigma^2/k}\sum \overline{\epsilon}_{i\bullet}^2 \sim \chi^2_b$$

Similarly,

$$\frac{1}{\sigma^2/b}\sum \overline{\epsilon}_{\bullet j}^2 \sim \chi^2_k$$

and

$$\frac{bk}{\sigma^2}\,\overline{\epsilon}_{\bullet\bullet}^2 \sim \chi^2_1$$

Putting these together using the Pythagorean Theorem we get:

$$\frac{1}{\sigma^2}\sum\sum E_{ij}^2 \sim \chi^2_{bk-k-b+1}$$

In other words, even under the alternate hypothesis (when the μ_j are different), we have:

$$\frac{SSE}{\sigma^2} \sim \chi^2_{(b-1)(k-1)}$$

So, $MSE = SSE/(b-1)(k-1)$ is a χ^2 estimator for σ^2 with $(b-1)(k-1)$ degrees of freedom.

2.4.6 F-distribution

I don't have time for a long explanation. Under the null hypothesis that the treatments are all equal,

$$\frac{MSTr}{MSE} \sim F_{k-1,(b-1)(k-1)}$$

2.4.7 Tukey's method

Again, I don't have time for a long explanation. The Tukey range is given by:

$$R_{crit} = S \cdot Q_{crit}$$

where

$$S = \sqrt{\frac{MSE}{b}}$$

and Q_{crit} is:

$$Q_{k,(b-1)(k-1),.95}$$