MIE237

Neil Montgomery
2016-04-08

the analysis of designed experiments

Formal definitions: factor, level

A is a controllable experimental condition.

A factor can take on two or more .

E.g., in a study of haul trucks "oil brand" could be a factor, with levels "Castrol", "Volvo", "Komatsu".

When experimental units are randomly assigned to levels of a factor and some output measure is observed, this is called a . The formal model is typically written as (more on this later):

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

We've seen the case of $i \in \{1, 2\}$—such an experiment would be analyzed using a two-sample t procedure.

In reality any dataset with one categorical "input" variable and one numerical "output" variable will be analysed the same as a formally designed experiment.

Typical dataset…

Truck.ID	Oil	Viscosity
HT 265	Volvo	25.5
HT 372	Castrol	25.7
HT 572	Komatsu	25.6
HT 908	Volvo	24.7
HT 201	Castrol	26.5
HT 898	Komatsu	25.4
HT 944	Volvo	24.4
HT 660	Castrol	22.8
HT 629	Komatsu	26.1
HT 61	Volvo	25.0
HT 205	Castrol	25.0
HT 176	Komatsu	25.9

One factor notation, models

"Balanced" case with equal sample size n for each of k levels for $N = nk$ total.

Levels:	**1**	**2**	**...**	**i**	**...**	**k**
	y_{11}	y_{21}	$\cdots$	y_{i1}	$\cdots$	y_{k1}
	y_{12}	y_{22}	$\cdots$	y_{i2}	$\cdots$	y_{k2}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	y_{1n}	y_{2n}	$\cdots$	y_{in}	$\cdots$	y_{kn}
Sample average:	$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$	$\cdots$	$\bar{y}_{i\cdot}$	$\cdots$	$\bar{y}_{k\cdot}$

Grand overall average: $\bar{y}_{\cdot\cdot}$

Models:

$$y_{ij} = \mu_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \text{ i.i.d. } N(0, \sigma^2)$$

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad \sum \alpha_i = 0 \qquad \varepsilon_{ij} \text{ i.i.d. } N(0, \sigma^2)$$

The main question

The main question is $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = 0$ versus the negation (equivalently: all the $\alpha_i = 0$.)

In other words "is the variation among all the y_{ij} due to the factor variable, or just due to random chance?". The analysis even follows this logic.

The variation among the y_{ij} is quantified as (as usual?):

$$(N-1) \cdot s_y^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} \left(y_{ij} - \overline{y}_{..}\right)^2$$

We will split this up into the "factor" part and the "random chance" part (like done in regression).

"Analysis of Variance" - I

Build up from the inside out. For any i and j fixed:

$$\begin{aligned}
\left(y_{ij} - \bar{y}_{..}\right)^2 &= \left(y_{ij} - \bar{y}_{i.} + \bar{y}_{i.} - \bar{y}_{..}\right)^2 \\
&= \left(y_{ij} - \bar{y}_{i.}\right)^2 + \left(\bar{y}_{i.} - \bar{y}_{..}\right)^2 + 2\left(y_{ij} - \bar{y}_{i.}\right)\left(\bar{y}_{i.} - \bar{y}_{..}\right)
\end{aligned}$$

Next, sum from $j = 1$ to n to get:

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

Finally, sum from $i = 1$ to k and rearrange:

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

(Note: unbalanced case)

Simply replace n with n_i (sample size for level i)

"Analysis of Variance" - II

Similar to the regression case:

$$\begin{array}{cccccc}
\sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{..}\right)^2 & = & \sum_{i=1}^{k} n\left(\bar{y}_{i.}-\bar{y}_{..}\right)^2 & + & + \sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i.}\right)^2 \\
SST & = & SSA & + & SSE \\
\chi^2_{nk-1} & = & \chi^2_{k-1} & + & \chi^2_{k(n-1)}
\end{array}$$

Note that $nk-1=k-1+k(n-1)$, as expected.

Call:

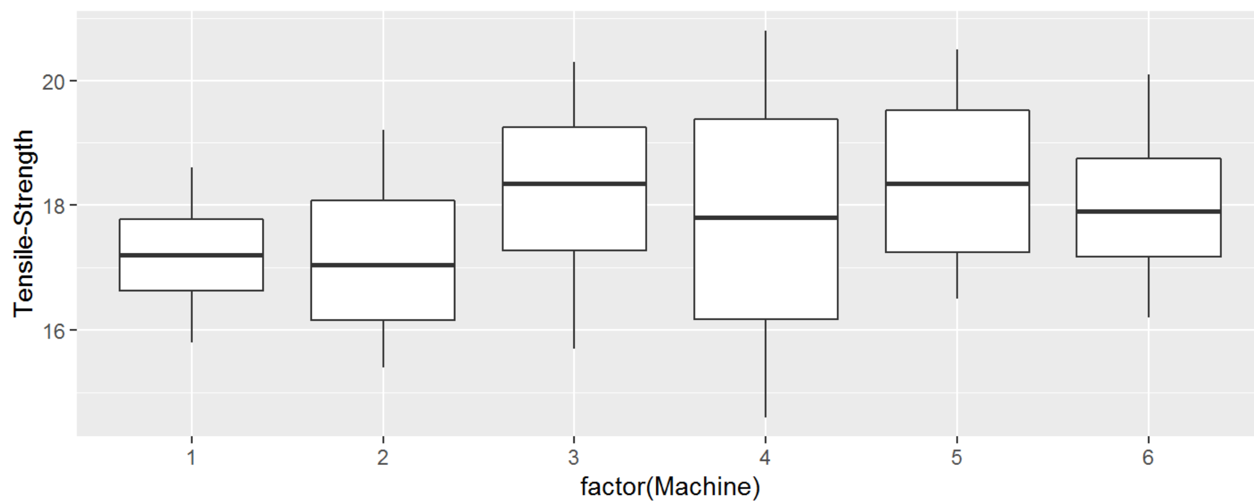
$$MSA=\frac{SSA}{k-1} \qquad \text{and} \qquad MSE=\frac{SSE}{k(n-1)}$$

"Analysis of Variance" - III

The analysis is based on:

$$F = \frac{MSA}{MSE} \sim F_{k-1,k(n-1)}$$

For example, consider exercise 13.1 "tensile strength of rubber seals". $k = 6$ machines are compared and $n = 4$ seals taken from each machine.



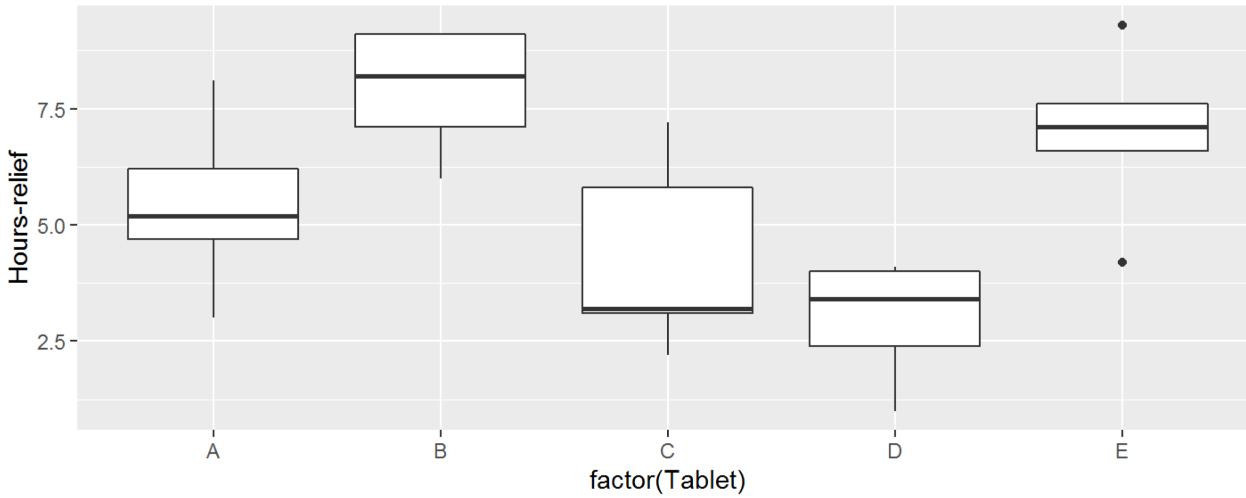
ex 13.1 analysis

```
summary(aov(`Tensile-Strength` ~ factor(Machine), seals))
```

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## factor(Machine)  5   5.34   1.068   0.307  0.902
## Residuals       18  62.64   3.480
```

Exercise 13.2

"Hours of relief for people with fevers". $k = 5$ tablets with $n = 5$ subjects each.




```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## factor(Tablet)  4  78.42  19.605   6.587 0.0015 **
## Residuals      20  59.53   2.977
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Relationship with other analyses

When $k = 2$ the usual analysis is done with a two-sample t test. The p-value obtained using the F approach will be   to the one using the t test (equal variance assumption version).

The analyis is   (mathematically) to the multiple regression with the factor variable coded into the required number of dummy variables.

Sadly the book chooses (which I have followed) to use k as the number of variables in a regression model as well as the number of levels for a factor. Of course, k levels translates into $k - 1$ dummy variables.

Model assumptions

The model assumptions are the same as for the two-sample t test: equal variances, normal errors.

The way to assess the model assumptions is similar, but adjusted for k levels rather than just 2.

For normality, use a normal quantile plot of the "residuals" $y_{ij} - \bar{y}_{i\cdot}$.

There is no good formal statistical test for equal variances. We'll use this heuristic. Calculate:

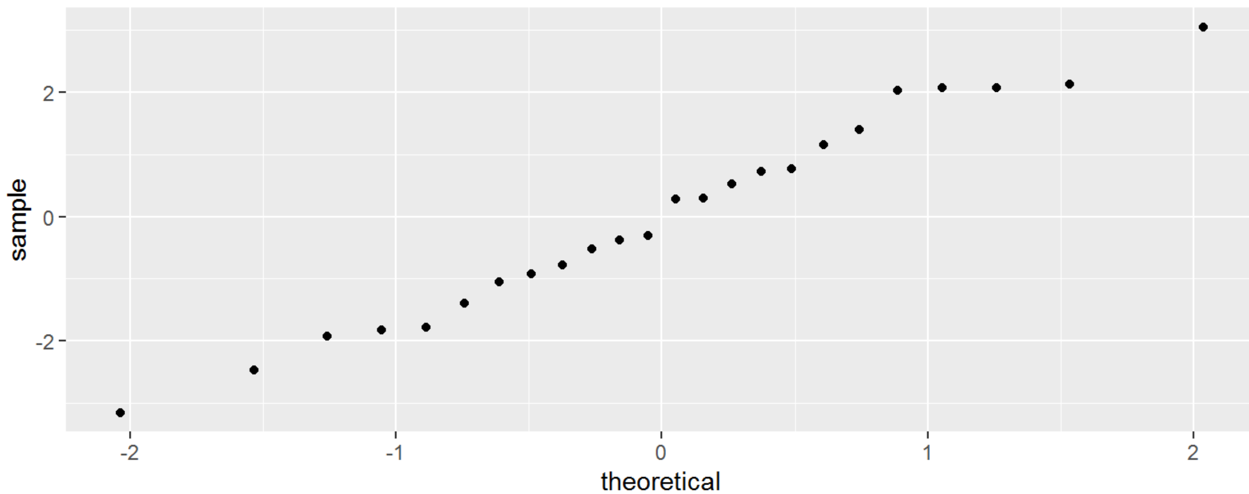
$$\frac{\max_i s_i^2}{\min_i s_i^2}$$

ANOVA equal variance assumption heuristic

Compare with:

- 9, when the residuals are normal the experiment is balanced
- 4, when the residuals are non-normal the experiment is (severely) unbalanced (but not both)
- 3, when the residuals are non-normal the experiment is (severely) unbalanced

Rubber seals assumptions check




```
## Source: local data frame [6 x 2]
##
##   Machine Variance
##     (int)    (dbl)
## 1       1 1.366667
## 2       2 2.709167
## 3       3 3.769167
## 4       4 7.216667
## 5       5 3.155833
## 6       6 2.662500
```

Multiple comparisons

The basic analysis does not give information about the most common follow-up question: which pairs of groups are different?

Naive approach: perform the $k(k-1)/2$ two-sample t tests of the pairs that interest us.

Problems: Type I error, and bias.

Solution: perform pairwise comparisons, holding them to a standard that controls for an overall "experiment-wise" error rate.

Tukey's procedure

Works for balanced experiments: $n_i = n$ for all i

Rather than comparing

$$\frac{\overline{y}_{i\cdot} - \overline{y}_{j\cdot}}{\sqrt{2 \cdot MSE/n}}$$

with a t distribution, we'll compare it with a special purpose (wider) distributioncalled the "studentized range distribution", which is the distribution of:

$$\frac{\max_i \overline{y}_{i\cdot} - \min_j \overline{y}_{j\cdot}}{\sqrt{MSE/n}}$$

Distribution parameters: k and $\nu = k(n-1)$

Tukey's procedure algorithm

This is classical hypothesis testing: fix α

Perform the overall F test. If "accept", **stop**.

Find the α upper-tail probability point $q[\alpha, k, \nu]$ from the table.

Multiply it by $\sqrt{MSE/n}$ to get the comparison threshold ("critical value")

Rank the $y_{i\cdot}$ (low to high) and compare.

Example: rubber seals

Done.

Example: tablets

$\alpha = 0.05$

$q(0.05, 5, 20) = 4.24$

Tablet	Mean
D	2.98
C	4.30
A	5.44
E	6.96
B	7.90

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## factor(Tablet)  4  78.42  19.605   6.587 0.0015 **
## Residuals      20  59.53   2.977
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Critical value: $\sqrt{\frac{2.977}{5}} = 3.2716759$