ANALYSIS OF EXPERIMENTS

Chapter 8: Analysis of Experiments

- ANOVA
- Completely Randomised Design
- Randomised Block Design
- Latin Square
- Factorial Experiments
- Interaction

Analysis of Variance

Analysis of Variance (ANOVA) can be considered as a version of a linear model (so not really anything new!) when all of the explanatory variables are categorical (the response is typically continuous).

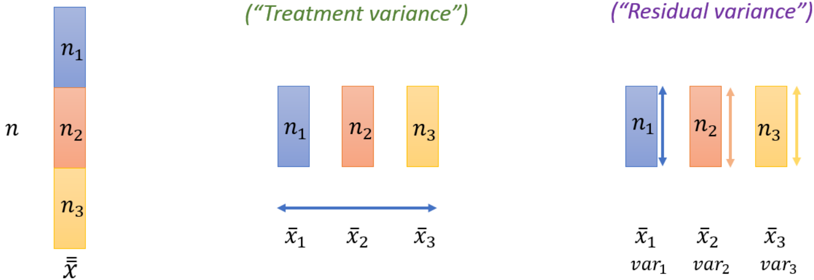
Their emphasis is a little different from linear regression models.

In linear regression models (with categorical explanatory variables) the emphasis is on testing *individual differences* between the means of the response for each level of the variable *compared to a baseline level*.

In ANOVA, the emphasis is on how much of the variation in the response variable can be attributed to variation in the explanatory variable.

ANOVA

Total variance $=$ *Variance between groups (“Treatment variance”)* $+$ *Variance within groups (“Residual variance”)*



$$\sum_{i=1}^{n_1+n_2+n_3} \frac{(x_i - \bar{\bar{x}})^2}{n_1+n_2+n_3-1} = n_1\frac{(\bar{x}_1-\bar{\bar{x}})^2}{n_1-1} + n_1\frac{(\bar{x}_2-\bar{\bar{x}})^2}{n_2-1} + n_1\frac{(\bar{x}_3-\bar{\bar{x}})^2}{n_3-1} + var_1 + var_2 + var_3$$

H_0: $\mu_1 = \mu_2 = \mu_3$

H_0: *At least one different*

$$F_{stat} = \frac{\dfrac{\text{Variance between groups}}{\#\ of\ groups - 1}}{\dfrac{\text{Variance within groups}}{\#\ of\ obs. - 1}} \sim F(\#\ of\ groups - 1,\ \#\ of\ obs. - 1)$$

One-way ANOVA

If we have a single categorical explanatory variable then the ANOVA is called one-way.

If that variable has only 2 levels then in fact the ANOVA is the same as a t-test for the differences between 2 means (not paired), which is in turn the same as the linear regression.

```
> set.seed(12345)
> irrigation=rep(c("water","Brawndo"),c(30,30))
> yield=ifelse(irrigation=="water",120,110)+rnorm(60,sd=15)
```

Simple Example

```
> t.test(yield~irrigation,var.equal=TRUE)

Two Sample t-test

data:  yield by irrigation
t = -1.4935, df = 58, p-value = 0.1407
alternative hypothesis: true difference in means is
not equal to 0
95 percent confidence interval:
  -15.232545   2.215061
sample estimates:
  mean in group Brawndo   mean in group water
114.6734              121.1821
```

Example continued

```
> Idiocracy=lm(yield~irrigation)
> summary(Idiocracy)

Call:
lm(formula = yield ~ irrigation)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      114.673      3.082  37.211   <2e-16 ***
irrigationwater    6.509      4.358   1.493    0.141

Residual standard error: 16.88 on 58 degrees of freedom
Multiple R-squared: 0.03703,  Adjusted R-squared: 0.02043
F-statistic:  2.23 on 1 and 58 DF,  p-value: 0.1407
```

Example continued

```
>summary(aov(yield~irrigation))

             Df Sum Sq Mean Sq F value Pr(>F)
irrigation    1    635   635.5    2.23  0.141
Residuals    58  16524   284.9

> anova(Idiocracy)

Analysis of Variance Table

Response: yield
           Df  Sum Sq Mean Sq F value Pr(>F)
irrigation  1   635.5  635.46  2.2304 0.1407
Residuals  58 16524.4  284.90
```

Example: 3-levels

Typically don't see ANOVA with one explanatory variable with only two levels (i.e. usually see more than 2 levels), because it is just the same as a t-test.

```
> set.seed(1234567)
> irrigation=rep(c("Aqua","Brawndo","Uber_Gro"),c(30,30,30)
> Means=rep(NA,90)
> Means[irrigation=="Aqua"]=120
> Means[irrigation=="Brawndo"]=115
> Means[irrigation=="Uber_Gro"]=125
> yield=Means+rnorm(90,sd=15)
```

Example continued

```
> Idiocracy=lm(yield~irrigation)
> summary(Idiocracy)

Call:
lm(formula = yield ~ irrigation)

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          121.412      2.791  43.499   <2e-16 ***
irrigationBrawndo     -7.661      3.947  -1.941   0.0555 .
irrigationUber_Gro     3.981      3.947   1.009   0.3160

Residual standard error: 15.29 on 87 degrees of freedom
Multiple R-squared: 0.09364,   Adjusted R-squared: 0.07281
F-statistic: 4.494 on 2 and 87 DF,  p-value: 0.01389
```

Example continued

```
> anova(Idiocracy)

Analysis of Variance Table

Response: yield
           Df  Sum Sq Mean Sq F value  Pr(>F)
irrigation  2  2100.7 1050.37  4.4942 0.01389 *
Residuals  87 20333.2  233.71
```

Multiple variables

If there are two explanatory variables it is called a two-way ANOVA.

```
> set.seed(1234567)
> irrigation=rep(c("Aqua","Brawndo","Uber_Gro"),c(30,30,30)
> soil=rep(rep(c("Clay","Loam"),3),rep(15,6))
> alpha=120
> MeansI=rep(NA,90)
> MeansI[irrigation=="Aqua"]=0
> MeansI[irrigation=="Brawndo"]=-5
> MeansI[irrigation=="Uber_Gro"]=+5
> MeansS=rep(NA,90)
> MeansS[soil=="Clay"]=0
> MeansS[soil=="Loam"]=-10
> yield=alpha+MeansI+MeansS+rnorm(90,sd=15)
```

Two-way example

```
> Idiocracy=lm(yield~irrigation+soil)
> summary(Idiocracy)

Call:
lm(formula = yield ~ irrigation + soil)

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         116.136      3.041  38.184   <2e-16 ***
irrigationBrawndo    -4.825      3.725  -1.295    0.199
irrigationUber_Gro    4.725      3.725   1.268    0.208
soilLoam             -3.953      3.041  -1.300    0.197

Residual standard error: 14.43 on 86 degrees of freedom
Multiple R-squared:  0.08766,  Adjusted R-squared: 0.05583
F-statistic: 2.754 on 3 and 86 DF,  p-value: 0.04738
```

Two-way example

```
> anova(Idiocracy)

Analysis of Variance Table

Response: yield
           Df  Sum Sq Mean Sq F value  Pr(>F)
irrigation  2  1368.2  684.08  3.2867 0.04212 *
soil        1   351.6  351.62  1.6894 0.19716
Residuals  86 17899.6  208.13
```

Two-way with interactions

We can include interaction terms in ANOVA.

```
> Idiocracy.int=lm(yield~irrigation+soil+irrigation*soil)
> summary(Idiocracy.int)

Call:
  lm(formula = yield ~ irrigation + soil + irrigation * soil)

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                  120.503      3.659  32.933   <2e-16 **
irrigationBrawndo             -9.679      5.175  -1.870   0.0649 .
irrigationUber_Gro            -3.521      5.175  -0.680   0.4981
soilLoam                     -12.686      5.175  -2.452   0.0163 *
irrigationBrawndo:soilLoam     9.707      7.318   1.326   0.1883
irrigationUber_Gro:soilLoam   16.492      7.318   2.254   0.0268 *

Residual standard error: 14.17 on 84 degrees of freedom
Multiple R-squared:  0.1402,Adjusted R-squared:  0.08901
F-statistic: 2.739 on 5 and 84 DF,  p-value: 0.02424
```

Two-way example

```
> anova(Idiocracy.int)

Analysis of Variance Table

Response: yield
Df  Sum Sq Mean Sq F value  Pr(>F)
irrigation      2  1368.2  684.08  3.4064 0.03780 *
soil            1   351.6  351.62  1.7509 0.18935
irrigation:soil  2  1030.6  515.32  2.5661 0.08285 .
Residuals      84 16868.9  200.82
```

Completely randomised designs

Statistical model:

$Y_{ij} = \mu_i + e_{ij}$, $e_{ij} \sim N(0, \sigma)$.

$H_0 : \mu_1 = \mu_2 = \ldots (= \mu)$.

$H_1 : H_0$ is not true.

Analysis: one-way ANOVA.

Randomised block design

ANOVA enables us to partition the total variation into three sources:

1. blocks (of little direct interest, but need to be accounted for);
2. treatments;
3. error.

Effect of drugs on lymphocyte levels in mice

4 mice from each of 5 litters; drugs randomised to mice within litters.

	Litter					
Drug	M	N	O	P	Q	$\bar{x}$
A	7.1	6.1	6.9	5.6	6.4	6.42
B	6.7	5.1	5.9	5.1	5.8	5.72
C	7.1	5.8	6.2	5.0	6.2	6.06
D	6.7	5.4	5.7	5.2	5.3	5.66
$\bar{x}$	6.90	5.60	6.18	5.23	5.93	5.97

Model:

$$\begin{aligned} \text{response} &= \text{overall mean} + \text{litter effect} + \text{drug effect} + \text{error} \\ Y_{ij} &= \mu + l_j + \alpha_i + e_{ij} \\ & i = 1, \ldots, 4, \quad j = 1, \ldots, 5, \quad e_{ij} \sim \mathrm{N}(0, \sigma) \end{aligned}$$

Hypothesis testing:

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 \quad (= 0)$. $\quad H_1 : H_0$ is not true.

Effect of drugs on lymphocyte levels in mice

```
> lymphocyte <- read.csv("../data/lymphocyte.csv")
> lymphocyte.lm = lm(count ~ litter+drug, data=lymphocyte)
> anova(lymphocyte.lm)

Analysis of Variance Table

Response: count
          Df Sum Sq Mean Sq F value    Pr(>F)
litter     4 6.4030 1.60075  30.155 3.544e-06 ***
drug       3 1.8455 0.61517  11.589 0.0007392 ***
Residuals 12 0.6370 0.05308
```

Randomised block design

What if blocks were wrongly ignored?

```
> lymphocyte.lm.1 <- lm(count ~ drug, data = lymphocyte)
> anova(lymphocyte.lm.1)
```

```
Analysis of Variance Table

Response: count
          Df Sum Sq Mean Sq F value Pr(>F)
drug       3 1.8455 0.61517  1.3981 0.2797
Residuals 16 7.0400 0.44000
```

Without blocks: residual MS = 0.440

With blocks: residual MS = 0.053

Without blocks:
Residual mean square is related to differences between *any* two mice, rather than differences between mice *within a litter*.

Latin square design

Model:

$$\begin{aligned} \text{response} &= \text{overall mean} + \text{row effect} + \text{column effect} \\ &\quad + \text{treatment effect} + \text{error} \\ Y_{ijk} &= \mu + r_i + c_j + \alpha_k + e_{ijk} \end{aligned}$$

ANOVA table for a Latin square design with t treatments:

Source	df	SS	MS	F	P
rows	$t-1$	SS_{row}	$SS_{row}/(t-1)$		
columns	$t-1$	SS_{col}	$SS_{col}/(t-1)$		
treatments	$t-1$	SS_{trt}	$SS_{trt}/(t-1)$	MS_{trt}/MS_{res}	P
residual	$(t-1)(t-2)$	SS_{res}	$SS_{res}/[(t-1)(t-2)]$		
total	t^2-1	SS_{tot}			

Latin square design

Food supplements and milk yield of cows:

Cow	Period I		Period II		Period III	
1	A	608	B	885	C	940
2	B	715	C	1087	A	766
3	C	844	A	711	B	832

```
> milk.yield.lm <- lm(yield ~ cow + period + supp, data = milk)
> anova(milk.yield.lm)

Analysis of Variance Table

Response: yield
          Df Sum Sq Mean Sq F value  Pr(>F)
cow        2   5900    2950  1.2183 0.45079
period     2  47214   23607  9.7490 0.09303 .
supp       2 103436   51718 21.3584 0.04473 *
Residuals  2   4843    2421
```

Latin square design

```
> tapply(milk.yield$yield, milk.yield$supp, mean)

       A        B        C
695.0000 810.6667 957.0000
```

Large differences, but not highly significant.

A single 3×3 Latin square usually does not give sufficient precision:

- It has only 3 replicates;
- The ANOVA has only 2 error DF (t value $= 4.303$);
- Most experiments need at least 10 error DF.

Factorial experiments

- Investigate the effect of two or more factors.
- Have the number of **treatment combinations** equal to the product of the number of levels of each factor.

  e.g. a $\mathbf{3} \times \mathbf{2} \times \mathbf{4}$ factorial experiment has 24 treatment combinations;
- Are written as a^b if each factor has the same number of levels,

  e.g. 2^3 for 3 factors at 2 levels.
- Have a **factorial treatment structure**; this doesn't affect the design, which relates to the *blocking* structure.

Factorial experiments

Quality of pancakes

Experiment to examine the effect of amount of whey and a baking supplement on quality of pancakes:

Amount of whey

	0%	10%	20%	30%	mean
no supplement	4.4	4.6	4.5	4.6	
	4.5	4.5	4.8	4.7	4.63
	4.3	4.8	4.8	5.1	
supplement	3.3	3.8	5.0	5.4	
	3.2	3.7	5.3	5.6	4.34
	3.1	3.6	4.8	5.3	
mean	3.80	4.17	4.87	5.12	4.49

Factorial experiments

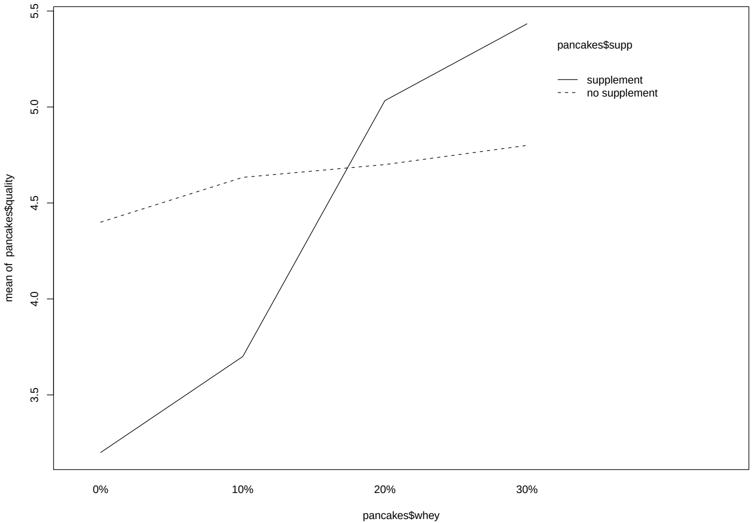
Quality of pancakes experiment:

```
> round(tapply(pancakes$quality, pancakes[, 1:2], mean), 2)
               whey
supp             0%  10%  20%  30%
  no supplement 4.4 4.63 4.70 4.80
  supplement    3.2 3.70 5.03 5.43
```

Interaction

- Interaction between two factors occurs when the differences between levels of one factor depend on the level of the other factor.
- The presence of interaction means that the effects are not *additive*.
- On an *interaction plot*, no interaction appears as parallel line segments. The greater the departure from "parallel" (or additivity), the greater the interaction.

Interaction plot



Interaction

Statistical model:

$$\begin{aligned} \text{quality} &= \text{overall mean} + \text{supplement effect} + \text{whey effect} \\ &\quad + \text{interaction effect} + \text{error} \\ y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \end{aligned}$$

Main effects:

- $\alpha_i,\ i = 1, 2$ (supplement effects);
- $\beta_j,\ j = 1$ to 4 (whey effects).

Interaction effects:

- γ_{ij} (additional effects arising from each combination of supplement and whey).

Interaction

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

Null hypotheses to test:

- $H_0 : \alpha_1 = \alpha_2 = 0$ (no main effect of supplement);
- $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (no main effect of whey);
- $H_0 : \gamma_{ij} = 0$ for all i, j (no interaction).

Interaction

```
> pancakes.lm <- lm(quality ~ supp * whey, data = pancakes)
> anova(pancakes.lm)
```

```
Analysis of Variance Table

Response: quality
          Df Sum Sq Mean Sq F value    Pr(>F)
supp       1 0.5104 0.51042  17.014 0.0007942 ***
whey       3 6.6912 2.23042  74.347 1.304e-09 ***
supp:whey  3 3.7246 1.24153  41.384 9.130e-08 ***
Residuals 16 0.4800 0.03000
```

- Small P-values $\Rightarrow$ reject all three null hypotheses;
- BUT there is significant interaction $\Rightarrow$ not useful to test main effects, because they are averaged across all the levels of the other factor;
- The null hypothesis concerning the interaction should be tested first;
- If it is accepted, then the null hypotheses about the main effects can be usefully tested.

Interaction

Inference when the interaction is significant:

Compare pairs of means for factor combinations.

Example: 95% confidence interval for
supp + 20% whey vs supp + 30% whey:

$$5.43 - 5.03 \pm t_{16}(.975) \times \sqrt{0.0300(\tfrac{1}{3} + \tfrac{1}{3})}$$

$$= 0.40 \pm 2.120 \times 0.141 = 0.40 \pm 0.30 = (0.10, 0.70).$$

Note that supplement + 30% whey gives the largest mean value for quality. The next largest is supplement + 20%. The above CI suggests that the former is significantly larger than the latter. Hence it will be significantly larger than any other.

Conclusion:

supplement + 30% whey produces the best quality pancakes.

Interaction

Creating a single factor `cookmethod`:

$4 \times 2 = 8$ combinations of supplement $\times$ whey.

An ANOVA with cookmethod as the only factor in the model results in:

```
> pancakes.lm.1 <- lm(quality ~ cookmethod, data = pancakes)
> anova(pancakes.lm.1)

Analysis of Variance Table

Response: quality
           Df Sum Sq Mean Sq F value    Pr(>F)
cookmethod  7 10.926  1.5609   52.03 7.94e-10 ***
Residuals  16  0.480  0.0300
```

Add SS and df for supp, whey, and the supp $\times$ whey interaction.

Factorial experiments

Degrees of freedom for analysing a factorial experiment

- The total df is 1 fewer than the total number of observations.
- The df for a main effect is 1 fewer than the number of levels of the factor.
- The df for an interaction term is the product of the df of the factors making up the interaction (provided there are no missing treatment combinations).

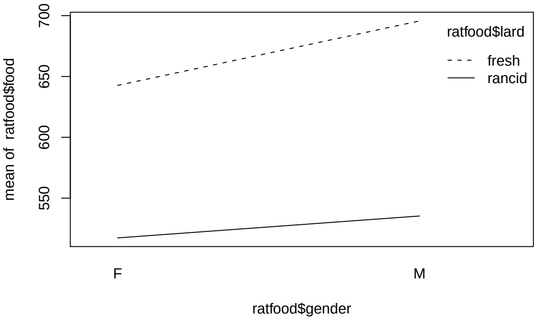
Interaction

Inference when the interaction is not significant:

Food consumption in male and female rats: fresh vs rancid lard.

	Lard		
Gender	Fresh	Rancid	row mean
M	709	592	
	679	538	
	699	476	
mean	695.7	535.3	615.5
F	657	508	
	594	505	
	677	539	
mean	642.7	517.3	580.0
column mean	669.2	526.3	

Interaction




```
Analysis of Variance Table

Response: food
            Df Sum Sq Mean Sq F value    Pr(>F)
gender       1   3781    3781  2.5925 0.1460358
lard         1  61204   61204 41.9685 0.0001925 ***
gender:lard  1    919     919  0.6300 0.4502546
Residuals    8  11667    1458
```

Interaction

Because the interaction is not significant, it is useful to compare means for the main effects.