# Analysis of Variance

The ASTA team

Contents

1 One way analysis of variance

1.1 Example

- The data set `chickwts` is available in R, and on the course webpage.
- Newly hatched chicks were randomly allocated into six groups, and each group was given a different feed supplement.

- Their weights in grams after six weeks are given along with feed types, i.e. we have a sample with corresponding measurements of 2 variables:

  - `weight`: a numeric variable giving the chick weight.
  - `feed`: a factor giving the feed type.

- Always start with some graphics:

```
library(mosaic)
gf_boxplot(weight ~ feed, data = chickwts)
```
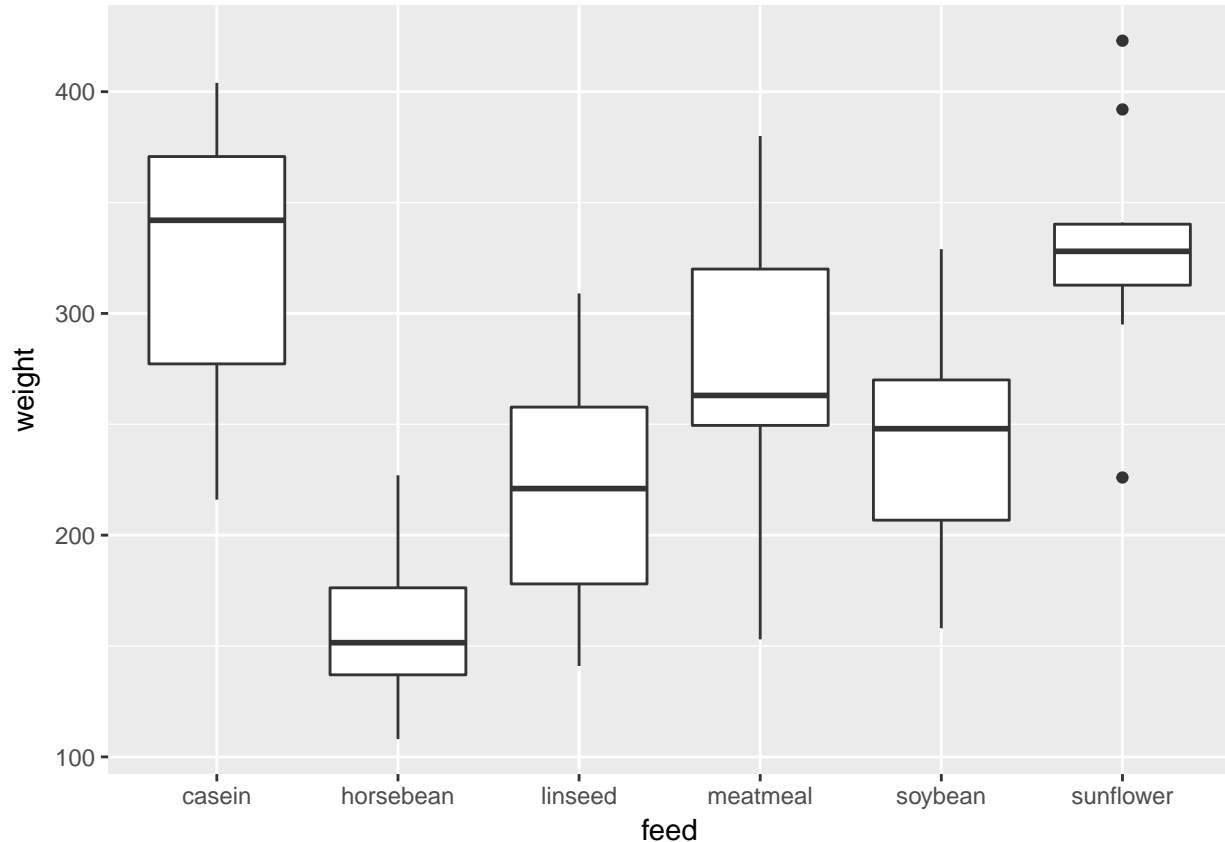

1.2 The ANOVA Model

- We measure the response y which in this case is `weight`.
- We want to study the effect of the factor x on y. In this case $x =$ `feed` and divides the sample in $g = 6$ groups.
- The mean responses within the groups are denoted $\mu_1, \mu_2, \ldots, \mu_g$.
- We will assume that

  - $y = \mu_x + \epsilon$, when y is a response in group x
  - ϵ are a sample from a population with mean zero and standard deviation σ.
  - The standard deviation for the population in each group is the same and equals σ
  - The response variable, y, is normal distributed within each group.

- The ANOVA test is a *test of independence* between the quantitative response variable and the qualitative explanatory variables.

2 Estimation of mean values

2.1 Estimates

- Least squares estimates for population means $\widehat{\mu}_x$ is given by the average of the response measurements in group x.
- For a given measured response y we let $\widehat{y}$ denote the model's prediction of y, i.e.

$$\widehat{y} = \widehat{\mu}_x$$

  if y is a response for an observation in group x.
- We use `mean` to find the mean, for each group:

```
mean(weight ~ feed, data = chickwts)
```

```
##    casein horsebean   linseed  meatmeal   soybean sunflower
##  323.5833  160.2000  218.7500  276.9091  246.4286  328.9167
```

- We can e.g. see that $\widehat{y} = 323.6$, when `feed=casein` but $\widehat{y} = 160.2$, when `feed=horsebean`.
- Is it a significant difference?

2.2 Contrast coding

- In many cases there is a group corresponding to "no treatment" and we are interested in the effect of different treatments.
- In this example we only have different `feeds`, which are sorted in lexicographical order by R, so `casein` is the reference.
- We can specify the model via:

  - `Intercept` corresponding to the mean response for the reference (`casein`).
  - For each of the other groups we have a **contrast**, which measures **the difference** between the mean value for that group and the reference group.

- For a given contrast we can calculate standard error, t-score and p-value, and thereby investigate whether there is a difference between this group and the reference group.
- In Agresti this is referred to as using **dummy variables**.

2.3 Example

```
model <- lm(weight ~ feed, data = chickwts)
summary(model)
```

```
##
## Call:
## lm(formula = weight ~ feed, data = chickwts)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -123.909  -34.413    1.571   38.170  103.091
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    323.583      15.834  20.436  < 2e-16 ***
## feedhorsebean -163.383      23.485  -6.957 2.07e-09 ***
## feedlinseed   -104.833      22.393  -4.682 1.49e-05 ***
## feedmeatmeal   -46.674      22.896  -2.039 0.045567 *
## feedsoybean    -77.155      21.578  -3.576 0.000665 ***
## feedsunflower    5.333      22.393   0.238 0.812495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.85 on 65 degrees of freedom
## Multiple R-squared:  0.5417, Adjusted R-squared:  0.5064
## F-statistic: 15.36 on 5 and 65 DF,  p-value: 5.936e-10
```

- We get information about contrasts and their significance:
- `Intercept` corresponding to `casein` has `weight` different from zero ($p < 2 \times 10^{-16}$).
- Weight difference between `casein` and `horsebean` is extremely significant (p=2×10^{-9}).
- There is no significant weight difference between `casein` and `sunflower` (p=81%).

3   Overall test for effect

3.1   Graphical representation of models

- We have two alternative explanations of the data.
- Simple model with one parameter (mean): "The feed type doesn't matter. The weight is just random around a common mean value".
- Complex model with six parameters (means): "The feed type is important. For each feed type we get a different mean value and the weights are random around these values."
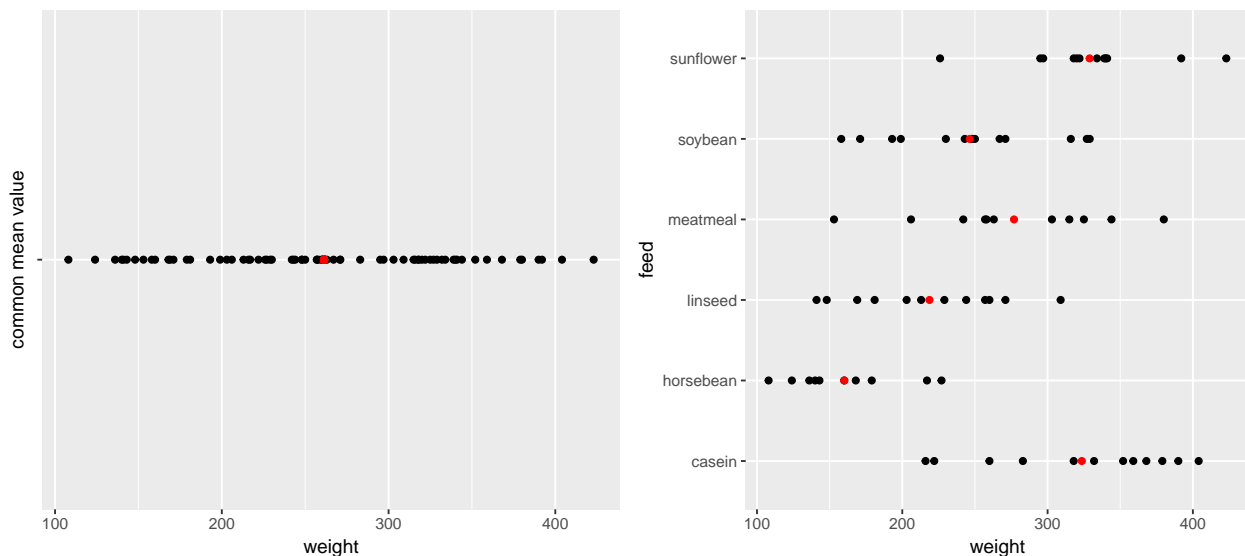
3.2 Hypotheses and test statistic

- Is the complex model significantly better (i.e. is there any effect of the explanatory grouping variable)? We can write the corresponding hypotheses in two different ways

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_g \quad \text{against} \quad H_a : \text{ At least 2 of the population means are different}$$

- Alternatively

$$H_0 : \text{ All contrasts are equal to zero.} \quad H_a : \text{ At least one contrast is non-zero.}$$

- We will (indirectly) use R^2 to do the test. If it is large, the complex model has good predictive power compared to the simple model. To judge significance we use

$$F_{obs} = \frac{(n-g)R^2}{(g-1)(1-R^2)} = \frac{(TSS - SSE)/(g-1)}{SSE/(n-g)}.$$

- Large values of R^2 implies large values of F_{obs}, which points to the alternative hypothesis.
- I.e. when we have calculated the observed value F_{obs}, then we have to find the probability that a new experiment would result in a larger value.

3.3 Interpretation of F statistic - Variance between/within groups

- It can be shown that the numerator of F_{obs} is a measure of **the variance between the groups**, i.e. how much "boxes" vary around the total average (the red line).

- Likewise it can be shown the denominator of F_{obs} is a measure for **the variance within groups**, i.e. how "tall" the boxes in the boxplot are.
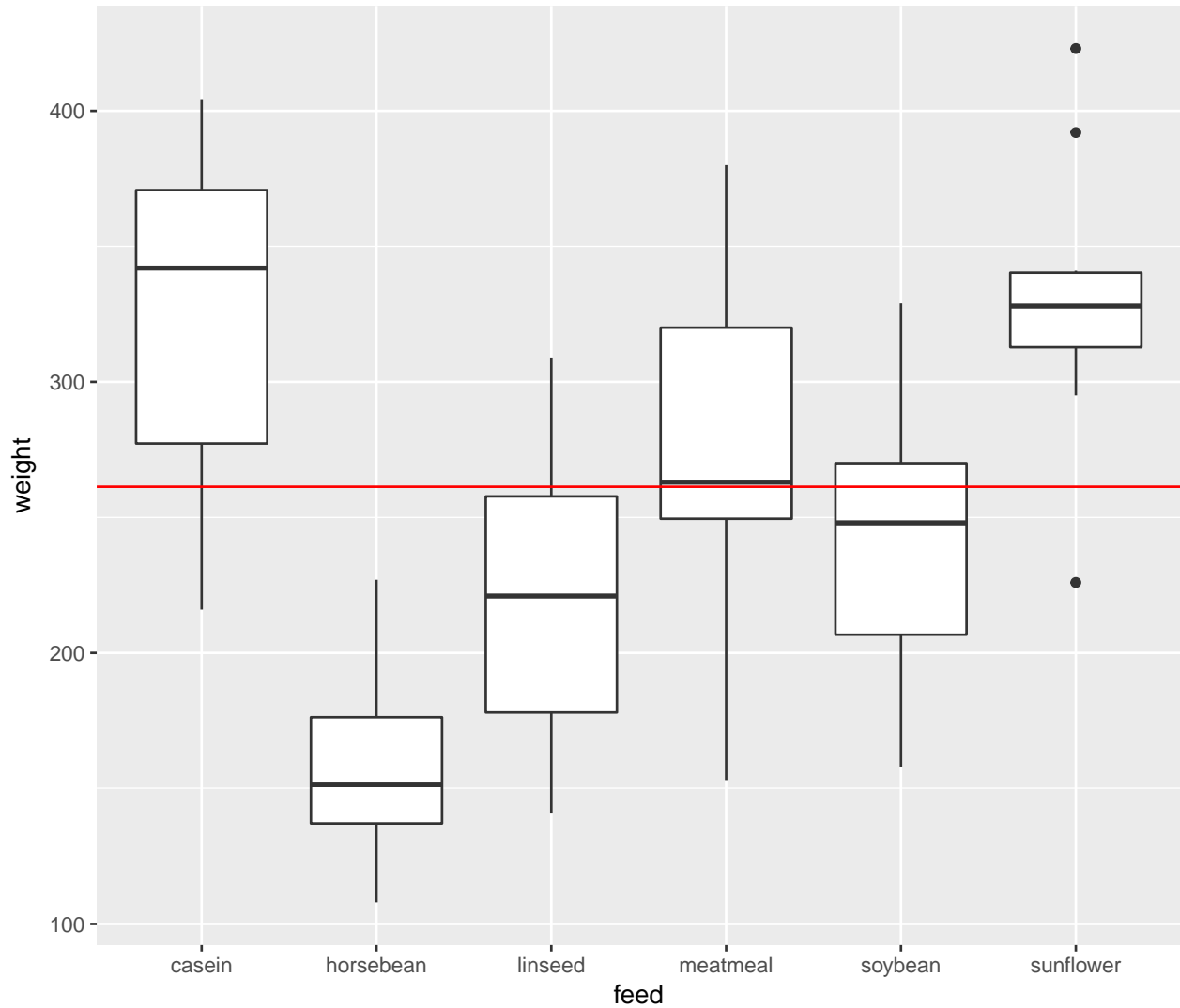
- If the boxes' deviations from the red line are to be explained by randomness, then the two types of variances should be of same magnitude. This is measured by the F-test statistic, which can be stated as

$$F_{obs} = \frac{\text{variance between groups}}{\text{variance within groups}}$$

3.4   Example

```
model <- lm(weight ~ feed, data = chickwts)
summary(model)
```

```
##
## Call:
## lm(formula = weight ~ feed, data = chickwts)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
```

```
## -123.909  -34.413    1.571   38.170  103.091
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    323.583     15.834  20.436  < 2e-16 ***
## feedhorsebean -163.383     23.485  -6.957 2.07e-09 ***
## feedlinseed   -104.833     22.393  -4.682 1.49e-05 ***
## feedmeatmeal   -46.674     22.896  -2.039 0.045567 *
## feedsoybean    -77.155     21.578  -3.576 0.000665 ***
## feedsunflower    5.333     22.393   0.238 0.812495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.85 on 65 degrees of freedom
## Multiple R-squared:  0.5417, Adjusted R-squared:  0.5064
## F-statistic: 15.36 on 5 and 65 DF,  p-value: 5.936e-10
```

- The last line gives us the value of $F_{obs} = 15.36$ and the corresponding p-value (5.9×10^{-10}). Clearly there is a significant difference between the types of `feed`.

4 Two way analysis of variance

4.1 Additive effects

- The data set `ToothGrowth` is available in **R** and on the webpage. For more info about this data, use `?ToothGrowth`.

- The data describes the tooth length in guinea pigs where some recvieve vitamin C treatment and others are given orange juice in different dosage.

- A total of 60 observations on 3 variables.

  - `[,1]` `len` The tooth length
  - `[,2]` `supp` The type of the supplement (`OJ` or `VC`)
  - `[,3]` `dose` The dosage (`LO, ME, HI`)

- We will study the response `len` with the predictors `supp` and `dose`.

- At first we look at the model with additive effects

  - `len`$=\mu +$ `"effect of supp"`$+$ `"effect of dose"` $+$ `error`

- This is also called the main effects model since it does not contain interaction terms.

- The parameter μ corresponds to the `Intercept` and is the mean tooth length in the reference group (supp `OJ`, dose `LO`).

- The effect of `supp` is the difference in mean when changing from `OJ` to `VC`.

- The effect of `dose` is the difference in mean when changing from `LO` to either`ME` or `HI`.

4.2 Dummy coding

- Let us introduce dummy variables:
    - $s_C = 1$ if supp VC and zero otherwise.
    - $d_M = 1$ if dose is ME and zero otherwise.
    - $d_H = 1$ if dose is HI and zero otherwise.

- Then we state the model

$$\text{length} = \mu + \beta_1 s_C + \beta_2 d_M + \beta_3 d_H + \text{error}.$$

- Interpretation:
    - μ is the expected tooth length when supp is OJ and dose is LO ($s_C = d_M = d_H = 0$)).
    - β_1 is the effect of supplement OJ to VC ($s_C = 1$).
    - β_2 is the effect of increasing dosage from LO to ME ($d_M = 1$).
    - β_3 is the effect of increasing dosage from LO to HI ($d_H = 1$).

4.3 Main effect model in R

- The main effects model is fitted by

```r
MainEff <- lm(len ~ supp + dose, data = ToothGrowth)
summary(MainEff)
```

```
##
## Call:
## lm(formula = len ~ supp + dose, data = ToothGrowth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.085  -2.751  -0.800   2.446   9.650
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.4550     0.9883  12.603  < 2e-16 ***
## suppVC       -3.7000     0.9883  -3.744 0.000429 ***
## doseME        9.1300     1.2104   7.543 4.38e-10 ***
## doseHI       15.4950     1.2104  12.802  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.828 on 56 degrees of freedom
## Multiple R-squared:  0.7623, Adjusted R-squared:  0.7496
## F-statistic: 59.88 on 3 and 56 DF,  p-value: < 2.2e-16
```

- The model has 4 parameters.
- The F test at the end compares with the (null) model with only one overall mean parameter. Does it seem like supp and dose has an additive effect?

4.4 Testing effect of supp

- Alternative model without effect of supp:

```
doseEff <- lm(len ~ dose, data = ToothGrowth)
```

- We can compare R^2 to see if `doseEff` (Model 1) is sufficent to explain the data compared to `MainEff` (Model 2). This is done by converting to F-statistic:

$$F_{obs} = \frac{(R_2^2 - R_1^2)/(df_1 - df_2)}{(1 - R_2^2)/df_2} = \frac{(SSE_1 - SSE_2)/(df_1 - df_2)}{(SSE_2)/df_2}.$$

- In **R** the calculations are done using `anova`:

```
anova(doseEff, MainEff)
```

```
## Analysis of Variance Table
##
## Model 1: len ~ dose
## Model 2: len ~ supp + dose
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     57 1025.78
## 2     56  820.43  1    205.35 14.017 0.0004293 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.5 Testing effect of dose

- Alternative model without effect of dose:

```
suppEff <- lm(len ~ supp, data = ToothGrowth)
anova(suppEff, MainEff)
```

```
## Analysis of Variance Table
##
## Model 1: len ~ supp
## Model 2: len ~ supp + dose
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     58 3246.9
## 2     56  820.4  2    2426.4 82.811 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
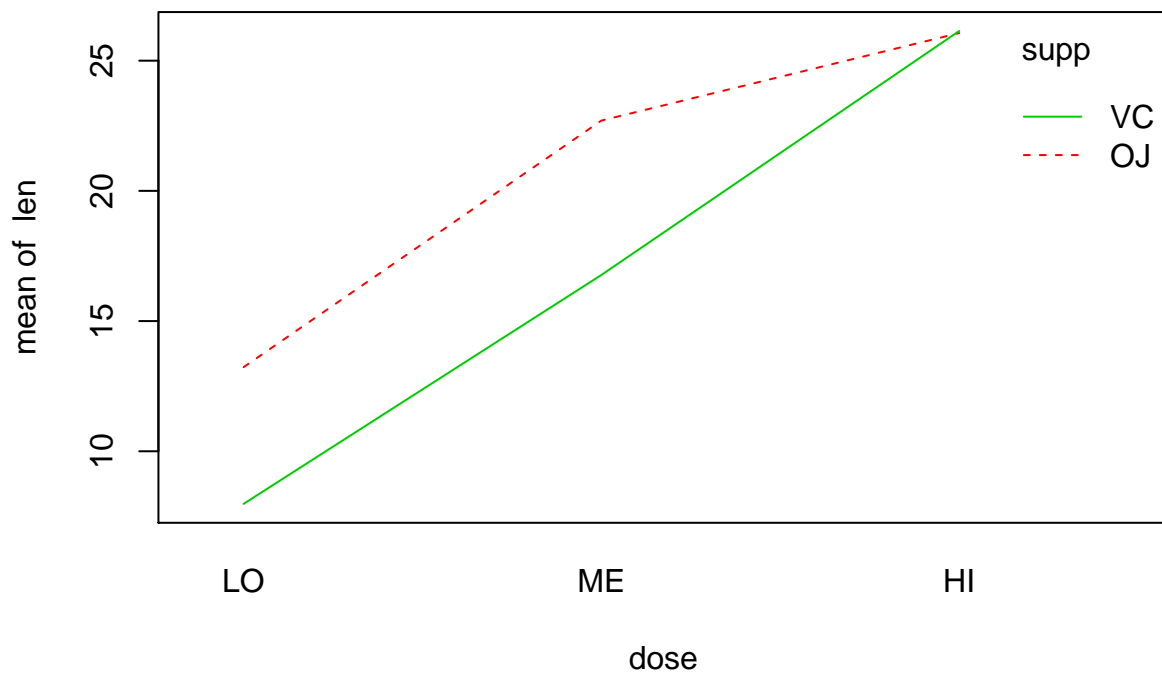
5 Interaction

5.1 Example

- We will extend the model by introducing an interaction between `supp` and `dose`.
- A graphical check for no interaction in the main effects model:
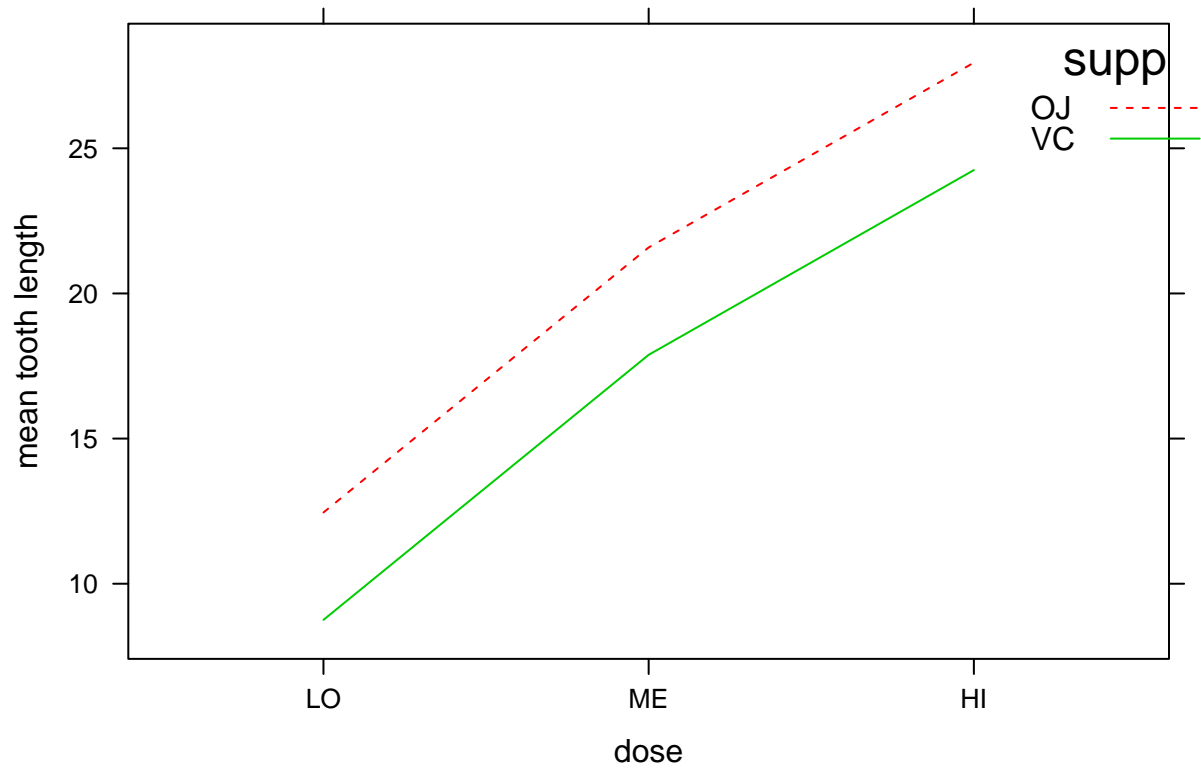
- Interaction plot:

```r
with(ToothGrowth, interaction.plot(dose, supp, len, col = 2:3))
```


- For each of the supplement types we plot the average number of tooth length as a function of dosage.
- If the main effects model is correct then the difference between supplements is the same for all levels of dosage, i.e. the curves should be parallel - except for noise.

10

- This does not seem to be the case.

5.2 Dummy coding

- The extended model can be formulated as

$$\texttt{length} = \mu + \beta_1 s_C + \beta_2 d_M + \beta_3 d_H + \beta_4 s_C d_M + \beta_5 s_C d_H + \texttt{error}$$

- Interpretation:
    - μ is the expected tooth length for `supp` OJ and `dose` LO ($s_C = d_M = d_H = 0$).
    - β_1 is the effect of changing from `supp` OJ to VC, `dose` is LO ($s_C = 1, d_M = d_H = 0$).
    - β_2 is the effect of increasing `dose` from LO to ME, when `supp` is OJ ($s_C = 0, d_M = 1$).
    - β_3 is the effect of increasing `dose` from LO to HI, when `supp` is OJ ($s_C = 0, d_H = 1$).
    - β_4 is an additional effect of both changing from `supp` OJ to VC and increasing `dose` from LO to ME ($s_C = 1, d_M = 1$)
    - β_5 is an additional effect of both changing from `supp` OJ to VC and increasing `dose` from LO to HI ($s_C = 1, d_H = 1$)

5.3 Example

- We fit the interaction model by changing plus to multiply in the model expression from before:

```
Interaction <- lm(len ~ supp*dose, data = ToothGrowth)
```

- Now we can think of an experiment with 6 groups corresponding to each combination of the predictors.
- Looking at the group averages it looks like, the supplement types behave quite differently depending on `dose`:

```
mean(len ~ supp + dose, data = ToothGrowth)
```

```
## OJ.LO VC.LO OJ.ME VC.ME OJ.HI VC.HI
## 13.23  7.98 22.70 16.77 26.06 26.14
```

- But is that significant?

```
anova(MainEff, Interaction)
```

```
## Analysis of Variance Table
##
## Model 1: len ~ supp + dose
## Model 2: len ~ supp * dose
##   Res.Df    RSS Df Sum of Sq     F  Pr(>F)
## 1     56 820.43
## 2     54 712.11  2    108.32 4.107 0.02186 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- With a p-value of 2.1860269% there is a significant interaction `supp:dose`, i.e. the lack of parallel curves in the interaction plot is significant.

```
summary(Interaction)
```

```
##
## Call:
## lm(formula = len ~ supp * dose, data = ToothGrowth)
##
## Residuals:
##     Min     1Q Median     3Q    Max
##   -8.20  -2.72  -0.27   2.65   8.27
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     13.230      1.148  11.521 3.60e-16 ***
## suppVC          -5.250      1.624  -3.233  0.00209 **
## doseME           9.470      1.624   5.831 3.18e-07 ***
## doseHI          12.830      1.624   7.900 1.43e-10 ***
## suppVC:doseME   -0.680      2.297  -0.296  0.76831
## suppVC:doseHI    5.330      2.297   2.321  0.02411 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.631 on 54 degrees of freedom
## Multiple R-squared:  0.7937, Adjusted R-squared:  0.7746
## F-statistic: 41.56 on 5 and 54 DF,  p-value: < 2.2e-16
```

- The additional effect of both changing from `supp` `OJ` to `VC` and increasing `dose` from `LO` to `MI` (β_5=`suppVC:doseHI`) is highly significant.