University of Zurich UZH

Makerere University

Data Analysis with R:

Day 6

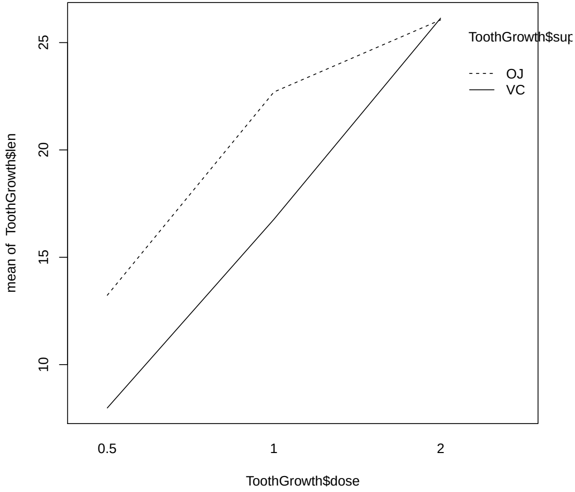
Sonja Hartnack, Terence Odoch & Muriel Buri

October 2017

Exercise 17

Interactions

```r
interaction.plot(ToothGrowth$dose, ToothGrowth$supp, ToothGrowth$len,
                 fixed = TRUE)
```



Two-way Interactions in R (1/3)

There are three different interactions:

- **Interaction between two categorical variables**
- Interaction between one continuous and one categorical variables
- Interaction between two continuous variables

Two-way Interactions in R (2/3)

Model specification in R: Be aware, an interaction is never tested without its corresponding main effects included in the model.

```
# Interaction between two categorical variables
# mod.dose.supp.int <- lm(len ~ dose.fac + supp + dose.fac:supp,
# data = ToothGrowth)
mod.dose.supp.int <- lm(len ~ dose.fac * supp, data = ToothGrowth)
# summary(mod.dose.supp.int)
```

$$
\begin{aligned}
y \sim\ & \beta_{\text{baseline((dose==low) \& (supp==OJ))}} + \beta_{\text{dose==med}} + \\
& \beta_{\text{dose==high}} + \beta_{\text{supp==VC}} \\
& + \beta_{\text{(dose==med) \& (supp==VC)}} \\
& + \beta_{\text{(dose==high) \& (supp==VC)}}
\end{aligned}
$$

Two-way Interactions in R (3/3)

```r
# Interaction between two categorical variables
# mod.dose.supp.int <- lm(len ~ dose.fac + supp + dose.fac:supp,
# data = ToothGrowth)
mod.dose.supp.int <- lm(len ~ dose.fac * supp, data = ToothGrowth)
summary(mod.dose.supp.int)


##
## Call:
## lm(formula = len ~ dose.fac * supp, 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 ***
## dose.facmed            9.470      1.624   5.831 3.18e-07 ***
## dose.fachigh          12.830      1.624   7.900 1.43e-10 ***
## suppVC                -5.250      1.624  -3.233  0.00209 **
## dose.facmed:suppVC    -0.680      2.297  -0.296  0.76831
## dose.fachigh:suppVC    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
```

Two-way Interactions in R: Interpretation of coefficients (1/2)

```
coef(mod.dose.supp.int)

##         (Intercept)          dose.facmed         dose.fachigh
##               13.23                 9.47                12.83
##              suppVC   dose.facmed:suppVC  dose.fachigh:suppVC
##               -5.25                -0.68                 5.33
```

$$y \sim \beta_{\text{baseline((dose==low) \& (supp==OJ))}} + \beta_{\text{dose==med}} + \beta_{\text{dose==high}} + \beta_{\text{supp==VC}} + \beta_{\text{(dose==med) \& (supp==VC)}} + \beta_{\text{(dose==high) \& (supp==VC)}}$$

Two-way Interactions in R: Interpretation of coefficients (2/2)

```
coef(mod.dose.supp.int)
```

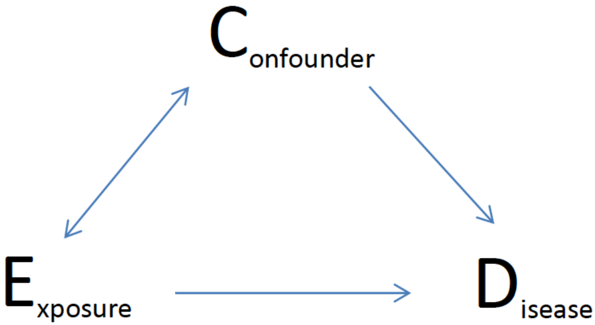
```
##         (Intercept)         dose.facmed        dose.fachigh
##               13.23                9.47               12.83
##              suppVC  dose.facmed:suppVC dose.fachigh:suppVC
##               -5.25               -0.68                5.33
```

- $\rightarrow$ `dose.facmed:suppVC`
  change in the slope between the `low` and `med` `dose.fac` group under the supplement type (`supp`) `VC` in comparison to the intercept. In other words, by changing the dose from `low` to `med` within the supplement group `VC`, the slope **de**creases by approx. -0.68.
- $\rightarrow$ `dose.fachigh:suppVC`
  change in the slope between the `low` and `high` `dose.fac` group under the supplement type (`supp`) `VC` in comparison to the intercept. In other words, by changing the dose from `low` to `high` within the supplement group `VC`, the slope **in**creases by approx. $+5.33$.

Exercise 18

Confounding

- **Confounding variable** - is associated with the exposure and the outcome
- **Confounding variable** - is not part of the causal path between exposure and the outcome



Some things to remember

- Not adjusting / controlling for a confounding variable may lead to bias results. It is good practice to present as well the crude (not adjusted ORs) and the adjusted ones. Adjustment is typically done if difference $> 10\%$.
- Check also for interaction (= effect modification) e. g. using logistic regression.
- Do not adjust for a variable C if it is a common effect of E and D (collider) or if it is in the causal pathway of E and D.