THE UNIVERSITY of EDINBURGH

Week 8: The Linear Model (ctd)

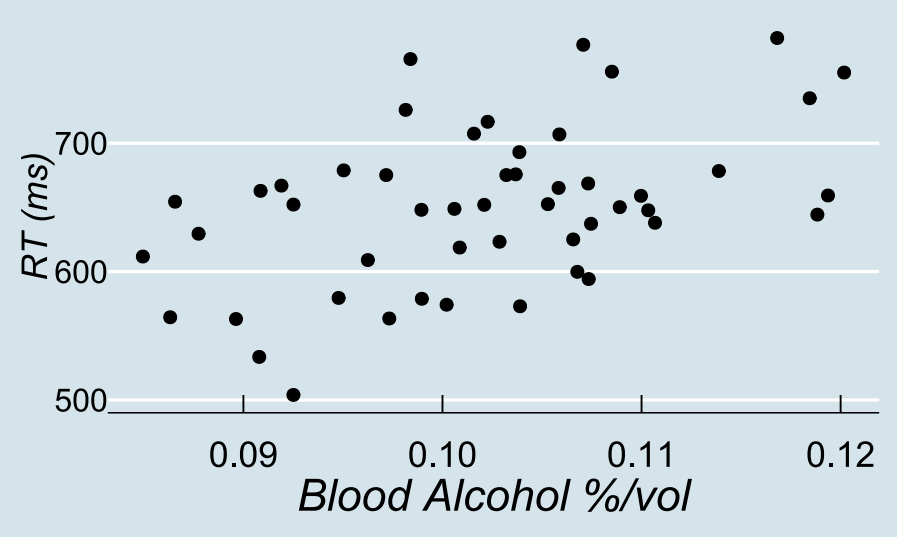
Univariate Statistics and Methodology using R

Department of Psychology
The University of Edinburgh

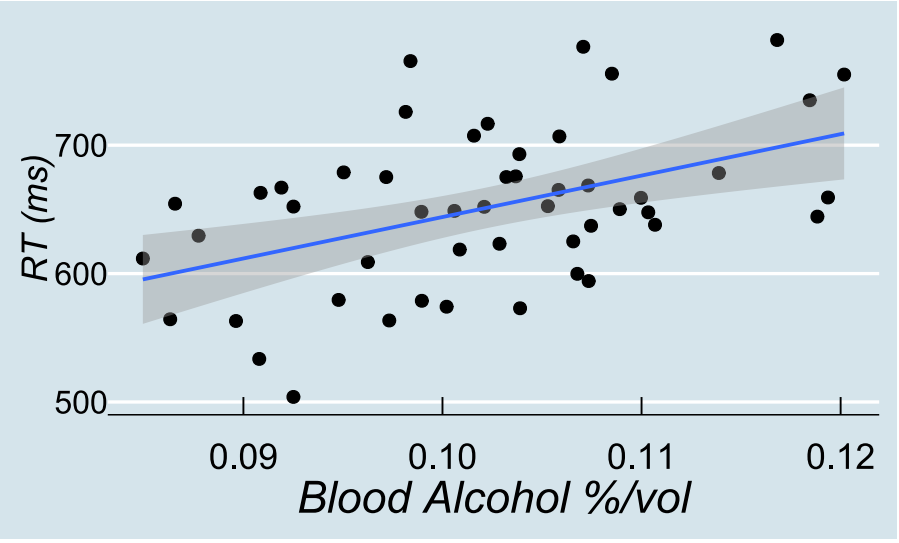
Part 1

Quick Refresh

Some new data



Some new data



"for every extra 0.01% blood alcohol, reaction time slows down by around 32 ms"

The Model

```
mod <- lm(RT~BloodAlc, data=dat)
summary(mod)
```

```
##
## Call:
## lm(formula = RT ~ BloodAlc, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -115.92  -40.42    1.05   42.93  126.64
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)      321         91    3.53  0.00093 ***
## BloodAlc        3228        888    3.64  0.00067 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.8 on 48 degrees of freedom
## Multiple R-squared:  0.216,    Adjusted R-squared:   0.2
## F-statistic: 13.2 on 1 and 48 DF,  p-value: 0.000673
```

Another (identical) Model

```
dat <- dat %>% mutate(BloodAlc100 = BloodAlc*100)
mod2 <- lm(RT~BloodAlc100, data=dat)
summary(mod2)
```

```
##
## Call:
## lm(formula = RT ~ BloodAlc100, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -115.92  -40.42    1.05   42.93  126.64
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   321.24      91.05    3.53  0.00093 ***
## BloodAlc100    32.28       8.88    3.64  0.00067 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.8 on 48 degrees of freedom
## Multiple R-squared:  0.216,    Adjusted R-squared:   0.2
## F-statistic: 13.2 on 1 and 48 DF,  p-value: 0.000673
```

Part 2

Checking Assumptions

Assumptions of Linear Models

required

- **linearity** of relationship(!)
- for the *residuals*:
  - **normality**
  - **homogeneity of variance**
  - **independence**

desirable

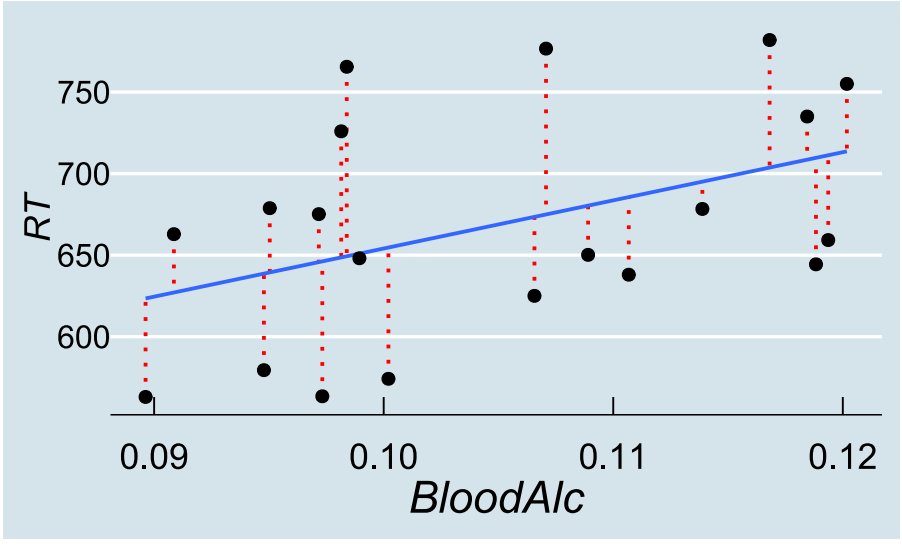
- no 'bad' (overly influential) observations

Residuals

$$y_i = b_0 + b_1 \cdot x_i + \epsilon_i$$

$$\epsilon \sim N(0, \sigma) \text{ independently}$$

- normally distributed (mean should be $\simeq$ zero)
- homogeneous (differences from $\hat{y}$ shouldn't be systematically smaller or larger for different x)
- independent (residuals shouldn't influence other residuals)



At A Glance

```
summary(mod)
```

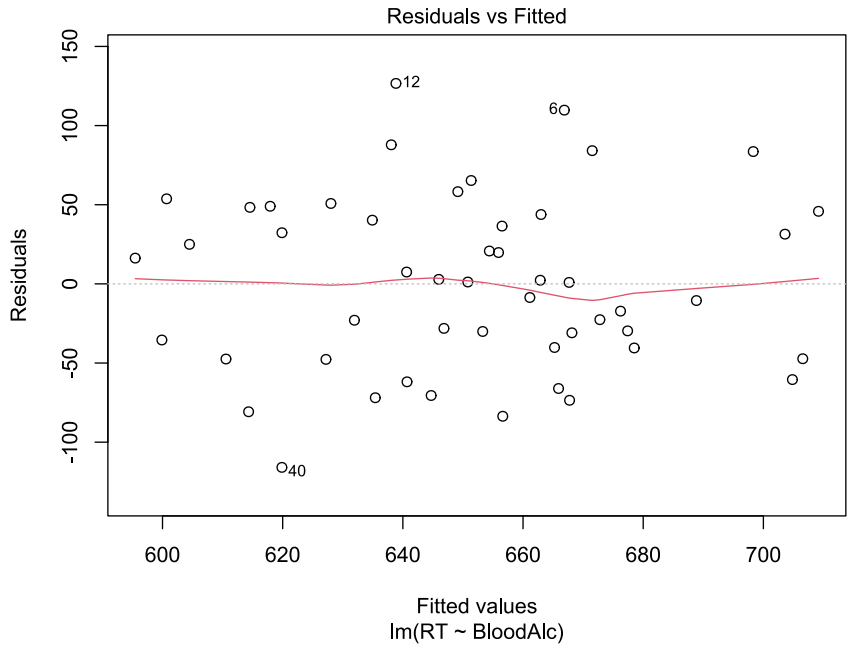
```
##
## Call:
## lm(formula = RT ~ BloodAlc, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -115.92  -40.42    1.05   42.93  126.64
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)      321         91    3.53  0.00093 ***
## BloodAlc        3228        888    3.64  0.00067 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.8 on 48 degrees of freedom
## Multiple R-squared:  0.216,    Adjusted R-squared:   0.2
## F-statistic: 13.2 on 1 and 48 DF,  p-value: 0.000673
```

In More Detail

linearity

```
plot(mod,which=1)
```

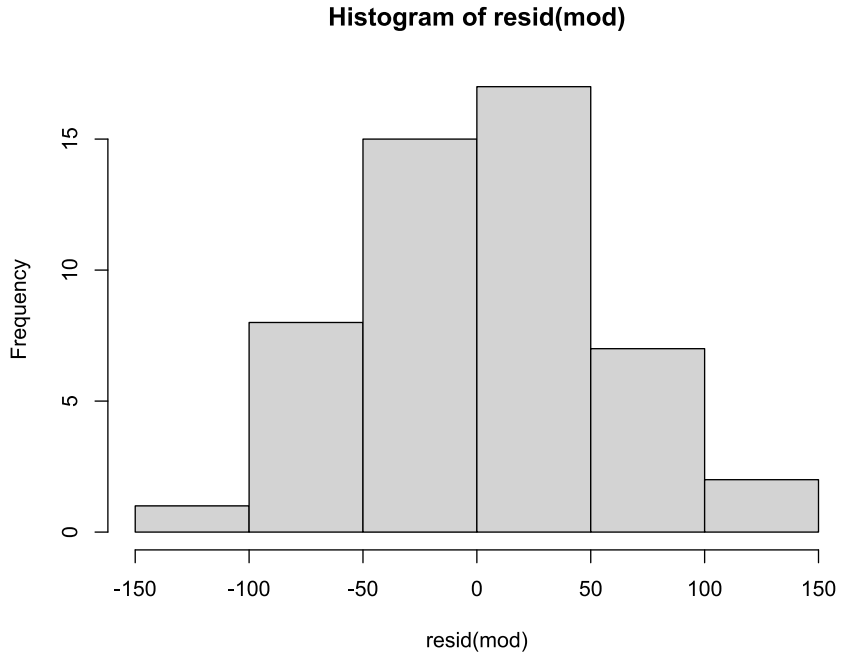
- plots residuals ϵ_i against fitted values $\hat{y}_i$
- the 'average residual' is roughly zero across $\hat{y}$, so relationship is likely to be linear



In More Detail

normality

```
hist(resid(mod))
```

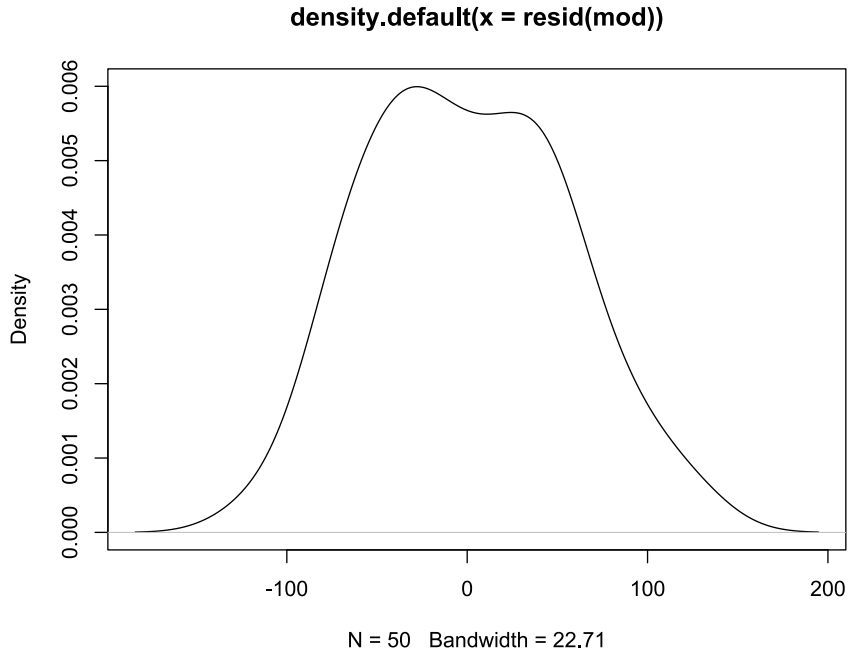


In More Detail

normality

```
plot(density(resid(mod)))
```

- check that residuals ϵ are approximately normally distributed
- in fact there's a better way of doing this...

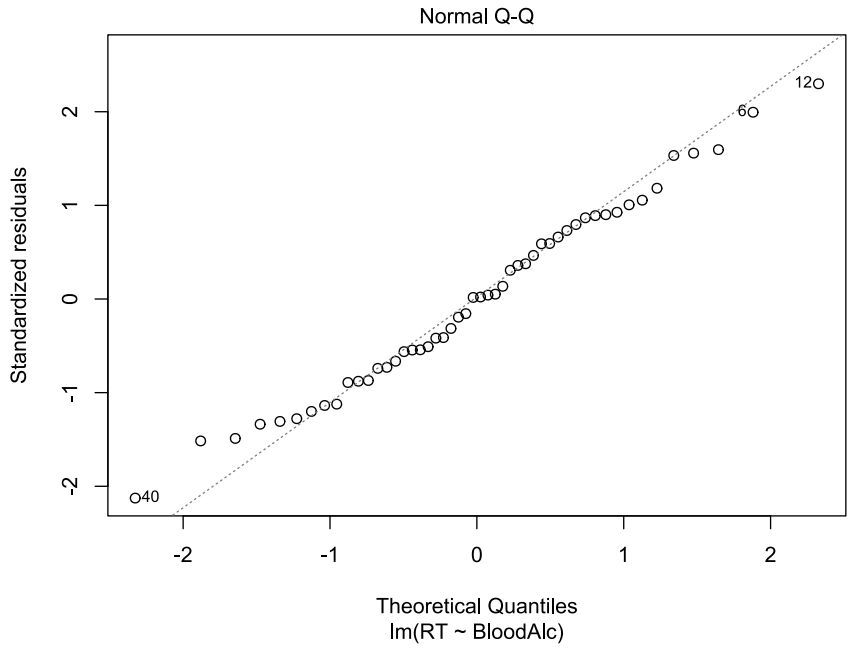


In More Detail

normality

```
plot(mod,which=2)
```

- **Q-Q plot** compares the residuals ϵ against a known distribution (here, normal)
- observations close to the straight line mean residuals are approximately normal
- numbered observations refer to *row numbers* in the original data, for checking



Q-Q Plots

y axis

- Our residuals, in terms of "standard deviations from the mean":
  $\text{standardized residual} = \frac{\text{residual} - mean(\text{residual})}{sd(\text{residual})}$

```
scale(resid(mod))
```

```
[1] -2.10063 -1.51541 -1.46399 -1.33279 -1.30350
[6] -1.27782 -1.19769 -1.12129 -1.09559 -0.86440
...
...
```

Q-Q Plots

y axis

- Our residuals, in terms of "standard deviations from the mean":
  $\text{standardized residual} = \frac{\text{residual} - mean(\text{residual})}{sd(\text{residual})}$

```
scale(resid(mod))
```

```
[1] -2.10063 -1.51541 -1.46399 -1.33279 -1.30350
[6] -1.27782 -1.19769 -1.12129 -1.09559 -0.86440
...
...
```

x axis

- we have 50 residuals.
- for a normal distribution, what values *should* 1/50th, 2/50th, 3/50th (etc) of the observations lie below?
- expressed in "standard deviations from the mean"

```
qnorm(c(1/50,2/50,3/50))
```

```
## [1] -2.054 -1.751 -1.555
```

Q-Q Plots

y axis

- Our residuals, in terms of "standard deviations from the mean":
  $\text{standardized residual} = \frac{\text{residual} - mean(\text{residual})}{sd(\text{residual})}$

```
scale(resid(mod))
```

```
[1] -2.10063 -1.51541 -1.46399 -1.33279 -1.30350
[6] -1.27782 -1.19769 -1.12129 -1.09559 -0.86440
...
...
```

- Q-Q Plot shows these values plotted against each other

x axis

- we have 50 residuals.
- for a normal distribution, what values *should* 1/50th, 2/50th, 3/50th (etc) of the observations lie below?
- expressed in "standard deviations from the mean"

```
qnorm(c(1/50,2/50,3/50))
```

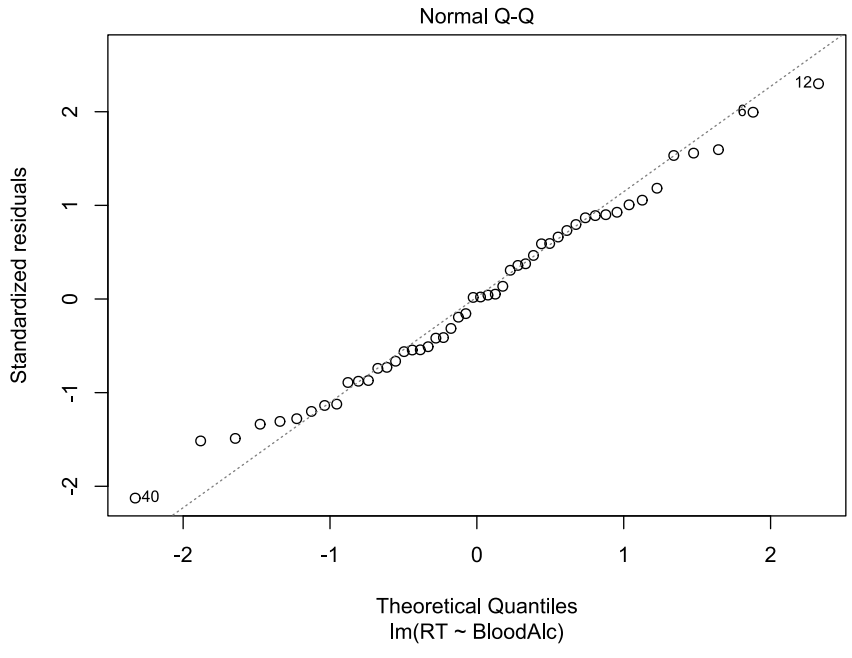
```
## [1] -2.054 -1.751 -1.555
```

In More Detail

normality

```
plot(mod,which=2)
```

- **Q-Q plot** compares the residuals ϵ against a known distribution (here, normal)
- observations close to the straight line mean residuals are approximately normal
- numbered observations refer to *row numbers* in the original data, for checking

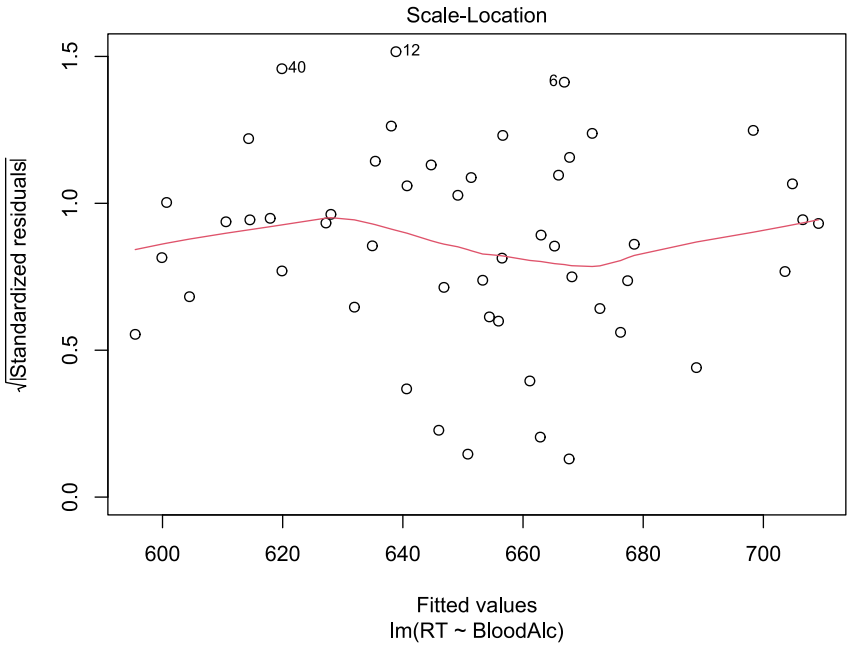


In More Detail

homogeneity of variance

```
plot(mod,which=3)
```

- the *size* of the residuals is approximately the same across values of $\hat{y}$



Visual vs Other Methods

- statistical ways of checking assumptions are introduced in the reading
- they tend to have limitations (for example, they're susceptible to sample size)
- nothing beats looking at plots like these (and `plot(<model>)` makes it easy)
- however, two things:

Visual vs Other Methods

- statistical ways of checking assumptions are introduced in the reading
- they tend to have limitations (for example, they're susceptible to sample size)
- nothing beats looking at plots like these (and `plot(<model>)` makes it easy)
- however, two things:

1. there are no criteria for deciding exactly when assumptions are sufficiently met
    - it's a matter of experience and judgement
2. we need to talk about **independence** of residuals

End of Part 2

Part 3

Independence, Influence

Independence

- no easy way to check independence of residuals
- in part, because it depends on the *source* of the observations

- one determinant might be a single person being observed multiple times
- e.g., my reaction times might tend to be slower than yours
  $\rightarrow$ multivariate statistics

Independence

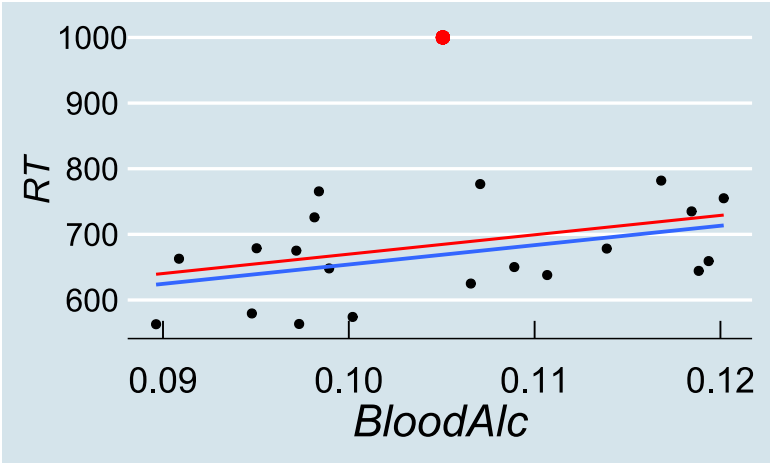
- another determinant might be *time*
- observations in a sequence might be autocorrelated
- can be checked using the Durbin-Watson Test from the `car` package

```
library(car)
dwt(mod)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1         -0.1377          2.22    0.42
##  Alternative hypothesis: rho != 0
```

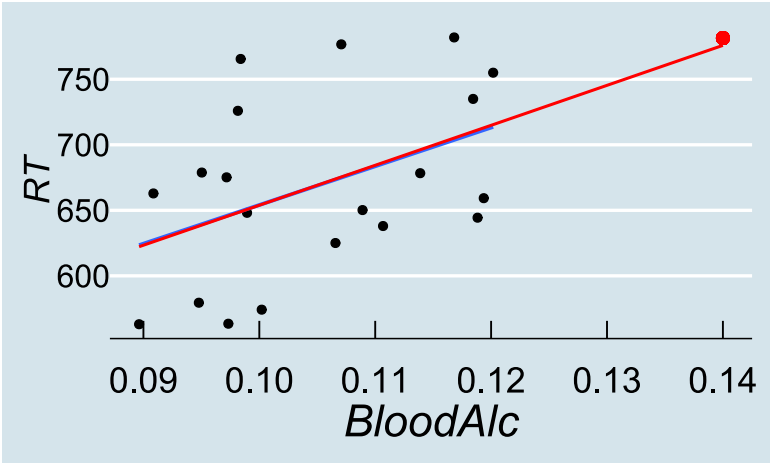
- shows no autocorrelation at lag 1

Influence



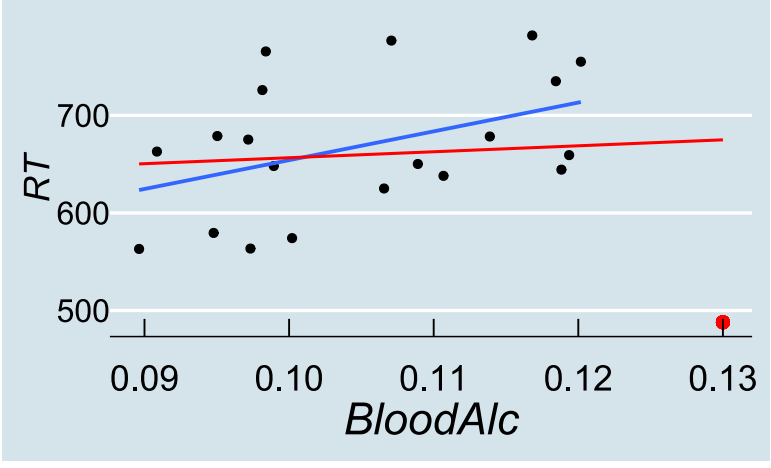
- even substantial **outliers** may only have small effects on the model
- here, only the intercept is affected

Influence



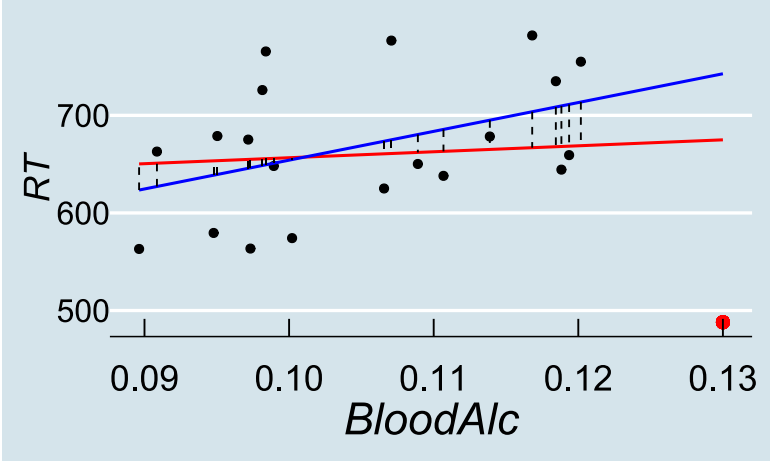
- observations with high **leverage** are inconsistent with other data, but may not be distorting the model

Influence



- we care about observations with high **influence** (outliers with high leverage)

Cook's Distance



- a standardised measure of "how much the model differs without observation i"

Cook's Distance

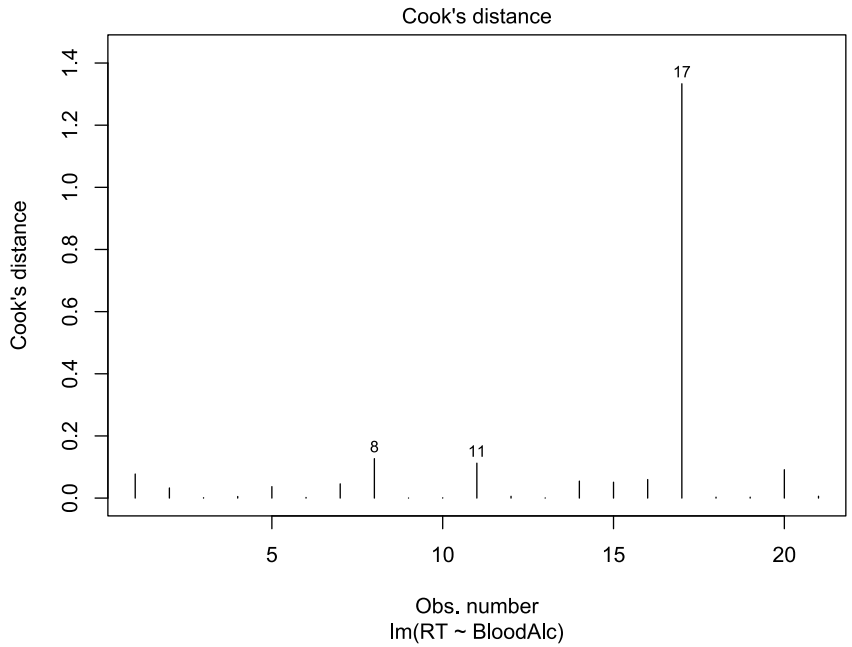
$$D_i = \frac{\sum_{j=1}^{n} (\hat{y}_j - \hat{y}_{j(i)})^2}{(p+1)\hat{\sigma}^2}$$

- $\hat{y}_j$ is the jth fitted value
- $\hat{y}_{j(i)}$ is the jth value from a fit which doesn't include observation i
- p is the number of regression coefficients
- $\hat{\sigma}^2$ is the estimated variance from the fit, i.e., mean squared error

Cook's Distance

```
plot(mod,which=4)
```

- observations labelled by row
- various rules of thumb, but start looking when Cook's Distance > 0.5



Learning to Read

- the Playmo School has been evaluating its reading programmes, using 50 students
- ages of students
- hours per week students spend reading of their own volition
- whether they are taught using phonics or whole-word methods
- **outcome: "reading age"**

Learning to Read

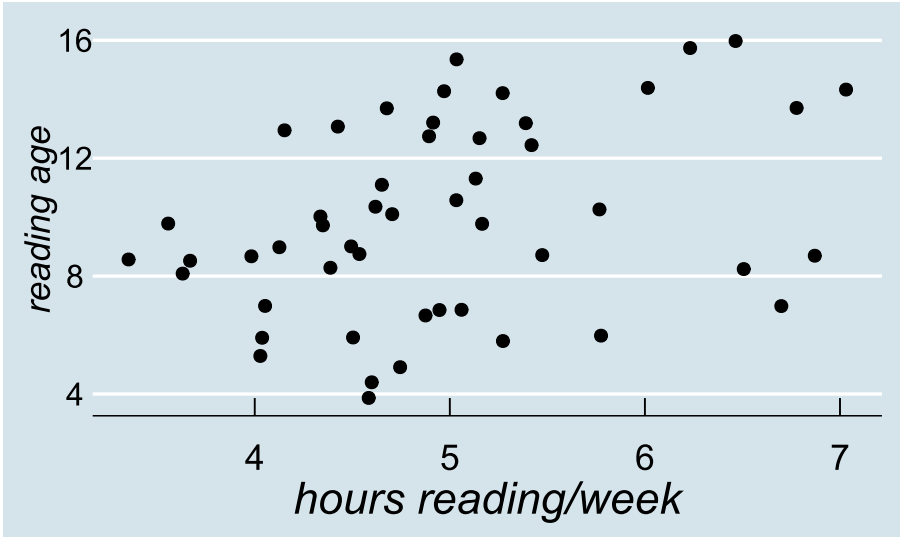
age	hrs_wk	method	R_AGE
10.115	4.971	phonics	14.272
9.940	4.677	phonics	13.692
6.060	4.619	phonics	10.353
9.269	4.894	phonics	12.744
10.991	5.035	phonics	15.353
6.535	5.272	word	5.798
8.150	6.871	word	8.691
7.941	4.053	word	6.988
8.233	5.474	word	8.713
6.219	4.038	word	5.908

Learning to Read

age	hrs_wk	method	R_AGE
10.115	4.971	phonics	14.272
9.940	4.677	phonics	13.692
6.060	4.619	phonics	10.353
9.269	4.894	phonics	12.744
10.991	5.035	phonics	15.353
6.535	5.272	word	5.798
8.150	6.871	word	8.691
7.941	4.053	word	6.988
8.233	5.474	word	8.713
6.219	4.038	word	5.908

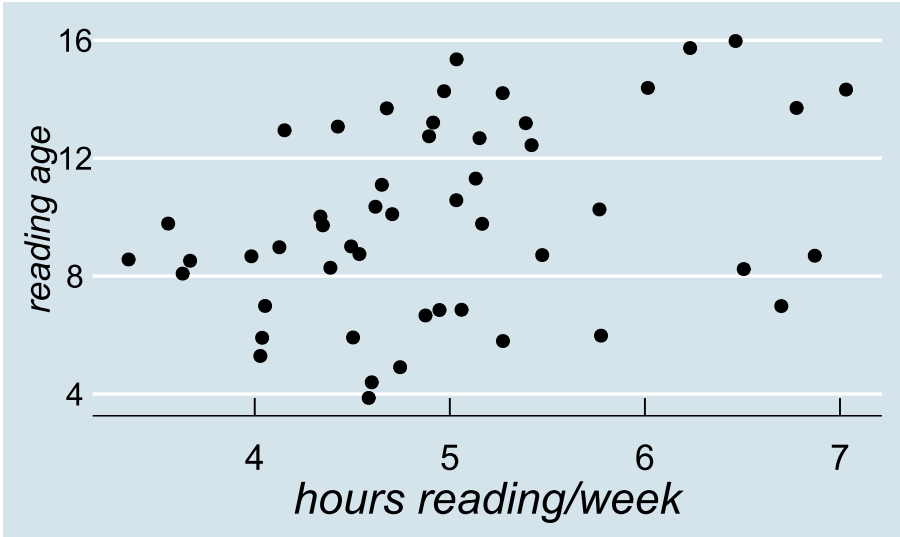
Does Practice Affect Reading Age?

```
p <- reading %>%
  ggplot(aes(x=hrs_wk,y=R_AGE)) +
  geom_point(size=3) +
  ylab("reading age") +
  xlab("hours reading/week")
p
```



Does Practice Affect Reading Age?

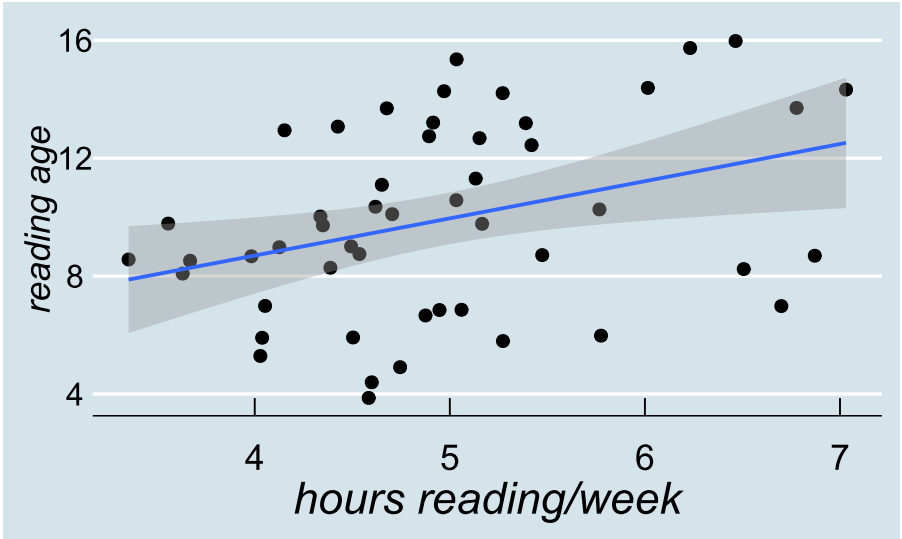
```
p <- reading %>%
  ggplot(aes(x=hrs_wk,y=R_AGE)) +
  geom_point(size=3) +
  ylab("reading age") +
  xlab("hours reading/week")
p
```



- hours per week is correlated with reading age: $r = 0.3483, p = 0.0132$
- we can use a linear model to say something about the effect size

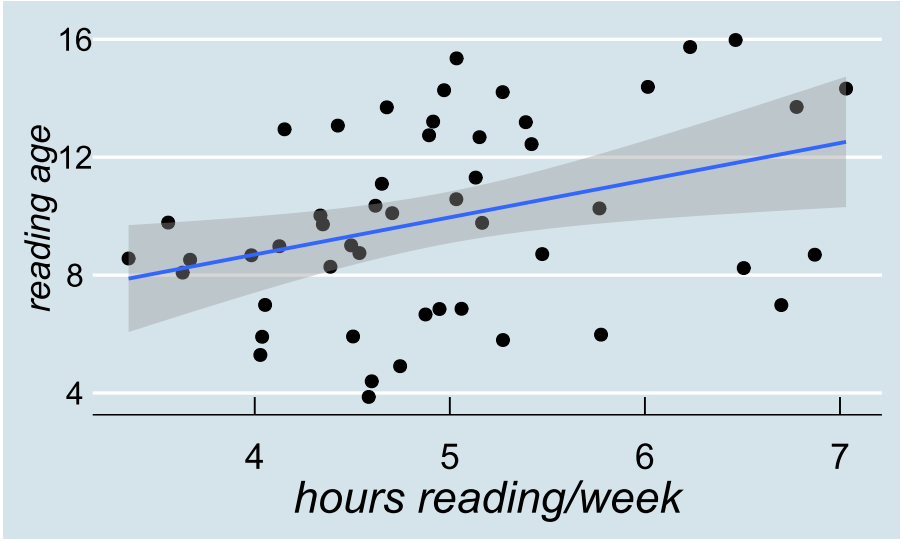
Does Practice Affect Reading Age?

```
p + geom_smooth(method="lm")
```



Does Practice Affect Reading Age?

```
p + geom_smooth(method="lm")
```



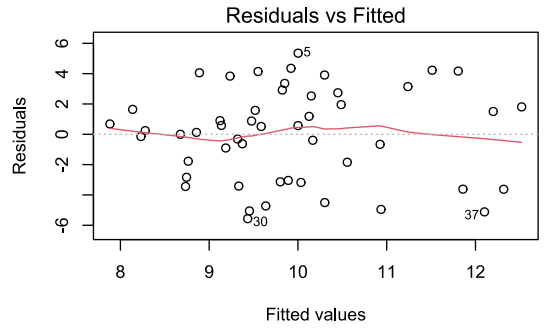
- each extra hour spent reading a week adds 1.26 years to reading age

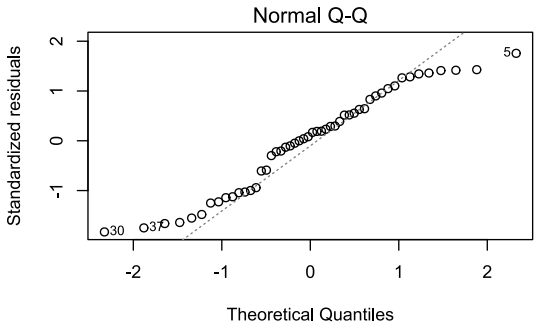
A Linear Model

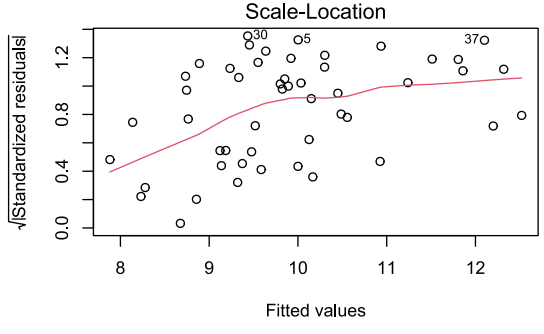
```
mod <- lm(R_AGE ~ hrs_wk, data=reading)
summary(mod)
```

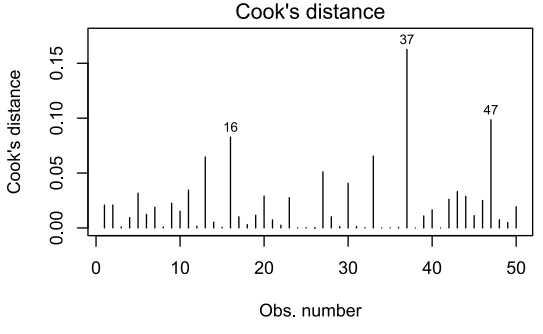
```
##
## Call:
## lm(formula = R_AGE ~ hrs_wk, data = reading)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.567 -2.991  0.378  2.385  5.351
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.65       2.47    1.48    0.146
## hrs_wk          1.26       0.49    2.57    0.013 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.08 on 48 degrees of freedom
## Multiple R-squared:  0.121,    Adjusted R-squared:  0.103
## F-statistic: 6.63 on 1 and 48 DF,  p-value: 0.0132
```

but...


Residuals vs Fitted
Residuals
Fitted values
5
30
37

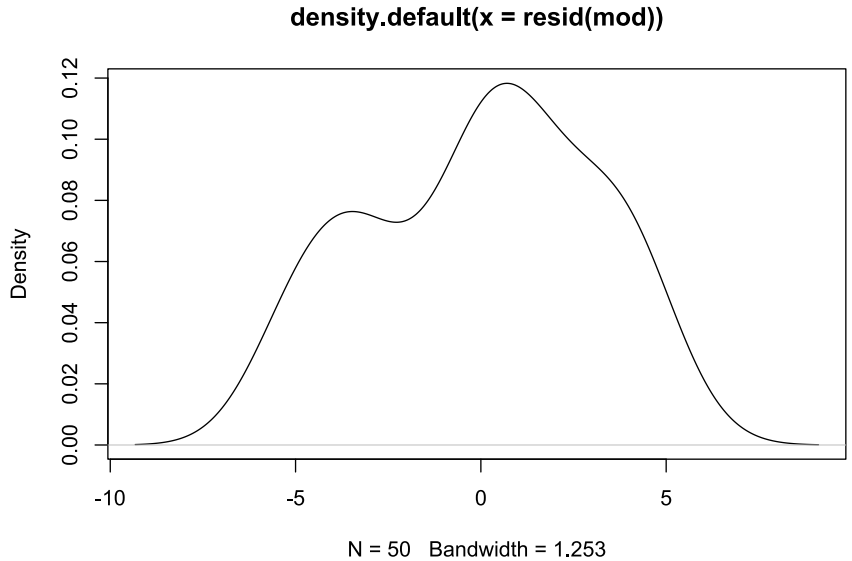

Normal Q-Q
Standardized residuals
Theoretical Quantiles
5
30
37


Scale-Location
√|Standardized residuals|
Fitted values
30
5
37


Cook's distance
Cook's distance
Obs. number
37
16
47

Assumptions Not Met!

- it seems that the assumptions aren't met for this model
- (another demonstration on the right)
- one reason for this can be because there's still systematic structure in the residuals
- i.e., *more than one thing* which can explain the variance



End of Part 3

Part 4

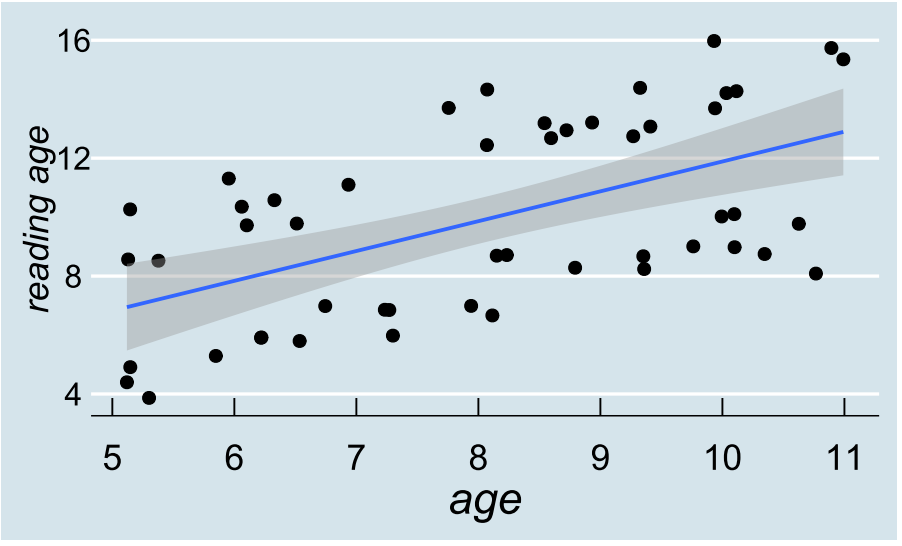
Multiple Regression

Adding Age into the Equation

- so far, have focused on effects of practice
- but presumably older children read better?

age	hrs_wk	method	R_AGE
10.115	4.971	phonics	14.272
9.940	4.677	phonics	13.692
6.060	4.619	phonics	10.353
7.941	4.053	word	6.988
8.233	5.474	word	8.713
6.219	4.038	word	5.908

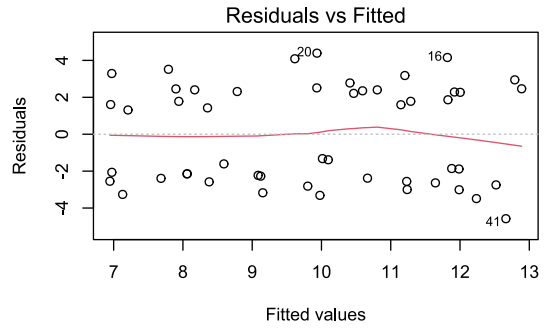
Another Model

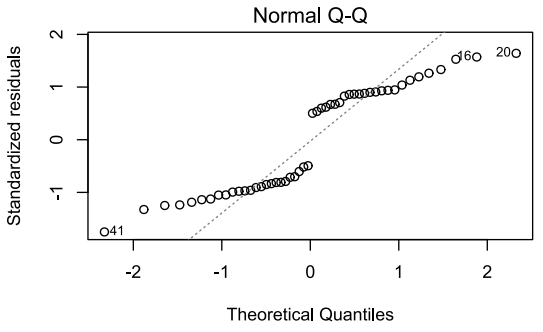


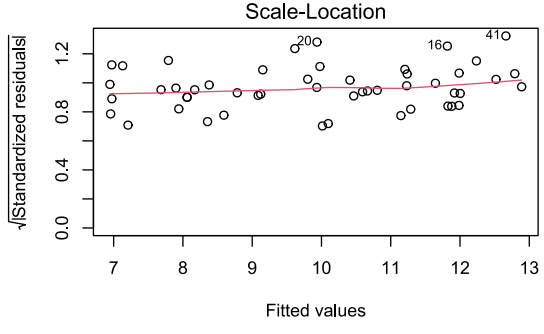
Another Model

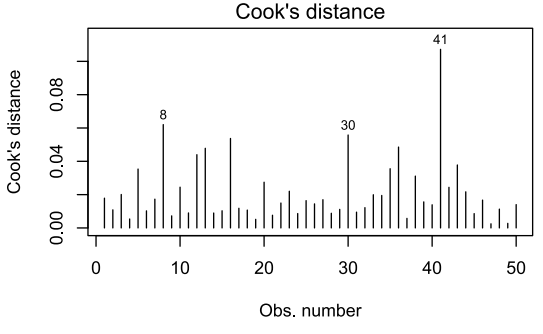
```
mod2 <- lm(R_AGE ~ age, data=reading)
summary(mod2)
```

```
##
## Call:
## lm(formula = R_AGE ~ age, data = reading)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.577 -2.509 -0.005  2.390  4.392
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.764      1.753    1.01     0.32
## age            1.012      0.212    4.76 0.000018 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.7 on 48 degrees of freedom
## Multiple R-squared:  0.321,    Adjusted R-squared:  0.307
## F-statistic: 22.7 on 1 and 48 DF,  p-value: 0.0000179
```


Residuals vs Fitted
Residuals
Fitted values
20
16
41


Normal Q-Q
Standardized residuals
Theoretical Quantiles
16
20
41


Scale-Location
√|Standardized residuals|
Fitted values
20
16
41


Cook's distance
Cook's distance
Obs. number
41
8
30

Two Models, No Answers

- we now have two models that don't map well to assumptions
- each suggests an effect
  - one of `age`
  - one of `hrs_wk`
- if we run them independently, the chances of a type 1 error are
  - $\frac{1}{20}$ (`mod`, including `hrs_wk`)
  - $\frac{1}{20}$ (`mod2`, including `age`)
- or $\frac{1}{10}$ overall

Two Models, No Answers

- we now have two models that don't map well to assumptions
- each suggests an effect
  - one of `age`
  - one of `hrs_wk`
- if we run them independently, the chances of a type 1 error are
  - $\frac{1}{20}$ (`mod`, including `hrs_wk`)
  - $\frac{1}{20}$ (`mod2`, including `age`)
- or $\frac{1}{10}$ overall

we need to test multiple predictors in *one* linear model

Model Equations Again

$$\text{outcome}_i = (\text{model})_i + \text{error}_i$$

$$y_i = b_0 \cdot 1 + b_1 \cdot x_i + \epsilon_i$$

Model Equations Again

$$\text{outcome}_i = (\text{model})_i + \text{error}_i$$

$$y_i = b_0 \cdot 1 + b_1 \cdot x_i + \epsilon_i$$

linear model with two predictors

$$y_i = b_0 \cdot 1 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} + \epsilon_i$$

$$\hat{y}_i = b_0 \cdot 1 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i}$$

Model Equations Again

$$\text{outcome}_i = (\text{model})_i + \text{error}_i$$

$$y_i = b_0 \cdot 1 + b_1 \cdot x_i + \epsilon_i$$

linear model with two predictors

$$y_i = b_0 \cdot 1 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} + \epsilon_i$$

$$\hat{y}_i = b_0 \cdot 1 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i}$$

```
y ~ 1 + x1 + x2
```

```
R_AGE ~ 1 + hrs_wk + age
```
or
```
R_AGE ~ hrs_wk + age
```
[1]

[1] we'll come back to why order can matter in a bit

Running a Multiple Regression

```
mod.m <- lm(R_AGE ~ age + hrs_wk, data=reading)
summary(mod.m)
```

```
##
## Call:
## lm(formula = R_AGE ~ age + hrs_wk, data = reading)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.385 -2.251  0.326  2.395  3.201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.423      2.472   -0.98    0.332
## age            0.938      0.206    4.55 0.000038 ***
## hrs_wk         0.964      0.418    2.31    0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.59 on 47 degrees of freedom
## Multiple R-squared:  0.39,    Adjusted R-squared:  0.364
## F-statistic:   15 on 2 and 47 DF,  p-value: 0.00000896
```

Running a Multiple Regression

```
## ...
## (Intercept)   -2.423      2.472   -0.98    0.332
## age            0.938      0.206    4.55 0.000038 ***
## hrs_wk         0.964      0.418    2.31    0.025 *
## ...
```

- there are *independent* effects of age and practice
    - reading age improves by 0.9378 for each year of age
    - reading age improves by 0.9636 for each weekly hour of practice
- note that the *intercept* (0 years old, 0 hours/week) is meaningless here

Running a Multiple Regression

```
## ...
## (Intercept)   -2.423      2.472   -0.98    0.332
## age            0.938      0.206    4.55 0.000038 ***
## hrs_wk         0.964      0.418    2.31    0.025 *
## ...
```

- there are *independent* effects of age and practice
    - reading age improves by 0.9378 for each year of age
    - reading age improves by 0.9636 for each weekly hour of practice
- note that the *intercept* (0 years old, 0 hours/week) is meaningless here
- important question: is this model *better* than a model based just on age?

Model Fit: R^2

```
## ...
## Residual standard error: 2.59 on 47 degrees of freedom
## Multiple R-squared:  0.39,    Adjusted R-squared:  0.364
## F-statistic:   15 on 2 and 47 DF,  p-value: 0.00000896
```

- in multiple regression, R^2 measures the fit of the entire model
  - sum of individual R^2s *if predictors not correlated*
- $R^2 = 0.3902$ looks better than the R^2 for mod2 (age as a predictor) of 0.3211
- but *any* predictor will improve R^2 (chance associations guarantee this)

```
mod2 <- lm(R_AGE ~ age, data=reading)
mod.2r <- update(mod2, ~ . + runif(50))
summary(mod.2r)
```

```
## ...
## Multiple R-squared:  0.361,    Adjusted R-squared:  0.334
## ...
```

Model Fit: F

```
## ...
## Residual standard error: 2.59 on 47 degrees of freedom
## Multiple R-squared:  0.39,    Adjusted R-squared:  0.364
## F-statistic:   15 on 2 and 47 DF,  p-value: 0.00000896
```

- in multiple regression, F tests the whether the model overall explains more variance than we would expect by chance.
- can be phrased as a model comparison:

```
null_mod <- lm(R_AGE ~ 1, data = reading)
mod.m <- lm(R_AGE ~ age + hrs_wk, data=reading)
anova(null_mod, mod.m)
```

```
## Analysis of Variance Table
##
## Model 1: R_AGE ~ 1
## Model 2: R_AGE ~ age + hrs_wk
##   Res.Df RSS Df Sum of Sq  F Pr(>F)
## 1     49 517
## 2     47 315  2       202 15  9e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing Models

- We can also use F ratios to compare models in terms of variance explained by each model:
  - Models must be "nested" - predictors of one model are a subset of predictors in the other.
  - Models must be fitted to the same data.

```
mod.r <- lm(R_AGE ~ age, data=reading)
mod.f <- lm(R_AGE ~ age + hrs_wk, data=reading)
anova(mod.r, mod.f)
```

```
## Analysis of Variance Table
##
## Model 1: R_AGE ~ age
## Model 2: R_AGE ~ age + hrs_wk
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1     48 351
## 2     47 315  1      35.7 5.33  0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Partitioning Variance

- Take one model, and examine variance explained by each predictor:

```
mod.f <- lm(R_AGE ~ age + hrs_wk, data=reading)
anova(mod.m)
```

```
## Analysis of Variance Table
##
## Response: R_AGE
##           Df Sum Sq Mean Sq F value    Pr(>F)
## age        1  166.0   166.0   24.75 0.0000092 ***
## hrs_wk     1   35.7    35.7    5.33     0.025 *
## Residuals 47  315.3     6.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Order Matters!

- `age` then `hrs_wk`

```
## Analysis of Variance Table
## ...
## Response: R_AGE
##           Df Sum Sq Mean Sq F value    Pr(>F)
## age        1  166.0   166.0   24.75 0.0000092 ***
## hrs_wk     1   35.7    35.7    5.33     0.025 *
## Residuals 47  315.3     6.7
## ---
```

- `hrs_wk` then `age`

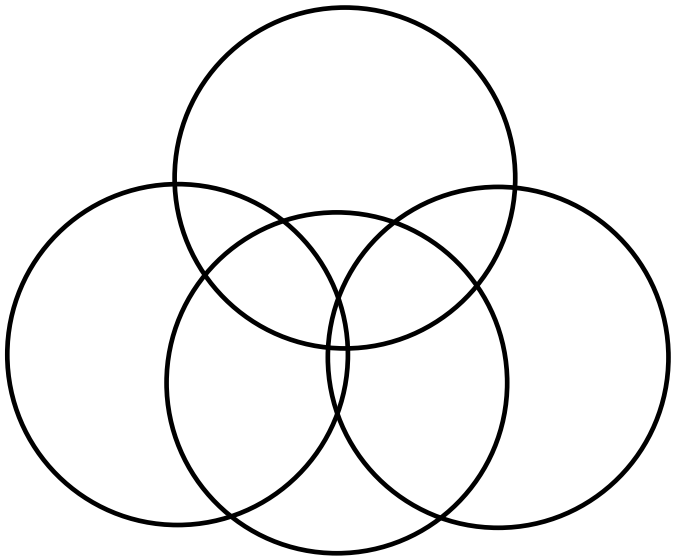
```
## Analysis of Variance Table
## ...
## Response: R_AGE
##           Df Sum Sq Mean Sq F value   Pr(>F)
## hrs_wk     1   62.7    62.7    9.35   0.0037 **
## age        1  139.0   139.0   20.72 0.000038 ***
## Residuals 47  315.3     6.7
## ---
```

Type 1 vs. Type 3 SS

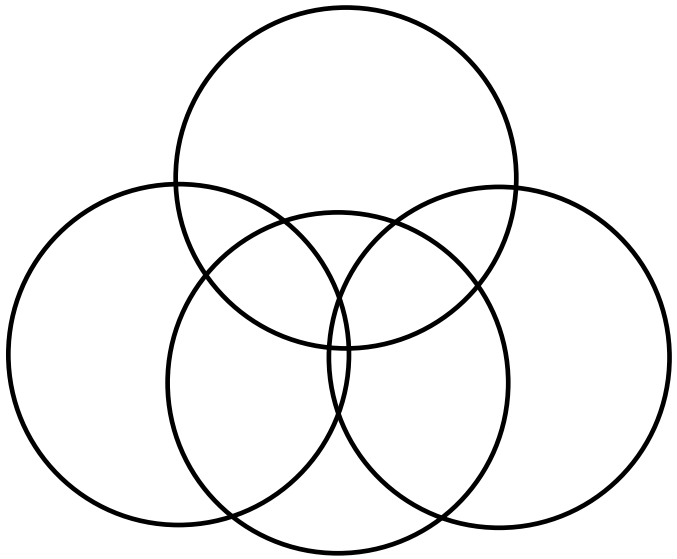
- order matters because R, by default, uses **Type 1** sums of squares for `anova()`
  - calculate each predictor's improvement to the model *in turn*
- compare to **Type 3** sums of squares
  - calculate each predictor's improvement to the model *taking all other predictors into account*
- huge debate about which is "better" (nobody likes Type 2)
- if using Type 1, *predictors should be entered into the model in a theoretically-motivated order*

Type 1 vs. Type 3 SS

type 1

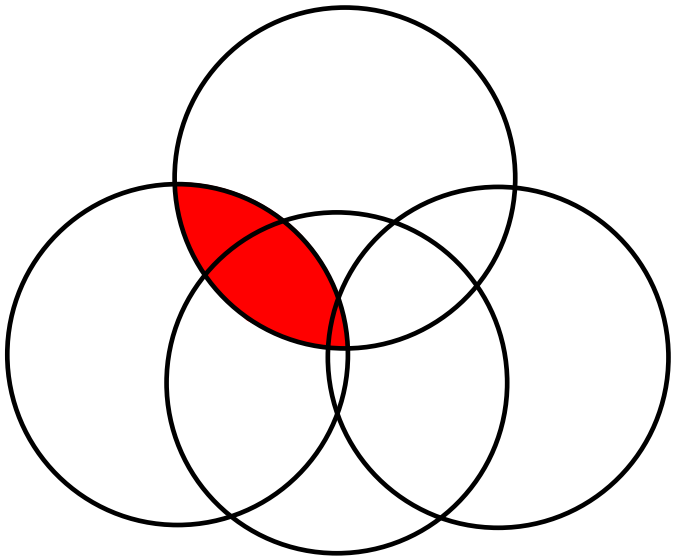


type 3

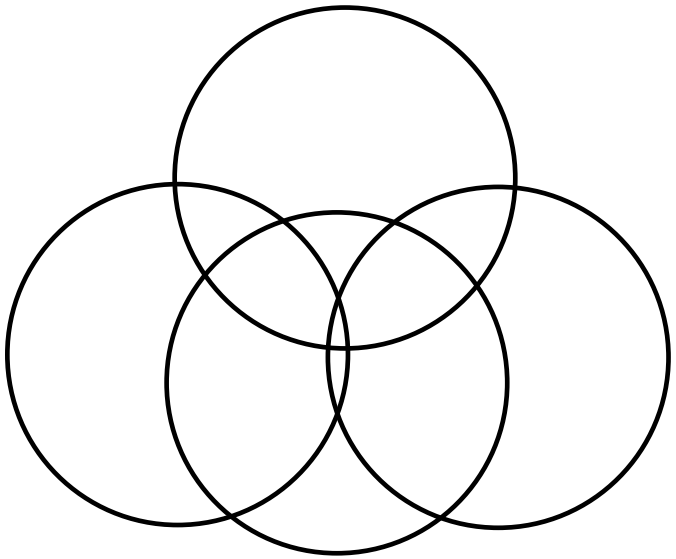


Type 1 vs. Type 3 SS

type 1

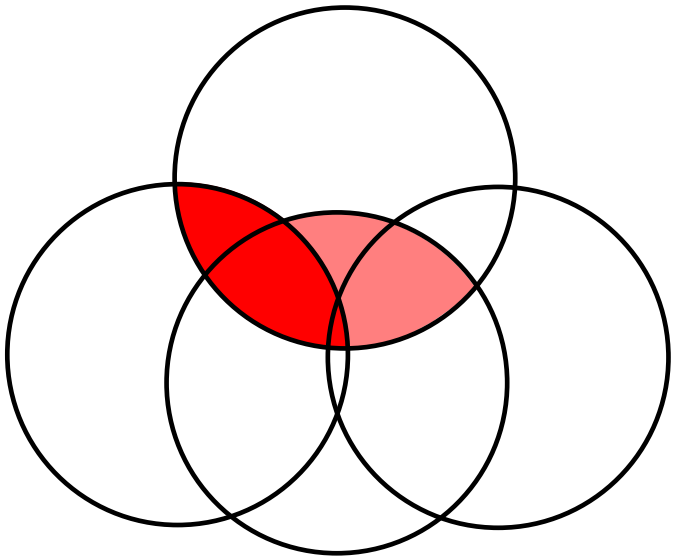


type 3

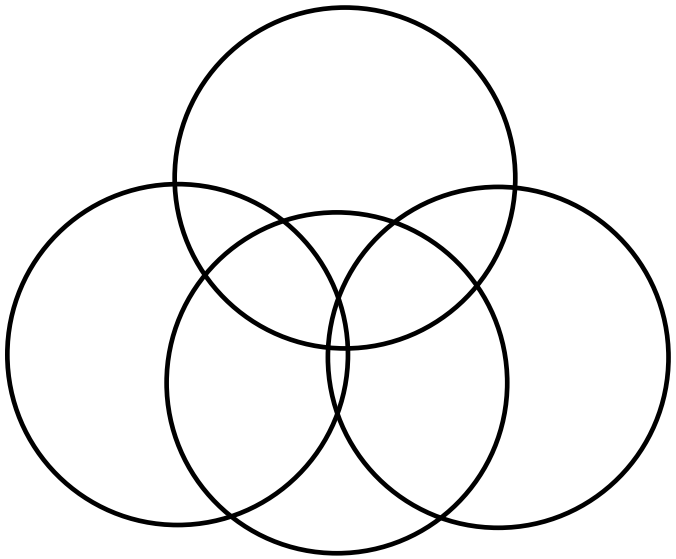


Type 1 vs. Type 3 SS

type 1

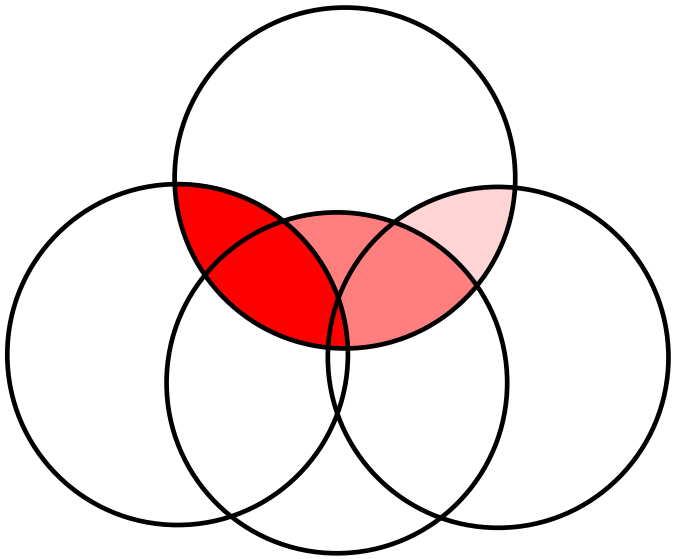


type 3

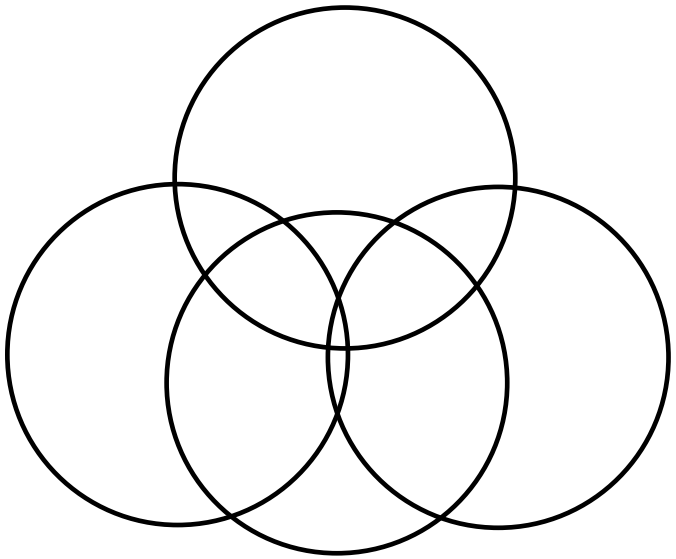


Type 1 vs. Type 3 SS

type 1

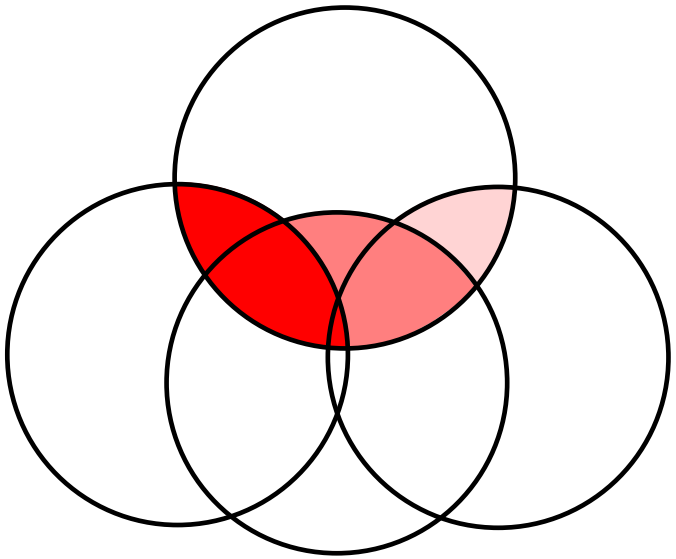


type 3

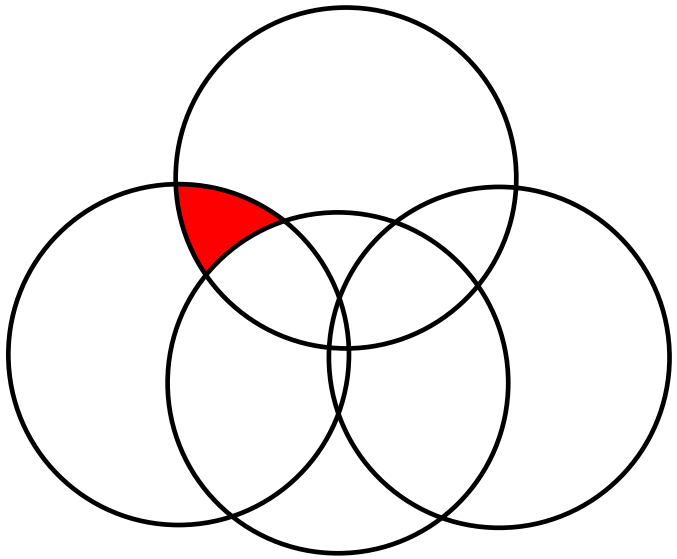


Type 1 vs. Type 3 SS

type 1

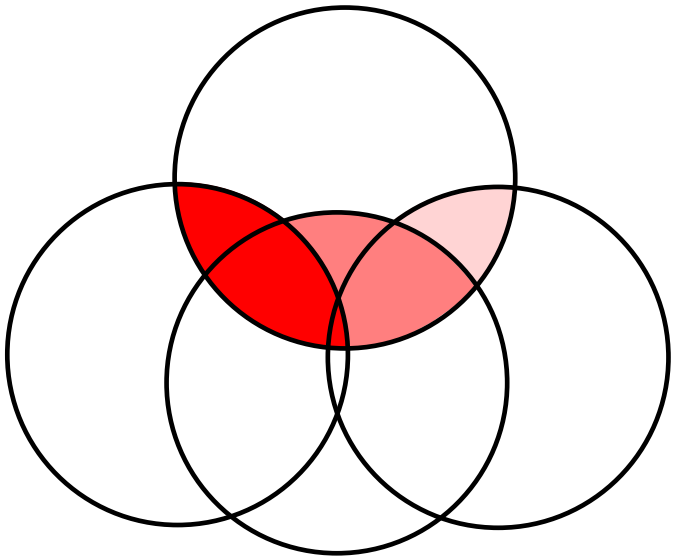


type 3

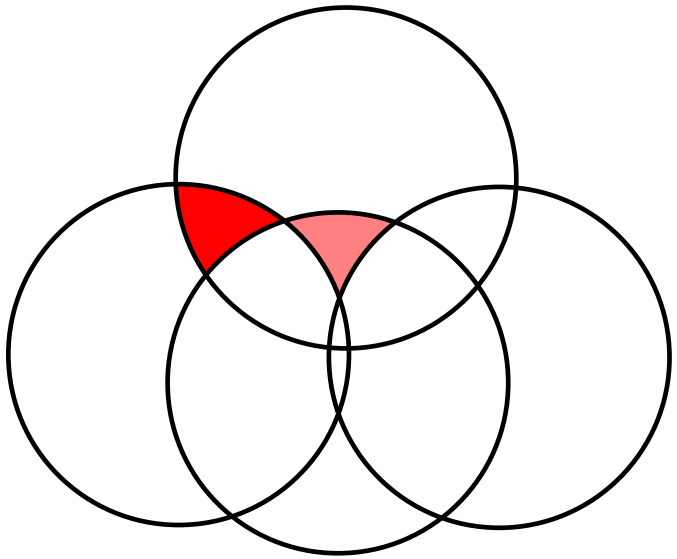


Type 1 vs. Type 3 SS

type 1

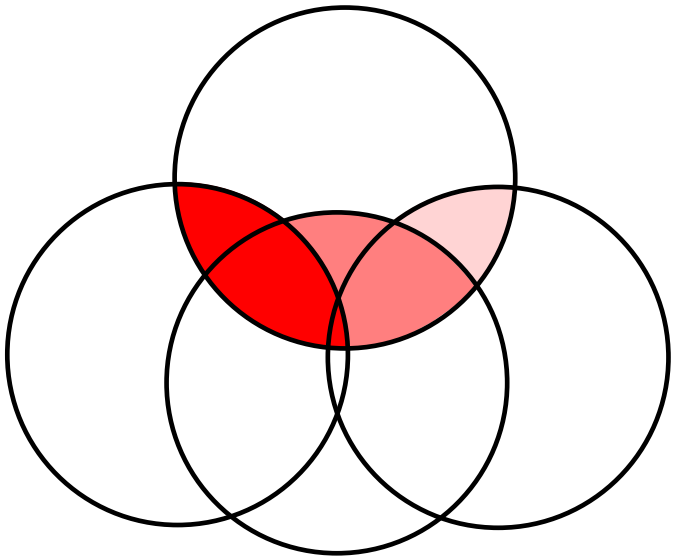


type 3

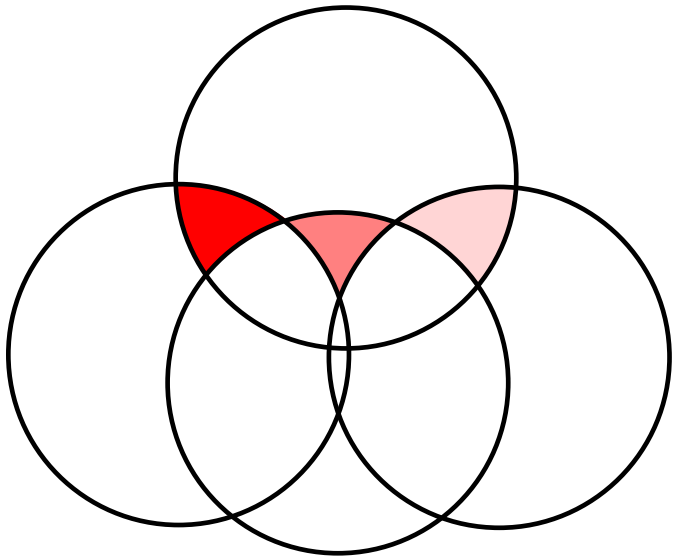


Type 1 vs. Type 3 SS

type 1



type 3



Type 1 vs. Type 3 SS

Type 1 - "Incremental" (Order Matters)

```
# age then hrs_wk:
anova(lm(R_AGE~age+hrs_wk,data=reading))
```

```
## Analysis of Variance Table
##
## Response: R_AGE
##           Df Sum Sq Mean Sq F value    Pr(>F)
## age        1  166.0   166.0   24.75 0.0000092 ***
## hrs_wk     1   35.7    35.7    5.33     0.025 *
## Residuals 47  315.3     6.7
## ---
```

```
# hrs_wk then age:
anova(lm(R_AGE~hrs_wk+age,data=reading))
```

```
## Analysis of Variance Table
##
## Response: R_AGE
##           Df Sum Sq Mean Sq F value   Pr(>F)
## hrs_wk     1   62.7    62.7    9.35   0.0037 **
## age        1  139.0   139.0   20.72 0.000038 ***
## Residuals 47  315.3     6.7
## ---
```

Type 3 - "Last one in"

```
mod.m <- lm(R_AGE~hrs_wk+age,data=reading)
drop1(mod.m, test="F")
```

```
## Single term deletions
##
## Model:
## R_AGE ~ hrs_wk + age
##        Df Sum of Sq RSS   AIC F value   Pr(>F)
## <none>              315  98.1
## hrs_wk  1      35.7 351 101.4    5.33    0.025 *
## age     1     139.0 454 114.3   20.72 0.000038 ***
## ---
```

So far..

What can we do with multiple regressions?

- Examine proportion of variance explained - R^2
- Test whether the model improves over chance - F test at the bottom of `summary(model)`
- Conduct comparisons between *nested* models
  - e.g. `lm(y ~ x1)` vs `lm(y ~ x1 + x2 + x3 + x4)`
  - using `anova(model1, model2)`
- Test the variance explained by each predictor in the model, either...
  - incrementally (in the order inputted into the model)
    `anova(model)`
  - after accounting for all other predictors
    `drop1(model, test = "F")`

Two Subtly Different Questions

> After accounting for `age`, *does* `hrs_wk` influence `R_AGE`?

```
mod.m <- lm(R_AGE~ age + hrs_wk, data=reading)
anova(mod.m)
```

```
## Analysis of Variance Table
##
## Response: R_AGE
##           Df Sum Sq Mean Sq F value    Pr(>F)
## age        1  166.0   166.0   24.75 0.0000092 ***
## hrs_wk     1   35.7    35.7    5.33     0.025 *
## Residuals 47  315.3     6.7
## ---
```

> After accounting for `age`, ***how*** *does* `hrs_wk` influence `R_AGE`?

```
mod.m <- lm(R_AGE~ age + hrs_wk, data=reading)
summary(mod.m)
```

```
## ...
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.423      2.472   -0.98    0.332
## age            0.938      0.206    4.55 0.000038 ***
## hrs_wk         0.964      0.418    2.31    0.025 *
## ---
## ...
```

The Two-Predictor Model

```
mod.m <- lm(R_AGE ~ age + hrs_wk, data=reading)
summary(mod.m)
```

```
##
## Call:
## lm(formula = R_AGE ~ age + hrs_wk, data = reading)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.385 -2.251  0.326  2.395  3.201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.423      2.472   -0.98    0.332
## age            0.938      0.206    4.55 0.000038 ***
## hrs_wk         0.964      0.418    2.31    0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.59 on 47 degrees of freedom
## Multiple R-squared:  0.39,    Adjusted R-squared:  0.364
## F-statistic:   15 on 2 and 47 DF,  p-value: 0.00000896
```

The Two-Predictor Model

End

Acknowledgements

- icons by Diego Lavecchia from the Noun Project