# STAT 516 Lec 11

Analysis of covariance (ANCOVA)

Karl Gregory

2026-11-16

# $VO_2$ max data from Kuehl (2000)

Change in $VO_2$ max of 12 males (after-minus-before) randomly assigned to two exercise programs (running, step aerobics). Ages recorded.

| *Group* | *Age* | *Change* | *Group* | *Age* | *Change* |
|---|---|---|---|---|---|
| Aerobic | 31 | 17.05 | Running | 23 | − 0.87 |
| | 23 | 4.96 | | 22 | − 10.74 |
| | 27 | 10.40 | | 22 | − 3.27 |
| | 28 | 11.05 | | 25 | − 1.97 |
| | 22 | 0.26 | | 27 | 7.50 |
| | 24 | 2.51 | | 20 | − 7.25 |
| Mean | 25.83 | 7.71 | | 23.17 | − 2.77 |
| Std. Err. | 1.40 | 2.55 | | 1.01 | 2.54 |

*Source*: D. Allen, Exercise Physiology, University of Arizona.

Which exercise program will lead to a greater average change in $VO_2$ max?

What role does age play?

# $VO_2$ max

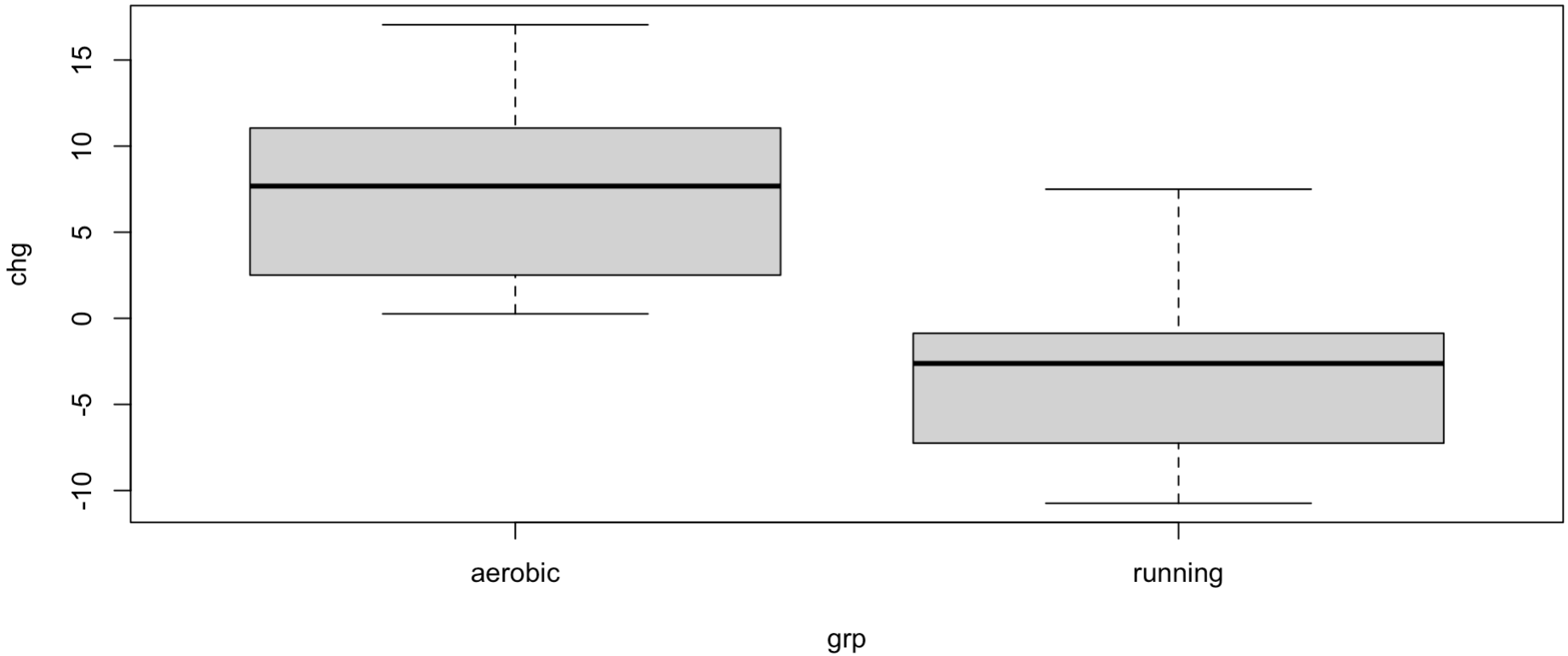
```r
vo2max <- data.frame(chg = c(17.05,4.96,10.40,11.05,0.26,2.51,
                             -0.87,-10.74,-3.27,-1.97,7.50,-7.25),
                     grp = as.factor(c(rep("aerobic",6),rep("running",6))),
                     age = c(31,23,27,28,22,24,23,22,22,25,27,20))

vo2max
```

```
      chg     grp age
1   17.05 aerobic  31
2    4.96 aerobic  23
3   10.40 aerobic  27
4   11.05 aerobic  28
5    0.26 aerobic  22
6    2.51 aerobic  24
7   -0.87 running  23
8  -10.74 running  22
9   -3.27 running  22
10  -1.97 running  25
11   7.50 running  27
12  -7.25 running  20
```

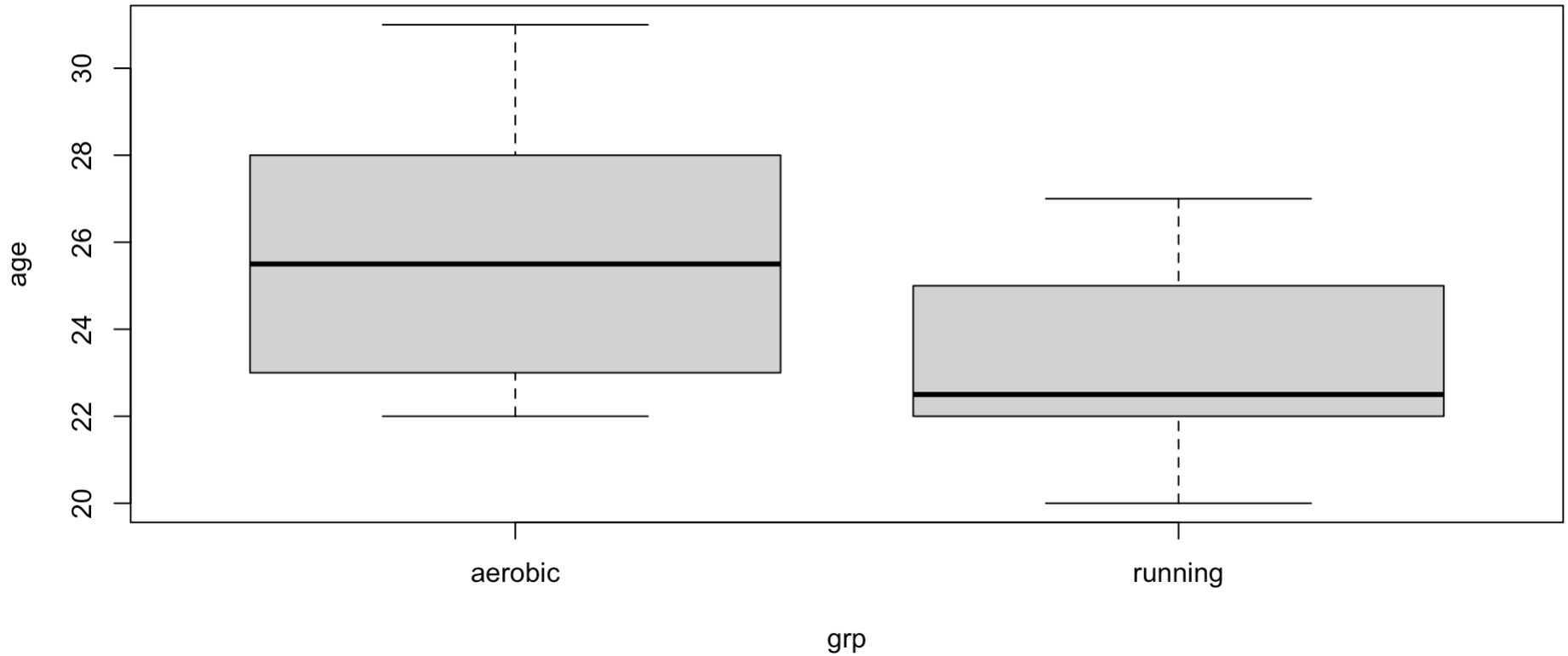
# $VO_2$ max

```
1 boxplot(chg ~ grp, data = vo2max)
```



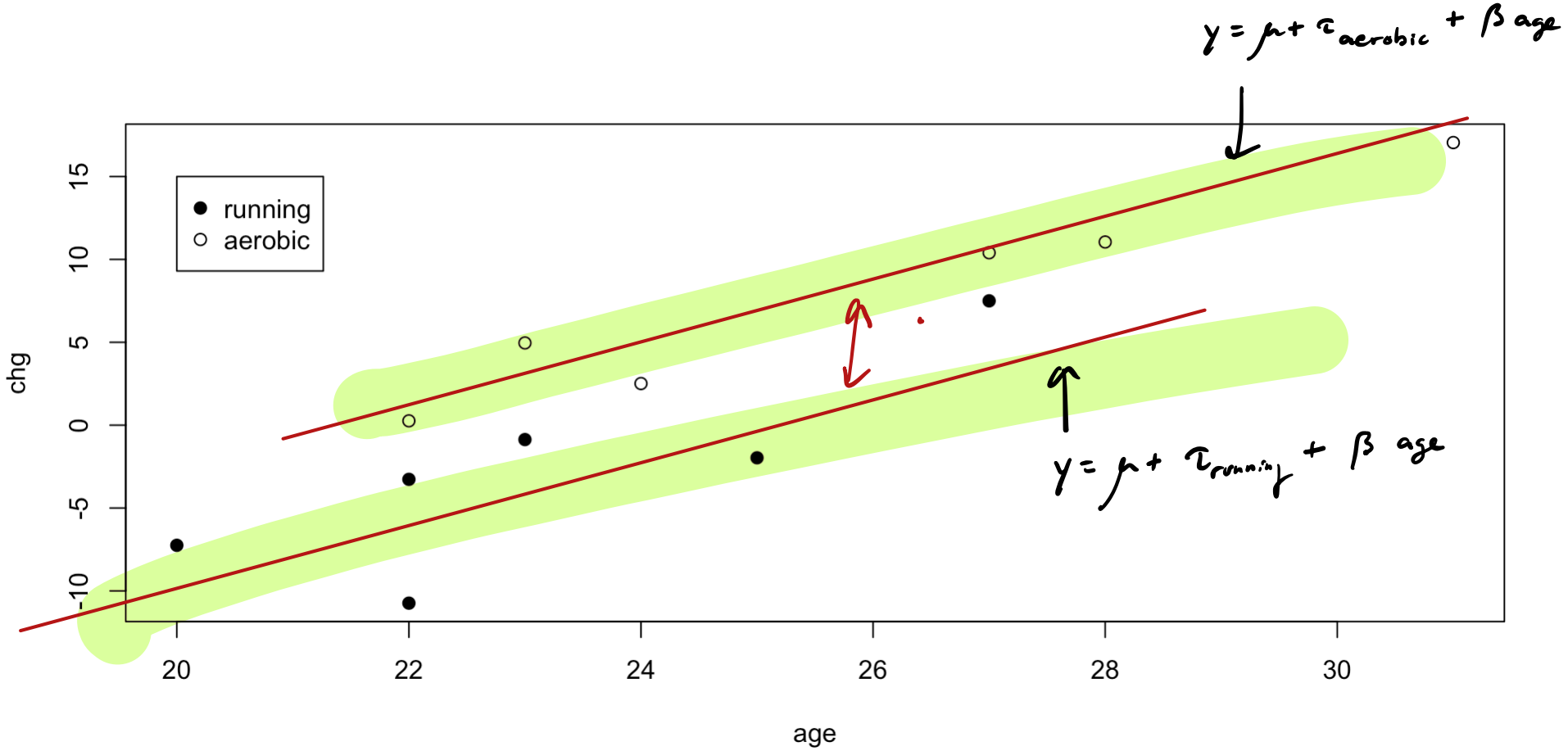
# $VO_2$ max

```
1 boxplot(age ~ grp, data = vo2max)
```



# $VO_2$ max

```
1 plot(chg ~ age, pch = ifelse(grp == "running",19,1), data = vo2max)
2 legend(x = 20,y = 15,legend = c("running","aerobic"), pch = c(19,1))
```




age of subject $j$ in group $i$

$$Y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij}$$

$i$ = treatment group

$j$ = subject within trt group

## Analysis of covariance

- Useful when EUs are not homogeneous.
- Measurement capturing EU inhomogeneity is called a covariate.
- Can isolate treatment effects even when EUs differ across treatment groups.

Eg. "age"

Analysis of Covariance model (ANCOVA)

slope parameter

covariate value of EU $j$ in group $i$

$$Y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij}$$

$i = 1, \ldots, a$ treatment group

$j = 1, \ldots, n_i$ EU in group $i$

# Single-slope analysis of covariance (ANCOVA) model

Assume

$$Y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij}$$

for $i = 1, \ldots, a$, $j = 1, \ldots, n_i$, where

- $Y_{ij}$ is the response of EU $j$ in treatment group $i$.
- $\mu$ is a baseline or overall mean.
- the $\tau_i$ are treatment effects.
- the $x_{ij}$ are covariate values measured on the EUs.
- $\beta$ is a slope coefficient expressing the effect of the covariate.
- the $\varepsilon_{ij}$ are independent $\text{Normal}(0, \sigma_\varepsilon^2)$ error terms.
- $n_1 + \cdots + n_a = N$. Unbalancedness not an issue.

Set $\mu_i = \mu + \tau_i$ for $i = 1, \ldots, a$.

# Goals in analysis of covariance

1. Estimate the parameters $\mu, \tau_1, \ldots, \tau_a$, and $\beta$.
2. Visualize the data.
3. Fit full and reduced models and collect error sums of squares.
4. Test whether the covariate has any effect.
5. Test for a treatment effect.
6. Adjust treatment group means for the inhomogeneity of the EUs.
7. Compare the adjusted group means.
8. Check whether model assumptions are satisfied.

Decompose variation in $Y_{ij}$ into sums of squares.

$H_0: \beta = 0$

$H_0: \tau_1 = \cdots = \tau_a$ no treatment effect

## Parameter constraints

- There are $a$ treatment groups and one covariate slope.
- The model has $a + 1$ parameters, which is one too-many.
- R will set $\tau_1 = 0$ so that it can estimate all the parameters.

Aerobic: $\hat{y} = -41.014 + 1.886 * age$

Running: $\hat{y} = -41.014 - 5.443 + 1.886 * age$

# $VO_2$ max

```
1 lm_out <- lm(chg ~ grp + age, data = vo2max)
2 summary(lm_out)
```

```
Call:
lm(formula = chg ~ grp + age, data = vo2max)

Residuals:
   Min     1Q Median     3Q    Max
 -5.77  -0.99   0.14   1.82   3.04

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -41.014      7.714   -5.32  0.00048 ***
grprunning    -5.443      1.796   -3.03  0.01425 *
age            1.886      0.295    6.39  0.00013 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.8 on 9 degrees of freedom
Multiple R-squared:  0.902, Adjusted R-squared:  0.88
F-statistic: 41.4 on 2 and 9 DF,  p-value: 2.89e-05
```

$\hat{\mu} = -41.014$

$\hat{\tau}_{running} = -5.443$

$\hat{\beta} = 1.886$

$\hat{\tau}_{aerobic} = 0$

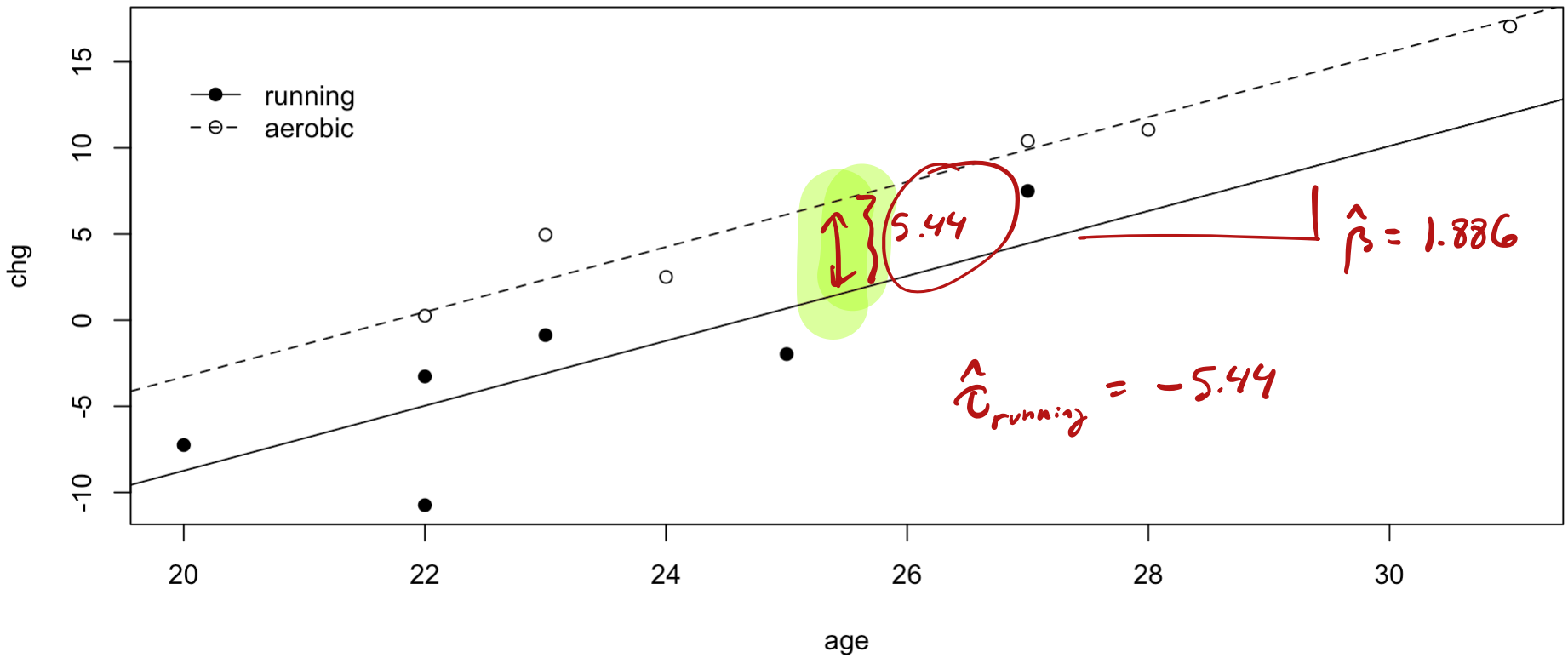
# $VO_2$ max

```
parms <- coef(lm_out)
parms
```

```
(Intercept)  grprunning         age
     -41.01       -5.44        1.89
```

# $VO_2$ max

```
1 plot(chg ~ age, pch = ifelse(grp == "running",19,1), data = vo2max)
2 abline(parms[1] + parms[2],parms[3])
3 abline(parms[1],parms[3],lty = 2)
4 legend(x = 20,y = 15,legend = c("running","aerobic"), pch = c(19,1), lty = c(1,2),bty = "n")
```



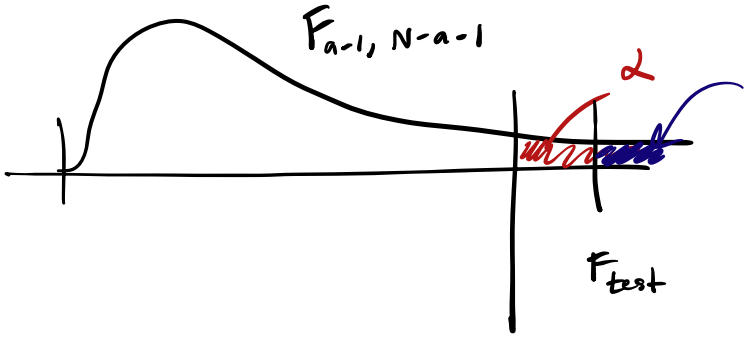
Full model: $Y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij} \rightarrow SS_{Error}(Full)$

$H_0: \tau_1 = \cdots = \tau_a$

Reduced model: $Y_{ij} = \mu + \beta x_{ij} + \varepsilon_{ij}$ $\quad SS_{Error}(Reduced)$

$$F_{test} = \frac{\left[SS_{Error}(Reduced) - SS_{Error}(Full)\right] / (a-1)}{SS_{Error}(Full) / (N-a-1)}$$

Reject $H_0$ if $F_{test} >$




p-value $= 1 - pf(F_{test}, a-1, N-a-1)$

$F_{a-1, N-a-1, \alpha} = qf(1-\alpha, a-1, N-a-1)$

$F_{2-1, 12-2-1, 0.05} = 5.12$

$F_{test} = 9.18$

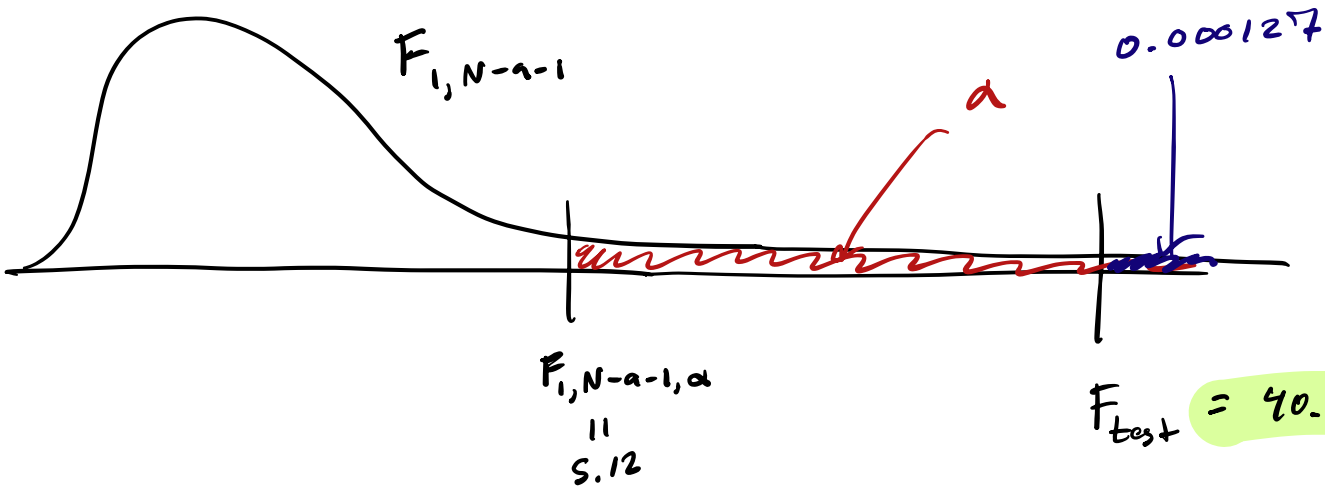
pval $= 0.0143$

$H_0: \beta = 0$

Reduced model: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ $\quad SS_{Error}(\text{Reduced})$

$$F_{test} = \frac{\left[SS_{Error}(\text{Reduced}) - SS_{Error}(\text{Full})\right] / 1}{SS_{Error}(\text{Full}) / (N - a - 1)}$$

$N = n_1 + \cdots + n_a$ = total # obs.



# Full and reduced models in ANCOVA

We construct test statistics by fitting full and reduced models:

- Full model: $Y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij}$
- Reduced model (no covariate effect): $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$
- Reduced model (no treatment effect): $Y_{ij} = \mu + \beta x_{ij} + \varepsilon_{ij}$

# Full- and reduced-model error sums of squares

| SS | Formula |
|---|---|
| $\mathrm{SS}_{\mathrm{Error}}(\mathrm{Full})$ | $\sum_{i=1}^{a}\sum_{j=1}^{n_{ij}}(Y_{ij}-(\hat{\mu}+\hat{\tau}_i+\hat{\beta}x_{ij}))^2$ |
| $\mathrm{SS}_{\mathrm{Error}}(\mathrm{No\ covariate})$ | $\sum_{i=1}^{a}\sum_{j=1}^{n_{ij}}(Y_{ij}-(\hat{\mu}+\hat{\tau}_i))^2$ |
| $\mathrm{SS}_{\mathrm{Error}}(\mathrm{No\ treatment})$ | $\sum_{i=1}^{a}\sum_{j=1}^{n_{ij}}(Y_{ij}-(\hat{\mu}+\hat{\beta}x_{ij}))^2$ |

Now define these difference in error sums of squares:

- $\mathrm{SS}_{\mathrm{Cov}} = \mathrm{SS}_{\mathrm{Error}}(\mathrm{No\ covariate}) - \mathrm{SS}_{\mathrm{Error}}(\mathrm{Full})$
- $\mathrm{SS}_{\mathrm{Trt}} = \mathrm{SS}_{\mathrm{Error}}(\mathrm{No\ treatment}) - \mathrm{SS}_{\mathrm{Error}}(\mathrm{Full})$

# ANCOVA table

| Source | Df | SS | MS | F value |
|---|---|---|---|---|
| Covariate | $1$ | $\mathrm{SS}_{\mathrm{Cov}}$ | $\mathrm{MS}_{\mathrm{Cov}}$ | $\mathrm{F}_{\mathrm{Cov}} = \mathrm{MS}_{\mathrm{Cov}} \,/\, \mathrm{MS}_{\mathrm{Error}}$ |
| Treatment | $a-1$ | $\mathrm{SS}_{\mathrm{Trt}}$ | $\mathrm{MS}_{\mathrm{Trt}}$ | $\mathrm{F}_{\mathrm{Trt}} = \mathrm{MS}_{\mathrm{Trt}} \,/\, \mathrm{MS}_{\mathrm{Error}}$ |
| Error | $N-a-1$ | $\mathrm{SS}_{\mathrm{Error}}$ | $\mathrm{MS}_{\mathrm{Error}}$ | |

1. Reject $H_0$: $\beta = 0$ at $\alpha$ if $\mathrm{F}_{\mathrm{Cov}} > F_{1,N-a-1,\alpha}$.
2. Reject $H_0$: $\mu_1 = \cdots = \mu_a$ at $\alpha$ if $\mathrm{F}_{\mathrm{Trt}} > F_{a-1,N-a-1,\alpha}$.

# $VO_2$ max

Using `anova()` on the `lm()` output gives the wrong SS (gives sequential).

Use `Anova()` from R package `car` on the `lm()` output.

```
1 library(car) # first time run install.packages("car")
2 # Use type = "II" or type = "III"
3 Anova(lm_out,type = "III")
```

```
Anova Table (Type III tests)

Response: chg
            Sum Sq Df F value  Pr(>F)
(Intercept)    221  1   28.27 0.00048 ***
grp             72  1    9.18 0.01425 *
age            319  1   40.78 0.00013 ***
Residuals       70  9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# $VO_2$ max

```r
lm_nocov <- lm(chg ~ grp, data = vo2max) # reduced model with no covariate
lm_notrt <- lm(chg ~ age, data = vo2max) # reduced model with no treatment
lm_full <- lm(chg ~ age + grp, data = vo2max) # full model

SSE_nocov <- sum(lm_nocov$resid^2)
SSE_notrt <- sum(lm_notrt$resid^2)
SSE_full <- sum(lm_full$resid^2)

SSCov <- SSE_nocov - SSE_full
SSTrt <- SSE_notrt - SSE_full
SSE <- SSE_full

a <- 2
N <- nrow(vo2max)

MSCov <- SSCov / 1
MSTrt <- SSTrt / (a-1)
MSE <- SSE / (N - a - 1)

FCov <- MSCov / MSE
FTrt <- MSTrt / MSE

pCov <- 1 - pf(FCov,1,N - a - 1)
```

# Treatment means in terms of parameter estimates

- We can write the treatment group means as

$$\bar{Y}_{i.} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}\bar{x}_{i.} \quad \text{for } i = 1, \ldots, a,$$

- We also define the covariate-adjusted means

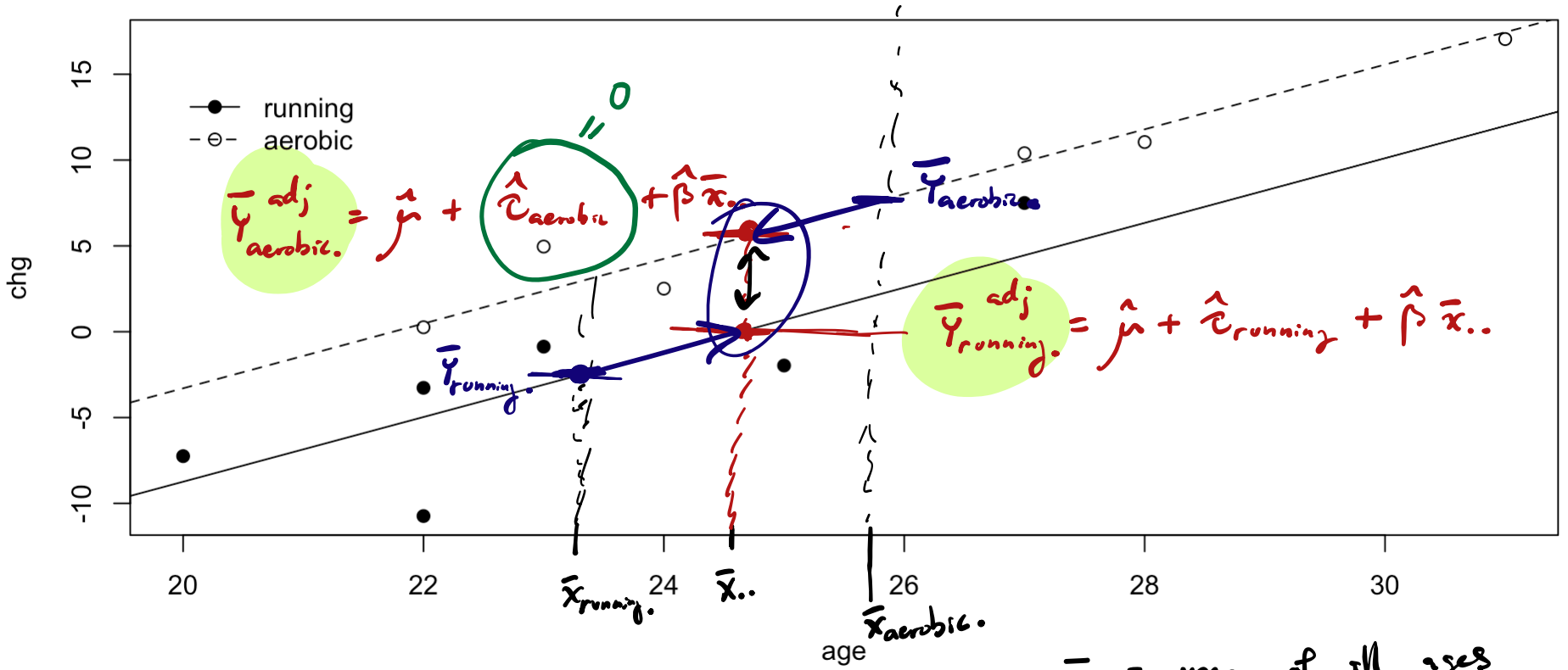
$$\bar{Y}_{i.}^{\text{adj}} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}\bar{x}_{..} \quad \text{for } i = 1, \ldots, a,$$

- We can equivalently write covariate-adjusted means as

$$\bar{Y}_{i.}^{\text{adj}} = \bar{Y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \quad \text{for } i = 1, \ldots, a.$$

# $VO_2$ max

```
1 plot(chg ~ age, pch = ifelse(grp == "running",19,1), data = vo2max)
2 abline(parms[1] + parms[2],parms[3])
3 abline(parms[1],parms[3],lty = 2)
4 legend(x = 20,y = 15,legend = c("running","aerobic"), pch = c(19,1), lty = c(1,2),bty = "n")
```




$x_{ij}$ is age of subj $j$ in group $i$.

$\bar{x}_{..}$ = mean of all ages

$\bar{x}_{aerobic.}$ = mean age in aerobic

$\bar{x}_{running.}$ = mean age in running

# Variances of covariate-adjusted means and their differences

| Contrast | Variance |
|---|---|
| $\bar{Y}_{i.}^{\text{adj}}$ | $\sigma^2 \left[ \frac{1}{n_i} + \frac{(\bar{x}_{i.} - \bar{x}_{..})^2}{E_{xx}} \right]$ |
| $\bar{Y}_{i.}^{\text{adj}} - \bar{Y}_{i'.}^{\text{adj}}$ | $\sigma^2 \left[ \frac{1}{n_i} + \frac{1}{n_{i'}} + \frac{(\bar{x}_{i.} - \bar{x}_{i'.})^2}{E_{xx}} \right]$ |

where $E_{xx} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2$.

# Some (unadjusted) CIs in the ANCOVA model

Define the true or population-level covariate-adjusted means as

$$\mu_i^{\text{adj}} = \mu + \tau_i + \beta\bar{x}_{..} \quad \text{for } i = 1, \ldots, a.$$

| **Target** | $(1-\alpha)100\%$ **confidence interval** |
|---|---|
| $\mu_i^{\text{adj}}$ | $\bar{Y}_{i.}^{\text{adj}} \pm t_{N-a-1,\alpha/2}\sqrt{\text{MS}_{\text{Error}}}\sqrt{\frac{1}{n_i} + \frac{(\bar{x}_{i.} - \bar{x}_{..})^2}{E_{xx}}}$ |
| $\mu_i^{\text{adj}} - \mu_{i'}^{\text{adj}}$ | $\bar{Y}_{i.}^{\text{adj}} - \bar{Y}_{i'.}^{\text{adj}} \pm t_{N-a-1,\alpha/2}\sqrt{\text{MS}_{\text{Error}}}\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}} + \frac{(\bar{x}_{i.} - \bar{x}_{}}{E_{xx}}}$ |

# $VO_2$ max

```r
y1. <- mean(vo2max$chg[vo2max$grp == "aerobic"])
y2. <- mean(vo2max$chg[vo2max$grp == "running"])

x1. <- mean(vo2max$age[vo2max$grp == "aerobic"])
x2. <- mean(vo2max$age[vo2max$grp == "running"])
x.. <- mean(vo2max$age)

bhat <- coef(lm_out)[3]

y1.adj <- y1. - bhat * (x1. - x..)
y2.adj <- y2. - bhat * (x2. - x..)
```

# $VO_2$ max

```r
n1 <- 6
n2 <- 6
alpha <- 0.05

tval <- qt(1-alpha/2,N-a-1)
age1 <- vo2max$age[vo2max$grp == "aerobic"]
age2 <- vo2max$age[vo2max$grp == "running"]
Ex1 <- sum((age1 - mean(age1))^2)
Ex2 <- sum((age2 - mean(age2))^2)
Exx <- Ex1 + Ex2

se1 <- sqrt(MSE) * sqrt(1/n1 + (x1. - x..)^2 / Exx)
lo1 <- y1.adj - tval * se1
up1 <- y1.adj + tval * se1

se2 <- sqrt(MSE) * sqrt(1/n2 + (x2. - x..)^2 / Exx)
lo2 <- y2.adj - tval * se2
up2 <- y2.adj + tval * se2

se12 <- sqrt(MSE) * sqrt(1/n1 + 1/n2 + (x1. - x2.)^2 / Exx)
lo12 <- y1.adj - y2.adj - tval * se12
up12 <- y1.adj - y2.adj + tval * se12
```

# $VO_2$ max

Mean change in $VO_2$ max in aerobic and running groups

| | Unadjusted | Age-adjusted(CI) |
|---|---|---|
| Aerobic | $\bar{Y}_{aerobic.}$ = 7.71 | 5.19 (2.46,7.92) |
| Running | $\bar{Y}_{running.}$ = -2.77 | -0.25 (-2.98,2.48) |
| Aerobic - Running | 10.47 | 5.44 (1.38,9.51) |

$\bar{Y}^{adj}_{aerobic.}$

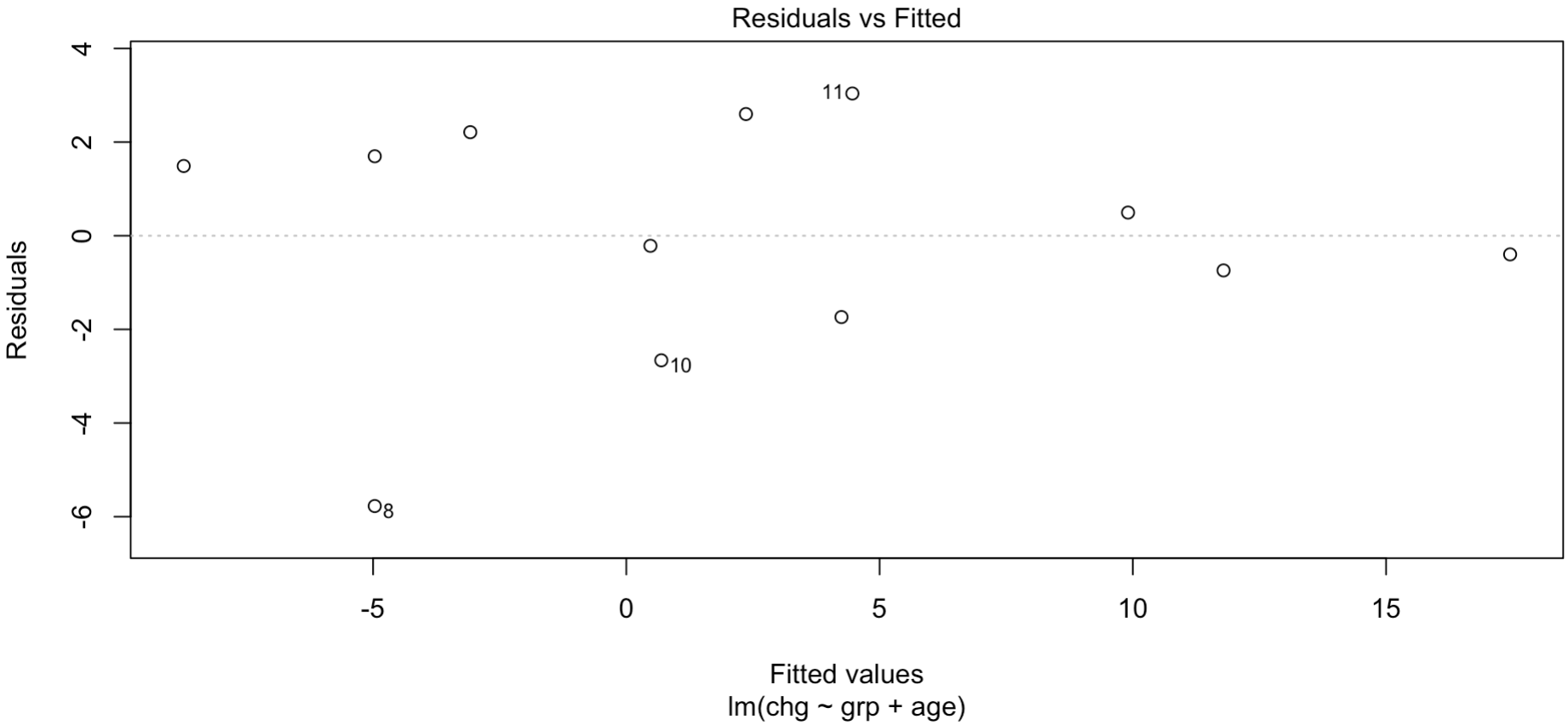
same as $\hat{\tau}_{running}$

$\bar{Y}^{adj}_{running.}$

# $VO_2$ max

Check constant variance assumption with residual vs fitted values plot:

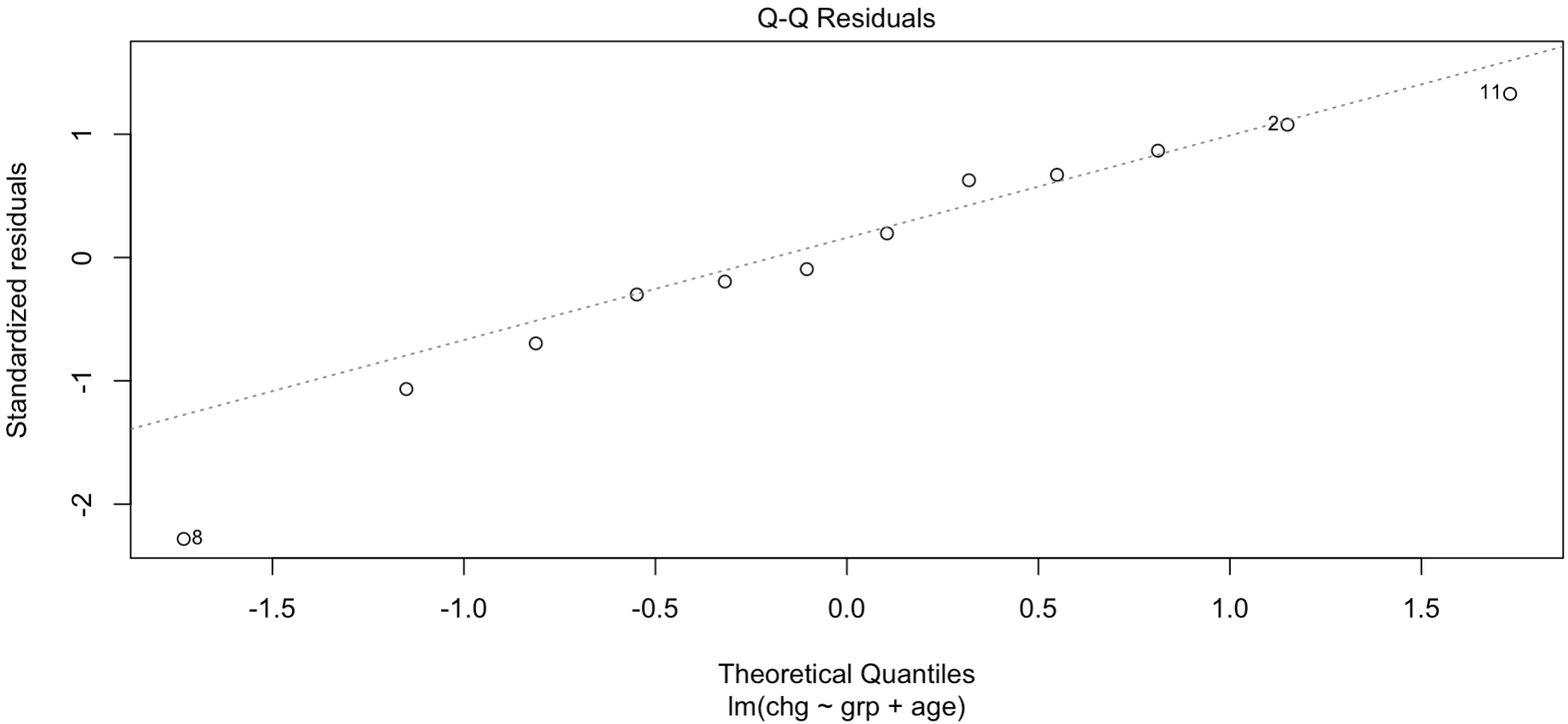
```
1 plot(lm_out, which = 1,add.smooth = F)
```



# $VO_2$ max

Check normality of error term with a normal QQ plot of the residuals:

```
1 plot(lm_out, which = 2)
```



# Soybean data from Dr. Longnecker's notes

Soybean plants assigned to three greenhouse conditions:

- Supplemental lighting (SL)
- Partial shading (PS)
- Control (C).

$a = 3$

The response was seed yield. The pre-treatment height of each plant was also recorded.

$x_{ij}$

| Yield | Height | TRT | Yield | Height | TRT | Yield | Height | TRT | Yield | Height | TRT | Yield | Height | TRT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12.2 | 45 | C | 12.4 | 52 | C | 11.9 | 42 | C | 11.3 | 35 | C | 11.8 | 40 | C |
| 12.1 | 48 | C | 13.1 | 60 | C | 12.7 | 61 | C | 12.4 | 50 | C | 11.4 | 33 | C |
| 12.3 | 48 | C | 12.2 | 51 | C | 12.6 | 56 | C | 13.2 | 65 | C | 12.3 | 51 | C |
| 16.6 | 63 | SL | 15.8 | 50 | SL | 16.5 | 63 | SL | 15.0 | 33 | SL | 15.4 | 38 | SL |
| 15.6 | 45 | SL | 15.8 | 50 | SL | 15.8 | 48 | SL | 16.0 | 50 | SL | 15.8 | 49 | SL |
| 15.0 | 35 | SL | 16.2 | 50 | SL | 16.7 | 62 | SL | 15.8 | 49 | SL | 15.9 | 52 | SL |
| 9.5 | 52 | PS | 9.5 | 54 | PS | 9.6 | 58 | PS | 8.8 | 45 | PS | 9.5 | 57 | PS |
| 9.8 | 62 | PS | 9.1 | 52 | PS | 10.3 | 67 | PS | 9.5 | 55 | PS | 8.5 | 40 | PS |
| 8.6 | 41 | PS | 10.4 | 67 | PS | 9.4 | 55 | PS | 10.2 | 66 | PS | 9.3 | 56 | PS |

Do the greenhouse conditions effect the seed yield?

What is the role of plant height (proxy for plant vigor)?

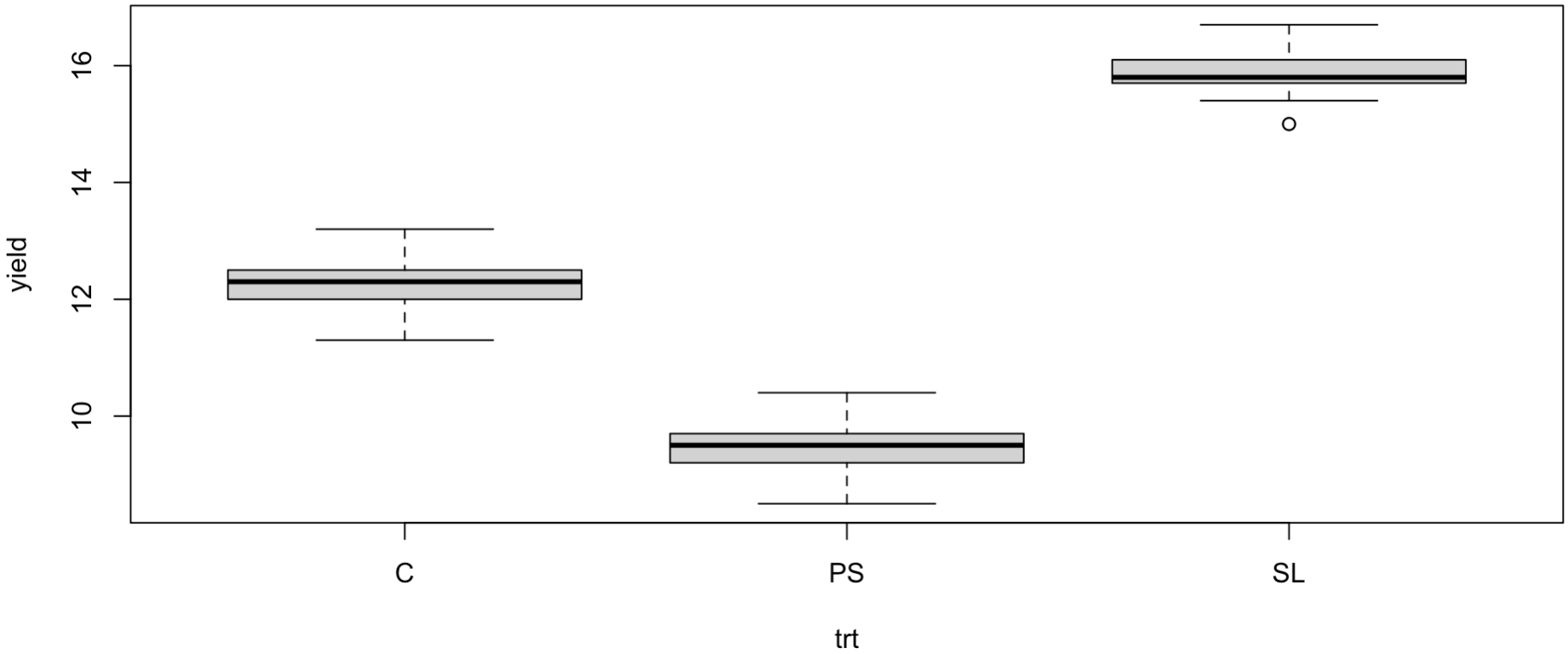
# Soybean data

```r
soybean <- data.frame(yield = c(12.2,12.1,12.3,16.6,15.6,15.0,9.5,9.8,8.6,
                                12.4,13.1,12.3,15.8,15.8,16.2,9.5,9.1,10.4,
                                11.9,12.7,12.6,16.5,15.8,16.7,9.6,10.3,9.4,
                                11.3,12.4,13.2,15.0,16.0,15.8,8.8,9.5,10.2,
                                11.8,11.4,12.3,15.4,15.8,15.9,9.5,8.5,9.3),
                      trt = as.factor(rep(c(rep("C",3),rep("SL",3),rep("PS",3)),5)),
                      height = c(45,48,48,63,45,35,52,62,41,
                                 52,60,51,50,50,50,54,52,67,
                                 42,61,56,63,48,62,58,67,55,
                                 35,50,65,33,50,49,45,55,66,
                                 40,33,51,38,49,52,57,40,56))
head(soybean,n = 12)
```

```
   yield trt height
1   12.2   C     45
2   12.1   C     48
3   12.3   C     48
4   16.6  SL     63
5   15.6  SL     45
6   15.0  SL     35
7    9.5  PS     52
8    9.8  PS     62
9    8.6  PS     41
10  12.4   C     52
11  13.1   C     60
12  12.3   C     51
```

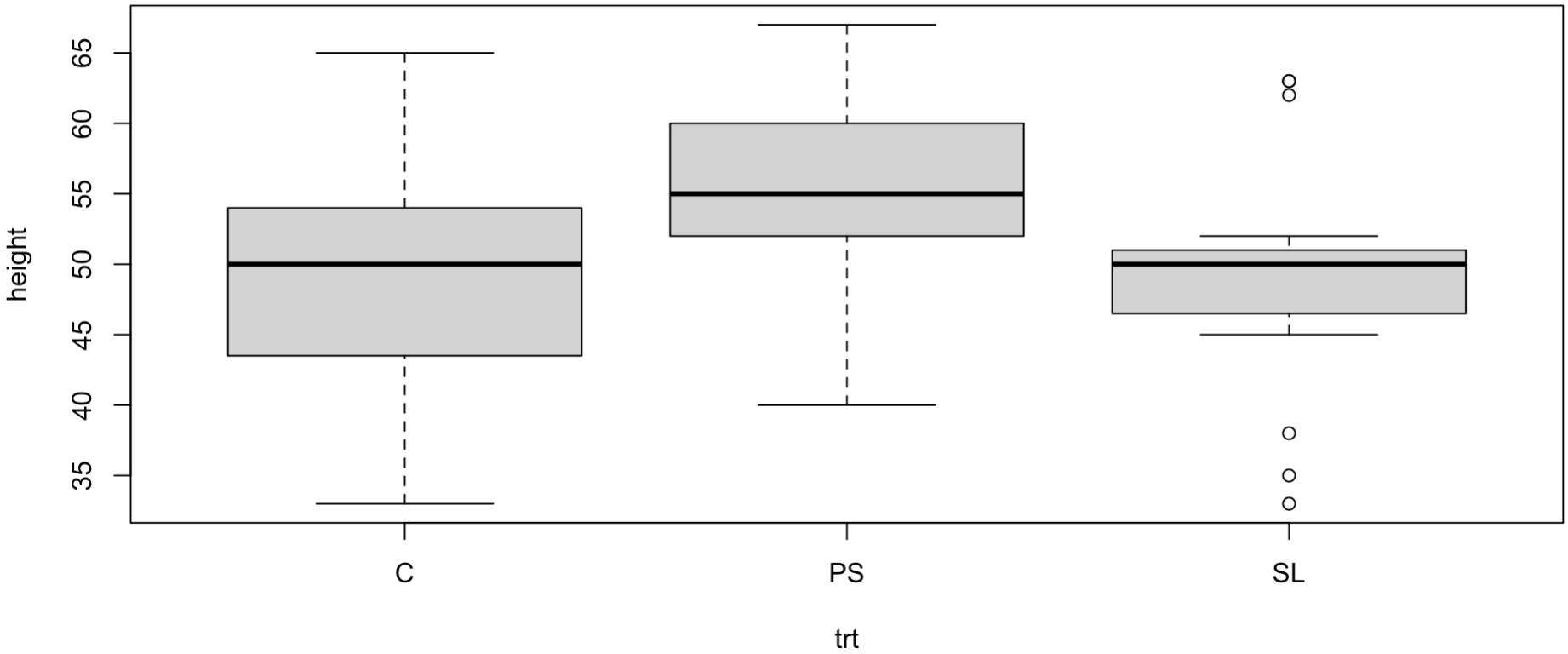
# Soybean data

```
1 boxplot(yield~trt, data = soybean)
```



# Soybean data

```
1 boxplot(height~trt, data = soybean)
```

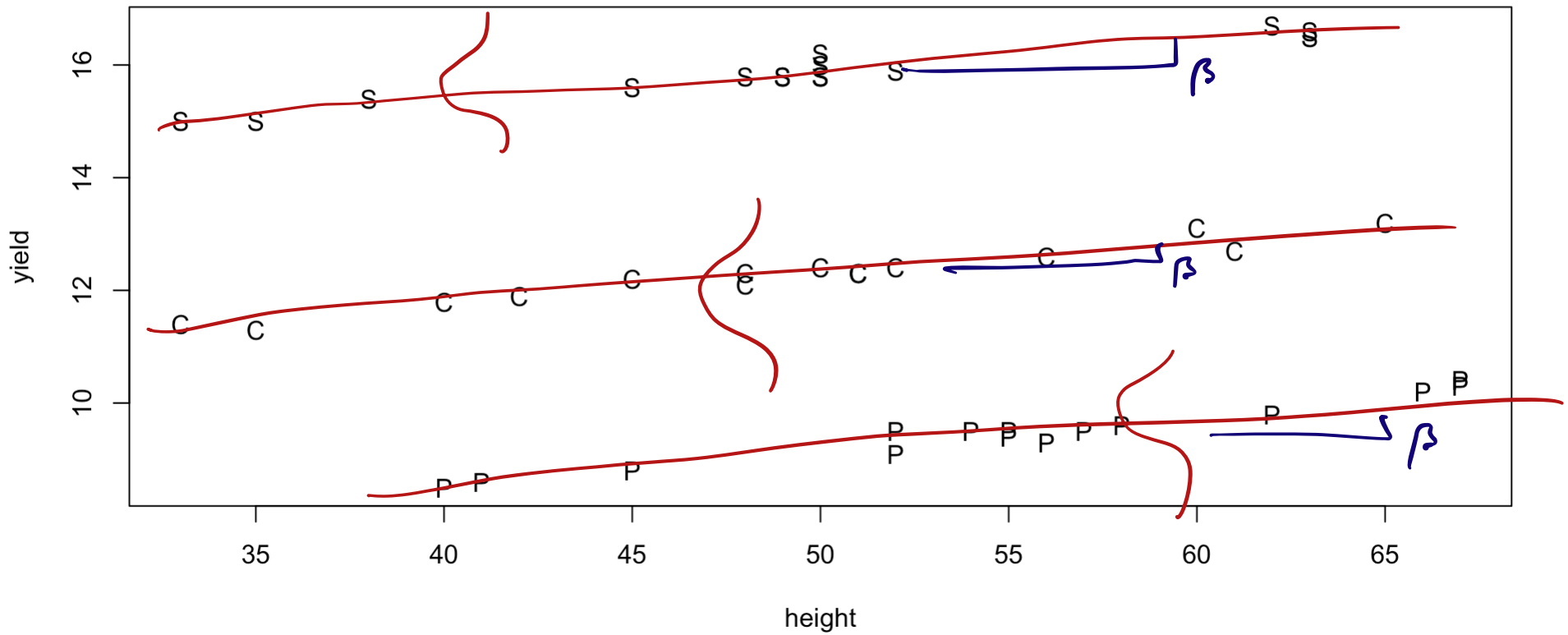


# Soybean data

$$Y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij}$$

$x_{ij}$: height

```
1 plot(yield ~ height, pch = as.character(trt), data = soybean)
```



# ANCOVA allowing different slopes in each group

Assume

$$Y_{ij} = \mu + \tau_i + (\beta + (\tau\beta)_i)x_{ij} + \varepsilon_{ij}$$

for $i = 1, \ldots, a$, $j = 1, \ldots, n_i$, where

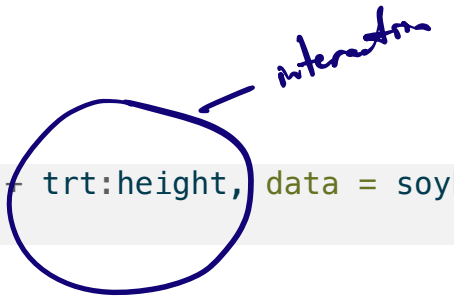
- $Y_{ij}$ is the response of EU $j$ in treatment group $i$.
- $\mu$ is a baseline or overall mean.
- the $\tau_i$ are treatment effects.
- the $x_{ij}$ are covariate values measured on the EUs.
- $\beta$ is a slope coefficient expressing the effect of the covariate.
- the $(\tau\beta)_i$ allow interaction between the treatment and the covariate.
- the $\varepsilon_{ij}$ are independent $\mathrm{Normal}(0, \sigma_\varepsilon^2)$ error terms.
- $n_1 + \cdots + n_a = N$. Unbalancedness not an issue.

Set $\mu_i = \mu + \tau_i$ and $\beta_i = \beta + (\tau\beta)_i$ for $i = 1, \ldots, a$.

## Parameter constraints in multiple slopes model

- There are $a$ treatment groups and $a$ covariate slopes.
- The model has $2(a+1)$ parameters, which is two too-many.
- R will set $\tau_1 = 0$ and $(\tau\beta)_1 = 0$ to make all parameters estimable.

# Soybean data




```
1 lm_out <- lm(yield ~ trt + height + trt:height, data = soybean)
2 summary(lm_out)
```

```
Call:
lm(formula = yield ~ trt + height + trt:height, data = soybean)

Residuals:
     Min       1Q   Median       3Q      Max
-0.23473 -0.08875 -0.00395  0.05764  0.29332

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.50057    0.18193   52.22  < 2e-16 ***
trtPS        -3.63959    0.28553  -12.75  1.7e-15 ***
trtSL         3.71305    0.25858   14.36  < 2e-16 ***
height        0.05630    0.00364   15.45  < 2e-16 ***
trtPS:height  0.00910    0.00537    1.69    0.098 .
trtSL:height -0.00244    0.00518   -0.47    0.641
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

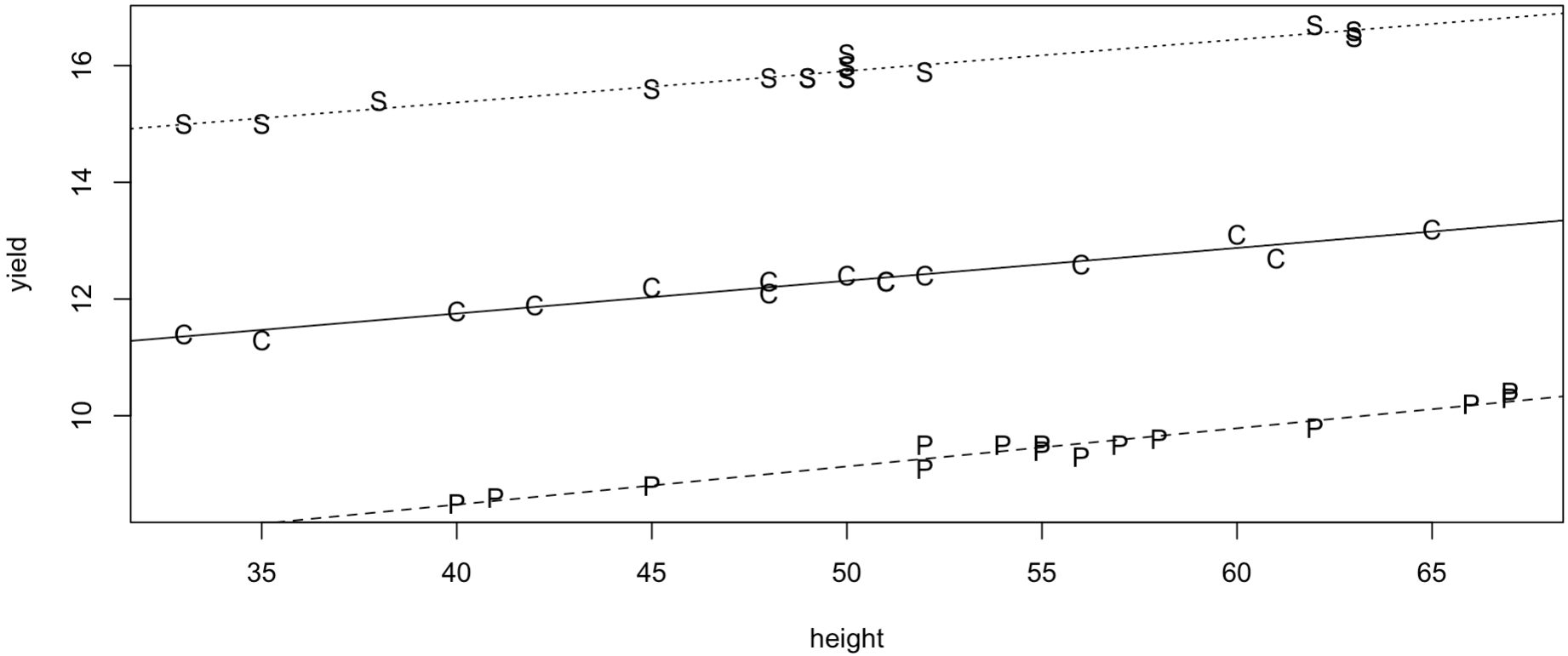
# Soybean data

```
parms <- coef(lm_out)
parms
```

```
(Intercept)        trtPS        trtSL       height trtPS:height trtSL:height
    9.50057     -3.63959      3.71305      0.05630      0.00910     -0.00244
```

# Soybean data

```
plot(yield ~ height, pch = as.character(trt), data = soybean)
abline(parms[1],parms[4],lty = 1)
abline(parms[1] + parms[2],parms[4] + parms[5],lty = 2)
abline(parms[1] + parms[3],parms[4] + parms[6],lty = 3)
```



# An F-test for equal slopes

Define the error sums of squares:

| SS | Formula |
|---|---|
| $\mathrm{SS}_{\mathrm{Error}}(\mathrm{Full})$ | $\sum_{i=1}^{a}\sum_{j=1}^{n_{ij}}(Y_{ij}-(\hat{\mu}+\hat{\tau}_i+(\hat{\beta}+(\hat{\tau\beta})_i)x_{ij}))^2$ |
| $\mathrm{SS}_{\mathrm{Error}}(\text{Equal slopes})$ | $\sum_{i=1}^{a}\sum_{j=1}^{n_{ij}}(Y_{ij}-(\hat{\mu}+\hat{\tau}_i+\hat{\beta}x_{ij}))^2$ |

residuals of full model

residuals of reduced model

Now set

$$F_{test} = \mathrm{F}_{\mathrm{T}\times\mathrm{C}} = \frac{[\mathrm{SS}_{\mathrm{Error}}(\text{Equal slopes})-\mathrm{SS}_{\mathrm{Error}}(\mathrm{Full})]/(a-1)}{\mathrm{SS}_{\mathrm{Error}}(\mathrm{Full})/(N-2a)}$$

Reject $H_0$: *Equal slopes* at $\alpha$ if $\mathrm{F}_{\mathrm{T}\times\mathrm{C}} > F_{a-1,N-2a,\alpha}$.

# Soybean data

```r
lm_eqslp <- lm(yield ~ trt + height, data = soybean) # equal slopes model
lm_full <- lm(yield ~ trt + height + trt:height, data = soybean) # full model

SSE_eqslp <- sum(lm_eqslp$resid^2)
SSE_full <- sum(lm_full$resid^2)

a <- 3
N <- nrow(soybean)

FTrtCov <- ((SSE_eqslp - SSE_full) / (a-1)) / (SSE_full / (N - 2*a))
pTrtCov <- 1 - pf(FTrtCov,a-1,N - 2*a)
```

# Soybean data

Same as interaction p value from `anova()` on the `lm()` output.

```
1 anova(lm_out)
```

```
Analysis of Variance Table

Response: yield
           Df Sum Sq Mean Sq F value Pr(>F)
trt         2  308.1   154.1 9770.80 <2e-16 ***
height      1   11.4    11.4  722.29 <2e-16 ***
trt:height  2    0.1     0.0    2.49  0.096 .
Residuals  39    0.6     0.0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

correct p-value for interaction

FTR $H_0$: equal slopes

$(\alpha\beta)_1 = \cdots = (\alpha\beta)_a$

$$Y_{ij} = \mu + \tau_i + (\beta + (\alpha\beta)_i) x_{ij} + \varepsilon_{ij}$$

$\mu_i$ $\beta_i$

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

$H_0: \beta_1 = \cdots = \beta_a$

## When slopes are unequal

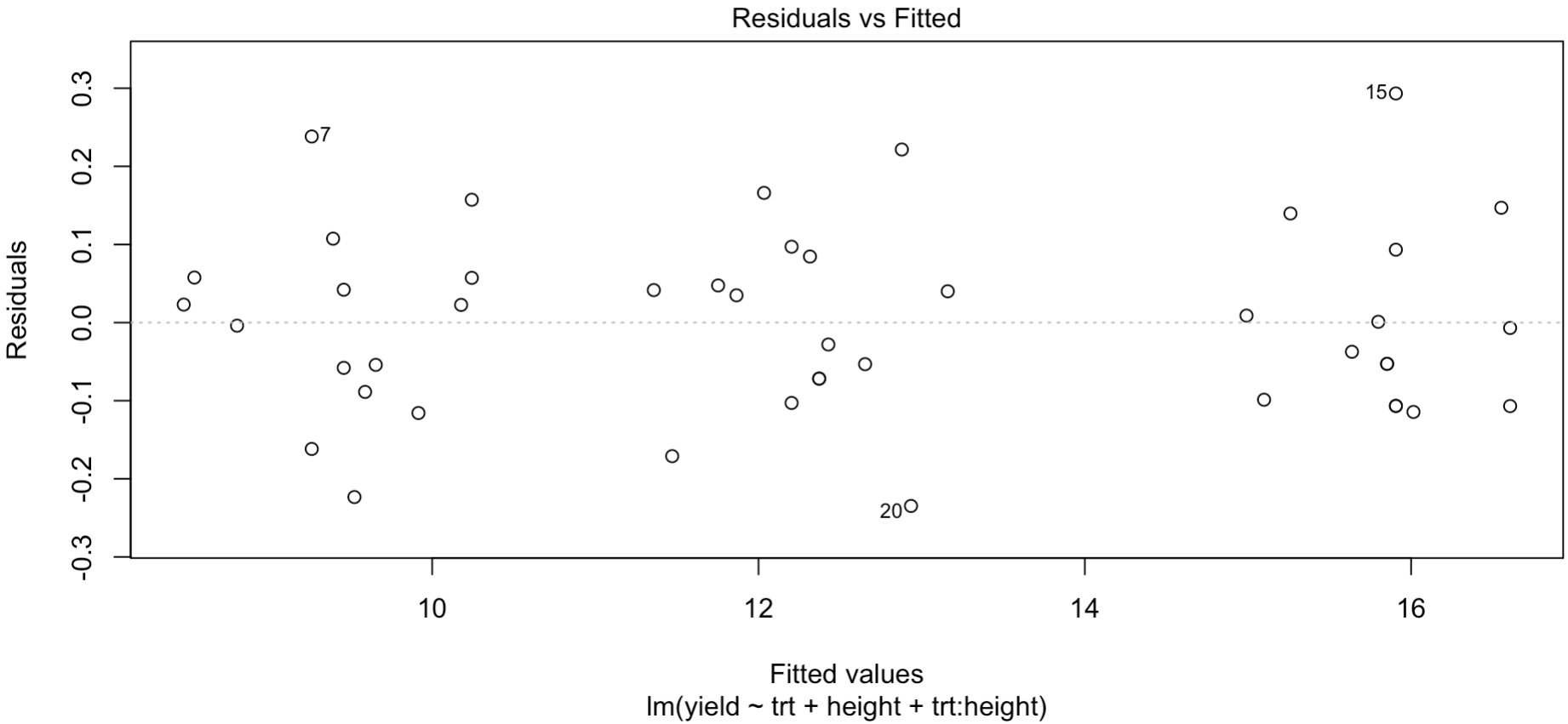
If we reject $H_0$: *Equal slopes*, then

- We can compute covariate-adjusted means with the different slopes.
- We can compare the covariate adjusted means—but the CI formulas are different from the ones above (details omitted).

# Soybean data

## Check for constant variance:

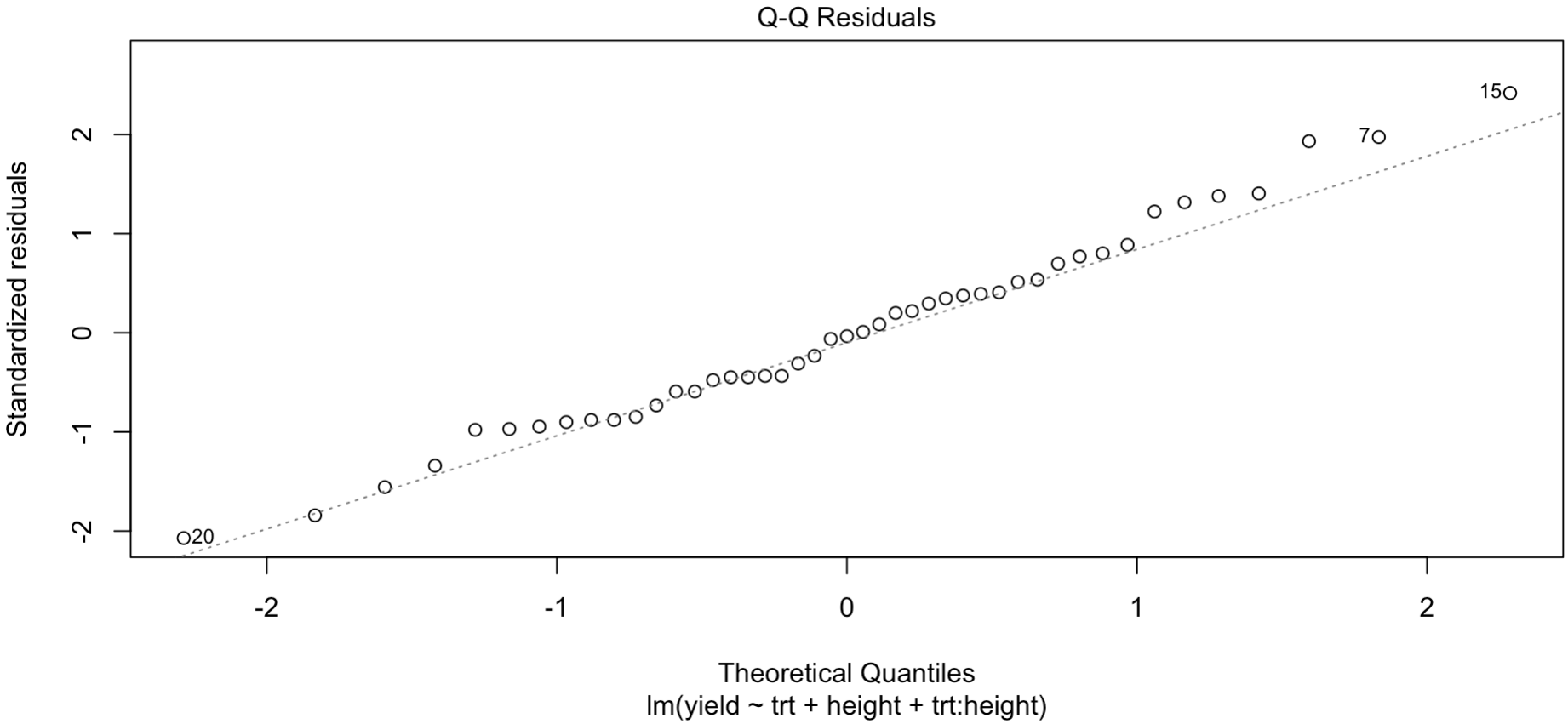
```
1 plot(lm_out, which = 1 , add.smooth = F)
```



# Soybean data

Check for normality of noise with normal QQ plot of the residuals:

```
1 plot(lm_out, which = 2)
```



# References

Kuehl, R. O. 2000. *Design of Experiments: Statistical Principles of Research Design and Analysis*. Duxbury/Thomson Learning.