Lesson 05

Christopher A. Swenson (chris@cswenson.com)

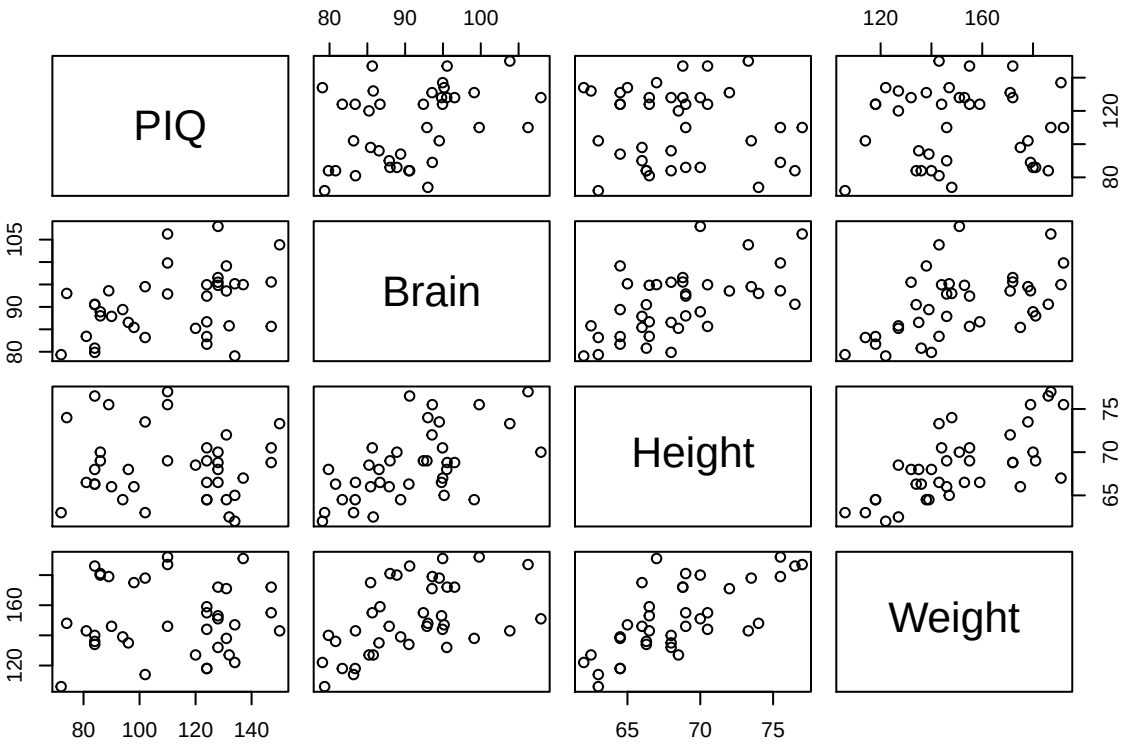
11/27/2021

IQ and physical characteristics

Load the iqsize data. Display a scatterplot matrix of the data. Fit a multiple linear regression model of PIQ on Brain, Height, and Weight. Display model results. Use the anova function to display anova table with sequential (type I) sums of squares. Use the Anova function from the car package to display anova table with adjusted (type III) sums of squares.

```
iqsize <- read.table("./Data/iqsize.txt", header=T)
attach(iqsize)

pairs(cbind(PIQ, Brain, Height, Weight))
```




```
model <- lm(PIQ ~ Brain + Height + Weight)
summary(model)
```

```
##
## Call:
## lm(formula = PIQ ~ Brain + Height + Weight)
##
## Residuals:
```

```
##    Min     1Q Median     3Q    Max
## -32.74 -12.09  -3.84  14.17  51.69
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.114e+02  6.297e+01   1.768 0.085979 .
## Brain        2.060e+00  5.634e-01   3.657 0.000856 ***
## Height      -2.732e+00  1.229e+00  -2.222 0.033034 *
## Weight       5.599e-04  1.971e-01   0.003 0.997750
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.79 on 34 degrees of freedom
## Multiple R-squared:  0.2949, Adjusted R-squared:  0.2327
## F-statistic: 4.741 on 3 and 34 DF,  p-value: 0.007215
```

```
# Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  1.114e+02  6.297e+01   1.768 0.085979 .
# Brain        2.060e+00  5.634e-01   3.657 0.000856 ***
# Height      -2.732e+00  1.229e+00  -2.222 0.033034 *
# Weight       5.599e-04  1.971e-01   0.003 0.997750
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# Residual standard error: 19.79 on 34 degrees of freedom
# Multiple R-squared:  0.2949,  Adjusted R-squared:  0.2327
# F-statistic: 4.741 on 3 and 34 DF,  p-value: 0.007215

anova(model) # Sequential (type I) SS
```

```
## Analysis of Variance Table
##
## Response: PIQ
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## Brain      1  2697.1 2697.09  6.8835 0.01293 *
## Height     1  2875.6 2875.65  7.3392 0.01049 *
## Weight     1     0.0    0.00  0.0000 0.99775
## Residuals 34 13321.8  391.82
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Analysis of Variance Table
# Response: PIQ
#           Df  Sum Sq Mean Sq F value  Pr(>F)
# Brain      1  2697.1 2697.09  6.8835 0.01293 *
# Height     1  2875.6 2875.65  7.3392 0.01049 *
# Weight     1     0.0    0.00  0.0000 0.99775
# Residuals 34 13321.8  391.82
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

library(car)
```

```
## Loading required package: carData
```

```r
Anova(model, type="III") # Adjusted (type III) SS
```

```
## Anova Table (Type III tests)
##
## Response: PIQ
##              Sum Sq Df F value    Pr(>F)
## (Intercept)  1225.2  1  3.1270 0.0859785 .
## Brain        5239.2  1 13.3716 0.0008556 ***
## Height       1934.7  1  4.9378 0.0330338 *
## Weight          0.0  1  0.0000 0.9977495
## Residuals   13321.8 34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Anova Table (Type III tests)
# Response: PIQ
#              Sum Sq Df F value    Pr(>F)
# (Intercept)  1225.2  1  3.1270 0.0859785 .
# Brain        5239.2  1 13.3716 0.0008556 ***
# Height       1934.7  1  4.9378 0.0330338 *
# Weight          0.0  1  0.0000 0.9977495
# Residuals   13321.8 34
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

detach(iqsize)
```

Underground air quality

Load the babybirds data. Display a scatterplot matrix of the data. Use the scatter3d function from the car package to create a 3D scatterplot of the data. Fit a multiple linear regression model of Vent on O2 and CO2. Display model results. Use the Anova function from the car package to display anova table with adjusted (type III) sums of squares.

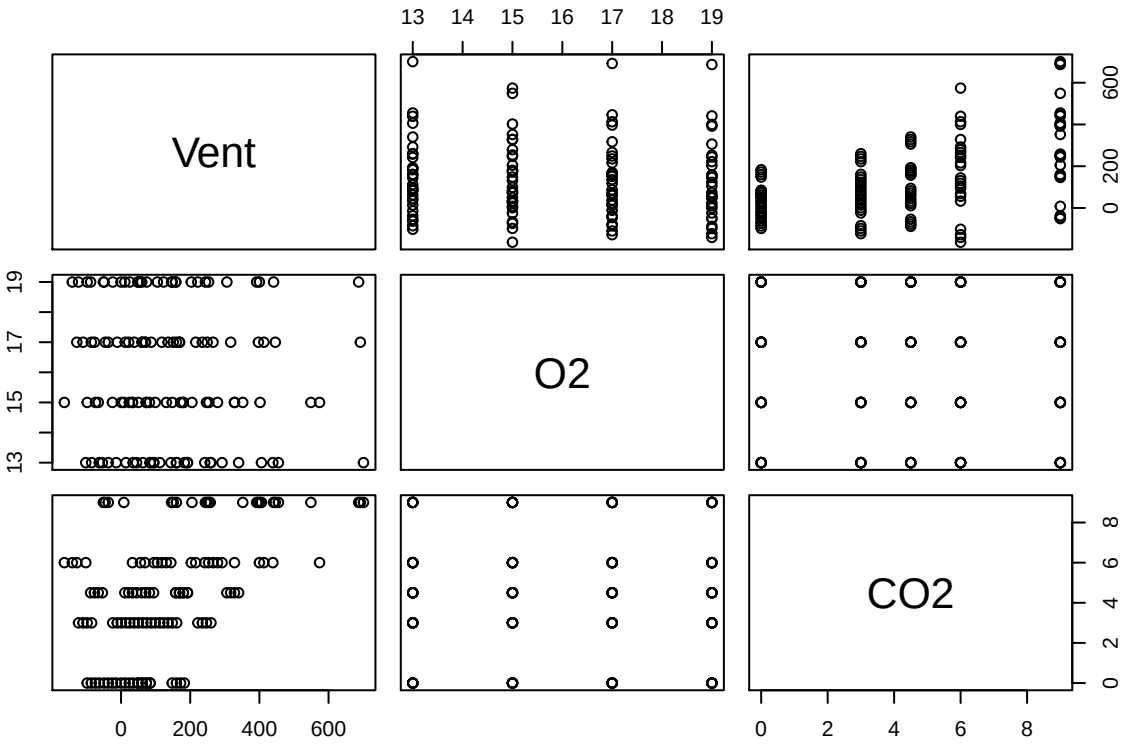
```r
babybirds <- read.table("./Data/babybirds.txt", header=T)
attach(babybirds)
```

```
## The following object is masked from package:datasets:
##
##     CO2
```

```r
pairs(cbind(Vent, O2, CO2))

library(car)
library(rgl) # need rgl library as well, install first
scatter3d(Vent ~ O2 + CO2)
```

```
## Loading required namespace: mgcv
```




```
#scatter3d(Vent ~ O2 + CO2, revolutions=3, speed=0.5, grid=F)

model <- lm(Vent ~ O2 + CO2)
summary(model)
```

```
##
## Call:
## lm(formula = Vent ~ O2 + CO2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -356.57  -96.50    8.73   84.68  422.44
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   85.901    106.006   0.810    0.419
## O2            -5.330      6.425  -0.830    0.408
## CO2           31.103      4.789   6.495  2.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 157.4 on 117 degrees of freedom
## Multiple R-squared:  0.2682, Adjusted R-squared:  0.2557
## F-statistic: 21.44 on 2 and 117 DF,  p-value: 1.169e-08
```

```
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)   85.901    106.006   0.810    0.419
# O2            -5.330      6.425  -0.830    0.408
# CO2           31.103      4.789   6.495  2.1e-09 ***
# ---
```

```
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# Residual standard error: 157.4 on 117 degrees of freedom
# Multiple R-squared:  0.2682,  Adjusted R-squared:  0.2557
# F-statistic: 21.44 on 2 and 117 DF,  p-value: 1.169e-08

Anova(model, type="III") # Adjusted (type III) SS
```

```
## Anova Table (Type III tests)
##
## Response: Vent
##              Sum Sq  Df F value    Pr(>F)
## (Intercept)   16262   1  0.6566    0.4194
## O2            17045   1  0.6883    0.4084
## CO2         1044773   1 42.1866 2.104e-09 ***
## Residuals   2897566 117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Anova Table (Type III tests)
# Response: Vent
#              Sum Sq  Df F value    Pr(>F)
# (Intercept)   16262   1  0.6566    0.4194
# O2            17045   1  0.6883    0.4084
# CO2         1044773   1 42.1866 2.104e-09 ***
# Residuals   2897566 117
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

detach(babybirds)
```

Soapsuds example (using matrices)

Load the soapsuds data. Fit a simple linear regression model of suds on soap and store the model matrix, X. Display model results. Calculate , and . Fit a multiple linear regression model with linearly dependent predictors.

```
soapsuds <- read.table("./Data/soapsuds.txt", header=T)
attach(soapsuds)

model <- lm(suds ~ soap, x=T)
summary(model)
```

```
##
## Call:
## lm(formula = suds ~ soap, x = T)
##
## Residuals:
##       1       2       3       4       5       6       7
## -2.3214  1.9286  0.1786  1.4286 -1.3214  1.9286 -1.8214
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.6786     4.2220  -0.634    0.554
## soap          9.5000     0.7553  12.579 5.64e-05 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.998 on 5 degrees of freedom
## Multiple R-squared:  0.9694, Adjusted R-squared:  0.9632
## F-statistic: 158.2 on 1 and 5 DF,  p-value: 5.639e-05
```

```
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  -2.6786     4.2220  -0.634    0.554
# soap          9.5000     0.7553  12.579 5.64e-05 ***

X <- model$x
t(X) %*% X
```

```
##             (Intercept)   soap
## (Intercept)         7.0  38.50
## soap               38.5 218.75
```

```
#             (Intercept)   soap
# (Intercept)         7.0  38.50
# soap               38.5 218.75

t(X) %*% suds
```

```
##             [,1]
## (Intercept)  347
## soap        1975
```

```
#             [,1]
# (Intercept)  347
# soap        1975

solve(t(X) %*% X)
```

```
##             (Intercept)       soap
## (Intercept)   4.4642857 -0.7857143
## soap         -0.7857143  0.1428571
```

```
#             (Intercept)       soap
# (Intercept)   4.4642857 -0.7857143
# soap         -0.7857143  0.1428571

solve(t(X) %*% X) %*% (t(X) %*% suds)
```

```
##                  [,1]
## (Intercept) -2.678571
## soap         9.500000
```

```
#                  [,1]
# (Intercept) -2.678571
# soap         9.500000

soap2 <- 2*soap
model <- lm(suds ~ soap + soap2)
summary(model)
```

```
##
```

```
## Call:
## lm(formula = suds ~ soap + soap2)
##
## Residuals:
##       1       2       3       4       5       6       7
## -2.3214  1.9286  0.1786  1.4286 -1.3214  1.9286 -1.8214
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.6786     4.2220  -0.634    0.554
## soap          9.5000     0.7553  12.579 5.64e-05 ***
## soap2             NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.998 on 5 degrees of freedom
## Multiple R-squared:  0.9694, Adjusted R-squared:  0.9632
## F-statistic: 158.2 on 1 and 5 DF,  p-value: 5.639e-05
```

```
# Coefficients: (1 not defined because of singularities)
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  -2.6786     4.2220  -0.634    0.554
# soap          9.5000     0.7553  12.579 5.64e-05 ***
# soap2             NA         NA      NA       NA

detach(soapsuds)
```

Pastry sweetness

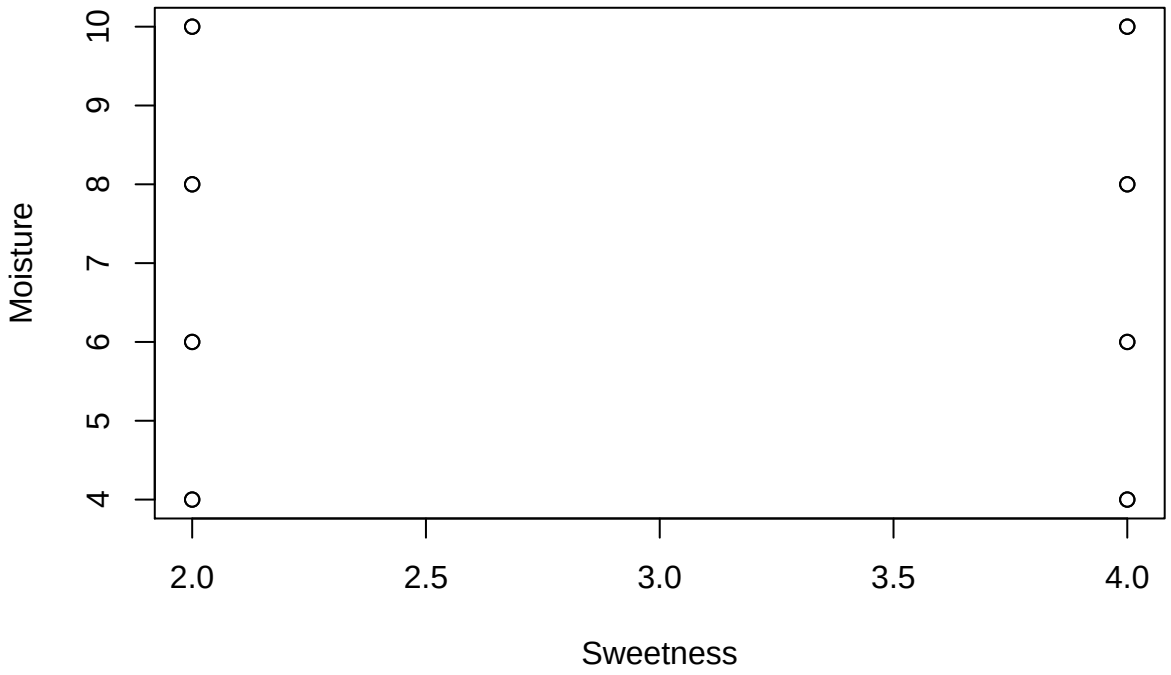
Load the pastry data. Calculate the correlation between the predictors and create a scatterplot. Fit a multiple linear regression model of Rating on Moisture and Sweetness and display the model results. Create a scatterplot of the data with points marked by Sweetness and two lines representing the fitted regression equation for each sweetness level. Fit a simple linear regression model of Rating on Moisture and display the model results. Fit a simple linear regression model of Rating on Sweetness and display the model results.

```
pastry <- read.table("./Data/pastry.txt", header=T)
attach(pastry)

cor(Sweetness, Moisture) # 0
```

```
## [1] 0
```

```
plot(Sweetness, Moisture)
```

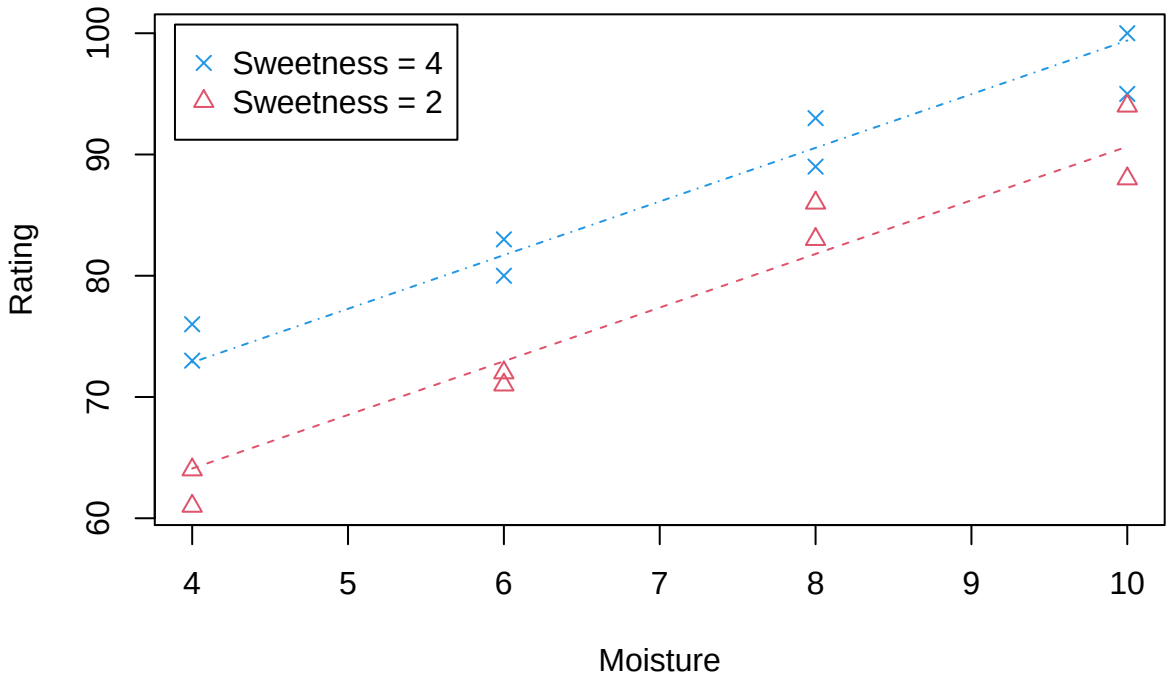



```
model.12 <- lm(Rating ~ Moisture + Sweetness)
summary(model.12)
```

```
##
## Call:
## lm(formula = Rating ~ Moisture + Sweetness)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
## Moisture      4.4250     0.3011  14.695 1.78e-09 ***
## Sweetness     4.3750     0.6733   6.498 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 on 13 degrees of freedom
## Multiple R-squared:  0.9521, Adjusted R-squared:  0.9447
## F-statistic: 129.1 on 2 and 13 DF,  p-value: 2.658e-09
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
# Moisture      4.4250     0.3011  14.695 1.78e-09 ***
# Sweetness     4.3750     0.6733   6.498 2.01e-05 ***
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# Residual standard error: 2.693 on 13 degrees of freedom
# Multiple R-squared:  0.9521,  Adjusted R-squared:  0.9447
# F-statistic: 129.1 on 2 and 13 DF,  p-value: 2.658e-09
```

```
plot(Moisture, Rating, type="n")
for (i in 1:16) points(Moisture[i], Rating[i], pch=Sweetness[i], col=Sweetness[i])
for (i in c(2,4)) lines(Moisture[Sweetness==i], fitted(model.12)[Sweetness==i],
                        lty=i, col=i)
legend("topleft", legend=c("Sweetness = 4",
                           "Sweetness = 2"),
       col=c(4,2), pch=c(4,2), inset=0.02)
```




```
model.1 <- lm(Rating ~ Moisture)
summary(model.1)
```

```
##
## Call:
## lm(formula = Rating ~ Moisture)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.475 -4.688 -0.100  4.638  7.525
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   50.775      4.395  11.554 1.52e-08 ***
## Moisture       4.425      0.598   7.399 3.36e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.349 on 14 degrees of freedom
## Multiple R-squared:  0.7964, Adjusted R-squared:  0.7818
## F-statistic: 54.75 on 1 and 14 DF,  p-value: 3.356e-06
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)   50.775      4.395  11.554 1.52e-08 ***
```

```
# Moisture       4.425      0.598   7.399 3.36e-06 ***
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# Residual standard error: 5.349 on 14 degrees of freedom
# Multiple R-squared:  0.7964,  Adjusted R-squared:  0.7818
# F-statistic: 54.75 on 1 and 14 DF,  p-value: 3.356e-06
```

```
model.2 <- lm(Rating ~ Sweetness)
summary(model.2)
```

```
##
## Call:
## lm(formula = Rating ~ Sweetness)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.375  -7.312  -0.125   8.688  16.625
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   68.625      8.610   7.970 1.43e-06 ***
## Sweetness      4.375      2.723   1.607     0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.89 on 14 degrees of freedom
## Multiple R-squared:  0.1557, Adjusted R-squared:  0.09539
## F-statistic: 2.582 on 1 and 14 DF,  p-value: 0.1304
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)   68.625      8.610   7.970 1.43e-06 ***
# Sweetness      4.375      2.723   1.607     0.13
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# Residual standard error: 10.89 on 14 degrees of freedom
# Multiple R-squared:  0.1557,  Adjusted R-squared:  0.09539
# F-statistic: 2.582 on 1 and 14 DF,  p-value: 0.1304
```

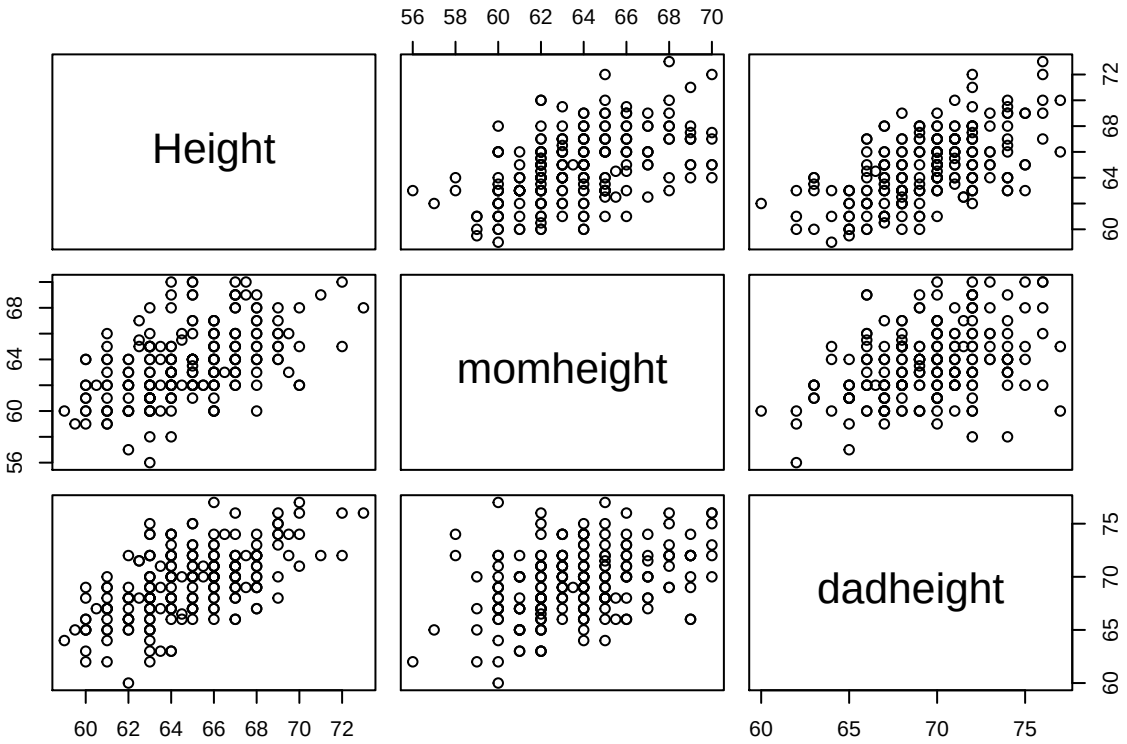
```
detach(pastry)
```

Female stat students

Load the statfemales data. Display a scatterplot matrix of the data. Fit a multiple linear regression model of Height on momheight and dadheight and display the model results. Create a residual plot.

```
statfemales <- read.table("./Data/stat_females.txt", header=T)
attach(statfemales)

pairs(cbind(Height, momheight, dadheight))
```



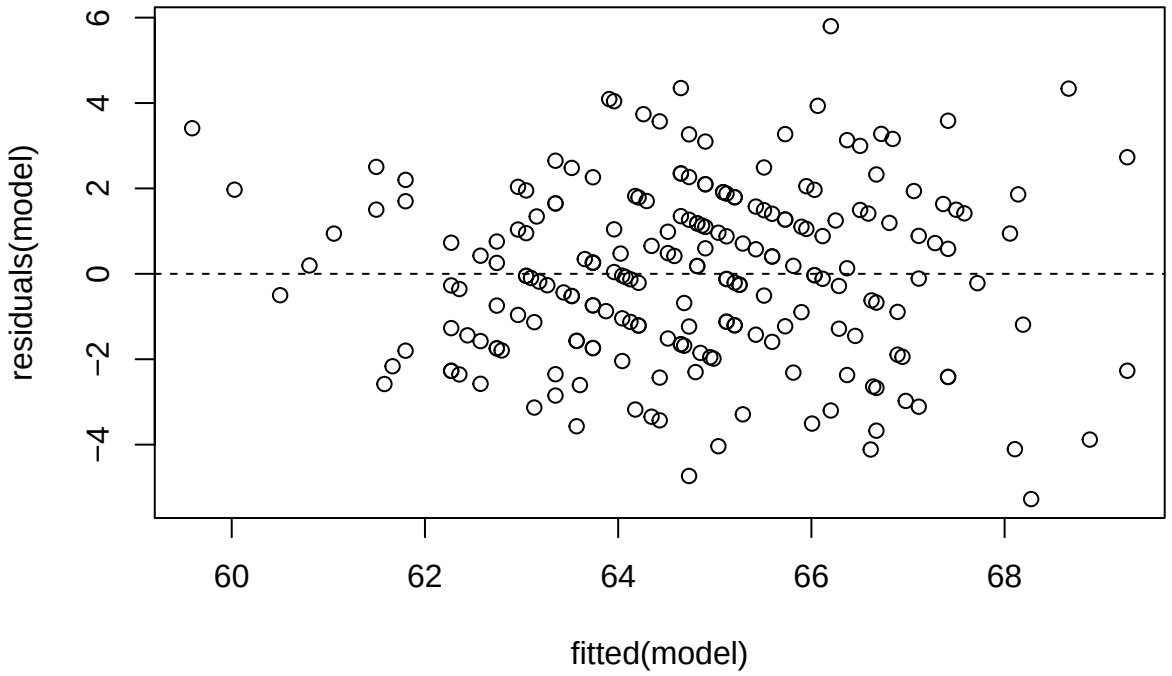

```
model <- lm(Height ~ momheight + dadheight)
summary(model)
```

```
##
## Call:
## lm(formula = Height ~ momheight + dadheight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.2748 -1.5562 -0.0372  1.4721  5.7993
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.54725    3.69278   5.023 1.08e-06 ***
## momheight    0.30351    0.05446   5.573 7.61e-08 ***
## dadheight    0.38786    0.04721   8.216 2.10e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.031 on 211 degrees of freedom
## Multiple R-squared:  0.4335, Adjusted R-squared:  0.4281
## F-statistic: 80.73 on 2 and 211 DF,  p-value: < 2.2e-16
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) 18.54725    3.69278   5.023 1.08e-06 ***
# momheight    0.30351    0.05446   5.573 7.61e-08 ***
# dadheight    0.38786    0.04721   8.216 2.10e-14 ***
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# Residual standard error: 2.031 on 211 degrees of freedom
```

```
# Multiple R-squared:  0.4335,  Adjusted R-squared:  0.4281
# F-statistic: 80.73 on 2 and 211 DF,  p-value: < 2.2e-16

plot(fitted(model), residuals(model),
     panel.last = abline(h=0, lty=2))
```




```
detach(statfemales)
```

Hospital infections

Load the infectionrisk data. Fit a multiple linear regression model of InfctRsk on Stay, Age, and Xray and display the model results.

```
infectionrisk <- read.table("./Data/infectionrisk.txt", header=T)
attach(infectionrisk)

model <- lm(InfctRsk ~ Stay + Age + Xray)
summary(model)
```

```
##
## Call:
## lm(formula = InfctRsk ~ Stay + Age + Xray)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.57423 -0.59005 -0.05824  0.51546  2.44614
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.036730   1.754704  -0.591   0.5571
## Stay         0.484895   0.101794   4.763 1.47e-05 ***
## Age         -0.014825   0.031201  -0.475   0.6366
## Xray         0.017418   0.007258   2.400   0.0199 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9863 on 54 degrees of freedom
## Multiple R-squared:  0.4589, Adjusted R-squared:  0.4288
## F-statistic: 15.26 on 3 and 54 DF,  p-value: 2.588e-07
```

```
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  1.001162   1.314724   0.761 0.448003
# Stay         0.308181   0.059396   5.189 9.88e-07 ***
# Age         -0.023005   0.023516  -0.978 0.330098
# Xray         0.019661   0.005759   3.414 0.000899 ***
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.085 on 109 degrees of freedom
# Multiple R-squared:  0.363,  Adjusted R-squared:  0.3455
# F-statistic:  20.7 on 3 and 109 DF,  p-value: 1.087e-10

detach(infectionrisk)
```

Physiological measurements (using matrices)

Load the bodyfat data. Fit a multiple linear regression model of BodyFat on Triceps, Thigh, and Midarm and store the model matrix, X. Display model results. Calculate MSE and and multiply them to find the the variance-covariance matrix of the regression parameters. Use the variance-covariance matrix of the regression parameters to derive: the regression parameter standard errors. covariances and correlations between regression parameter estimates.

```
bodyfat <- read.table("./Data/bodyfat.txt", header=T)
attach(bodyfat)

model <- lm(Bodyfat ~ Triceps + Thigh + Midarm, x=T)
summary(model)
```

```
##
## Call:
## lm(formula = Bodyfat ~ Triceps + Thigh + Midarm, x = T)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7263 -1.6111  0.3923  1.4656  4.1277
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  117.085     99.782   1.173    0.258
## Triceps        4.334      3.016   1.437    0.170
## Thigh         -2.857      2.582  -1.106    0.285
## Midarm        -2.186      1.595  -1.370    0.190
##
## Residual standard error: 2.48 on 16 degrees of freedom
## Multiple R-squared:  0.8014, Adjusted R-squared:  0.7641
## F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  117.085     99.782   1.173    0.258
# Triceps        4.334      3.016   1.437    0.170
# Thigh         -2.857      2.582  -1.106    0.285
# Midarm        -2.186      1.595  -1.370    0.190
#
# Residual standard error: 2.48 on 16 degrees of freedom
# Multiple R-squared:  0.8014,  Adjusted R-squared:  0.7641
# F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06

anova(model)
```

```
## Analysis of Variance Table
##
## Response: Bodyfat
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Triceps    1 352.27  352.27 57.2768 1.131e-06 ***
## Thigh      1  33.17   33.17  5.3931   0.03373 *
## Midarm     1  11.55   11.55  1.8773   0.18956
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value    Pr(>F)
# Triceps    1 352.27  352.27 57.2768 1.131e-06 ***
# Thigh      1  33.17   33.17  5.3931   0.03373 *
# Midarm     1  11.55   11.55  1.8773   0.18956
# Residuals 16  98.40    6.15
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

MSE <- sum(residuals(model)^2)/model$df.residual # 6.150306
X <- model$x
XTXinv <- solve(t(X) %*% X)
#             (Intercept)    Triceps       Thigh      Midarm
# (Intercept)  1618.86721 48.8102522 -41.8487041 -25.7987855
# Triceps        48.81025  1.4785133  -1.2648388  -0.7785022
# Thigh         -41.84870 -1.2648388   1.0839791   0.6657581
# Midarm        -25.79879 -0.7785022   0.6657581   0.4139009

sqrt(MSE*diag(XTXinv)) # standard errors of the regression parameters
```

```
## (Intercept)     Triceps       Thigh      Midarm
##   99.782403    3.015511    2.582015    1.595499
```

```
# (Intercept)     Triceps       Thigh      Midarm
#   99.782403    3.015511    2.582015    1.595499

MSE*XTXinv[2,3] # cov(b1, b2) = -7.779145
```

```
## [1] -7.779145
```

```
XTXinv[2,3]/sqrt(XTXinv[2,2]*XTXinv[3,3]) # cor(b1, b2) = -0.9991072
```

```
## [1] -0.9991072
```

```
detach(bodyfat)
```

Peruvian blood pressure

Load the peru data. Calculate FracLife variable. Fit full multiple linear regression model of Systol on nine predictors. Fit reduced multiple linear regression model of Systol on four predictors. Calculate SSE for the full and reduced models. Calculate the general linear F statistic by hand and find the p-value. Use the anova function with full and reduced models to display F-statistic and p-value directly.

```
peru <- read.table("./Data/peru.txt", header=T)
attach(peru)

FracLife <- Years/Age

model.1 <- lm(Systol ~ Age + Years + FracLife + Weight + Height + Chin +
              Forearm + Calf + Pulse)
summary(model.1)
```

```
##
## Call:
## lm(formula = Systol ~ Age + Years + FracLife + Weight + Height +
##     Chin + Forearm + Calf + Pulse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.3442  -6.3972   0.0507   5.7292  14.5257
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  146.81907   48.97096   2.998 0.005526 **
## Age           -1.12144    0.32741  -3.425 0.001855 **
## Years          2.45538    0.81458   3.014 0.005306 **
## FracLife    -115.29395   30.16900  -3.822 0.000648 ***
## Weight         1.41393    0.43097   3.281 0.002697 **
## Height        -0.03464    0.03686  -0.940 0.355194
## Chin          -0.94369    0.74097  -1.274 0.212923
## Forearm       -1.17085    1.19329  -0.981 0.334612
## Calf          -0.15867    0.53716  -0.295 0.769810
## Pulse          0.11455    0.17043   0.672 0.506818
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.655 on 29 degrees of freedom
## Multiple R-squared:  0.6674, Adjusted R-squared:  0.5641
## F-statistic: 6.465 on 9 and 29 DF,  p-value: 5.241e-05
```

```
#               Estimate Std. Error t value Pr(>|t|)
# (Intercept)  146.81907   48.97096   2.998 0.005526 **
# Age           -1.12144    0.32741  -3.425 0.001855 **
# Years          2.45538    0.81458   3.014 0.005306 **
# FracLife    -115.29395   30.16900  -3.822 0.000648 ***
# Weight         1.41393    0.43097   3.281 0.002697 **
# Height        -0.03464    0.03686  -0.940 0.355194
# Chin          -0.94369    0.74097  -1.274 0.212923
```

```
# Forearm       -1.17085    1.19329  -0.981 0.334612
# Calf          -0.15867    0.53716  -0.295 0.769810
# Pulse          0.11455    0.17043   0.672 0.506818

anova(model.1)
```

```
## Analysis of Variance Table
##
## Response: Systol
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Age        1    0.22    0.22  0.0030  0.956852
## Years      1   82.55   82.55  1.1019  0.302514
## FracLife   1 3112.41 3112.41 41.5449 4.728e-07 ***
## Weight     1  706.54  706.54  9.4311  0.004603 **
## Height     1    1.68    1.68  0.0224  0.882117
## Chin       1  297.68  297.68  3.9735  0.055704 .
## Forearm    1  113.91  113.91  1.5205  0.227440
## Calf       1   10.01   10.01  0.1336  0.717420
## Pulse      1   33.84   33.84  0.4518  0.506818
## Residuals 29 2172.58   74.92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df  Sum Sq Mean Sq F value    Pr(>F)
# Age        1    0.22    0.22  0.0030  0.956852
# Years      1   82.55   82.55  1.1019  0.302514
# FracLife   1 3112.41 3112.41 41.5449 4.728e-07 ***
# Weight     1  706.54  706.54  9.4311  0.004603 **
# Height     1    1.68    1.68  0.0224  0.882117
# Chin       1  297.68  297.68  3.9735  0.055704 .
# Forearm    1  113.91  113.91  1.5205  0.227440
# Calf       1   10.01   10.01  0.1336  0.717420
# Pulse      1   33.84   33.84  0.4518  0.506818
# Residuals 29 2172.58   74.92

model.2 <- lm(Systol ~ Age + Years + FracLife + Weight)
summary(model.2)
```

```
##
## Call:
## lm(formula = Systol ~ Age + Years + FracLife + Weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.890  -5.976   0.058   5.407  16.835
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  116.8354    21.9797   5.316 6.69e-06 ***
## Age           -0.9507     0.3164  -3.004 0.004971 **
## Years          2.3393     0.7714   3.032 0.004621 **
## FracLife    -108.0728    28.3302  -3.815 0.000549 ***
## Weight         0.8324     0.2754   3.022 0.004742 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 8.795 on 34 degrees of freedom
## Multiple R-squared:  0.5974, Adjusted R-squared:   0.55
## F-statistic: 12.61 on 4 and 34 DF,  p-value: 2.142e-06
```

```
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  116.8354    21.9797   5.316 6.69e-06 ***
# Age           -0.9507     0.3164  -3.004 0.004971 **
# Years          2.3393     0.7714   3.032 0.004621 **
# FracLife    -108.0728    28.3302  -3.815 0.000549 ***
# Weight         0.8324     0.2754   3.022 0.004742 **

anova(model.2)
```

```
## Analysis of Variance Table
##
## Response: Systol
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Age        1    0.22    0.22  0.0029  0.957480
## Years      1   82.55   82.55  1.0673  0.308840
## FracLife   1 3112.41 3112.41 40.2409 3.094e-07 ***
## Weight     1  706.54  706.54  9.1350  0.004742 **
## Residuals 34 2629.71   77.34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df  Sum Sq Mean Sq F value    Pr(>F)
# Age        1    0.22    0.22  0.0029  0.957480
# Years      1   82.55   82.55  1.0673  0.308840
# FracLife   1 3112.41 3112.41 40.2409 3.094e-07 ***
# Weight     1  706.54  706.54  9.1350  0.004742 **
# Residuals 34 2629.71   77.34

(2629.71-2172.58)/(34-29) / (2172.58/29) # F = 1.220371
```

```
## [1] 1.220371
```

```
pf(1.220371, 5, 29, lower.tail=F) # p-value = 0.3247213
```

```
## [1] 0.3247213
```

```
anova(model.2, model.1)
```

```
## Analysis of Variance Table
##
## Model 1: Systol ~ Age + Years + FracLife + Weight
## Model 2: Systol ~ Age + Years + FracLife + Weight + Height + Chin + Forearm +
##     Calf + Pulse
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     34 2629.7
## 2     29 2172.6  5    457.12 1.2204 0.3247
```

```
#   Res.Df    RSS Df Sum of Sq      F Pr(>F)
# 1     34 2629.7
# 2     29 2172.6  5    457.12 1.2204 0.3247

detach(peru)
```

Measurements of college students

Load the Physical data. Fit full multiple linear regression model of Height on LeftArm, LeftFoot, HeadCirc, and nose. Create a residual plot. Fit reduced multiple linear regression model of Height on LeftArm and LeftFoot. Calculate SSE for the full and reduced models. Calculate the general linear F statistic by hand and find the p-value. Use the anova function with full and reduced models to display F-statistic and p-value directly. Calculate partial R-squared for (LeftArm | LeftFoot).

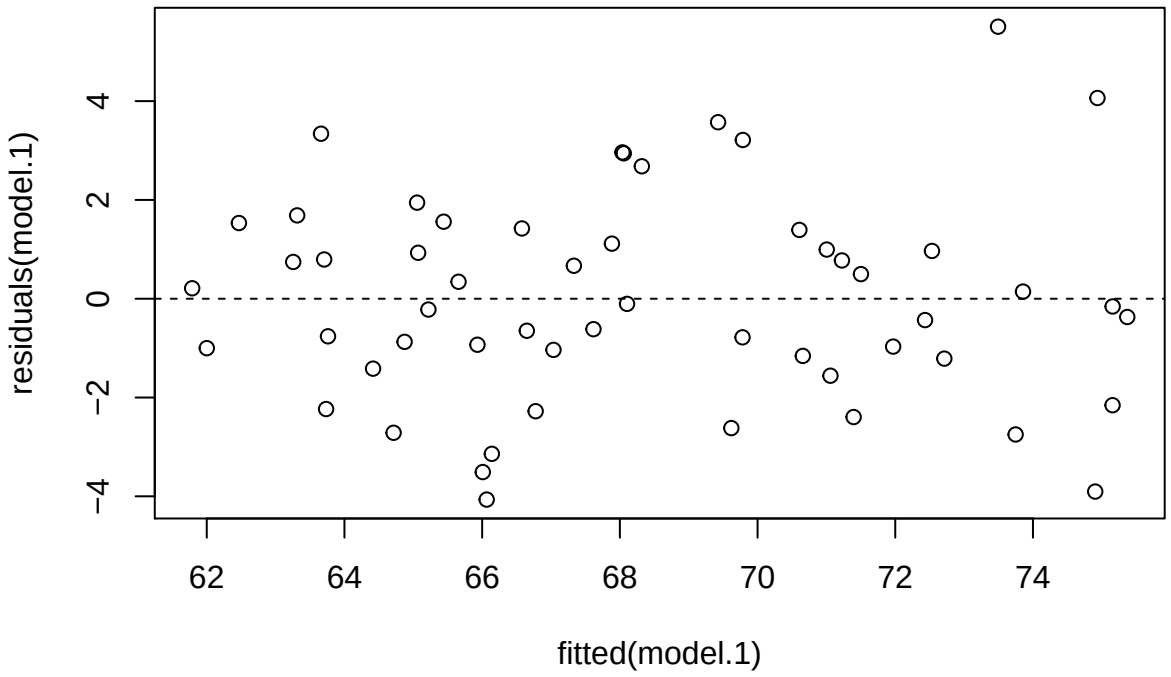
```
physical <- read.table("./Data/Physical.txt", header=T)
attach(physical)

model.1 <- lm(Height ~ LeftArm + LeftFoot + HeadCirc + nose)
summary(model.1)
```

```
##
## Call:
## lm(formula = Height ~ LeftArm + LeftFoot + HeadCirc + nose)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0660 -1.1850 -0.1569  1.2539  5.5071
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.50265    7.83031   2.363   0.0221 *
## LeftArm      0.80205    0.17074   4.697 2.09e-05 ***
## LeftFoot     0.99730    0.16230   6.145 1.30e-07 ***
## HeadCirc     0.08052    0.14952   0.539   0.5926
## nose        -0.14740    0.49233  -0.299   0.7659
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.183 on 50 degrees of freedom
## Multiple R-squared:  0.774,  Adjusted R-squared:  0.7559
## F-statistic: 42.81 on 4 and 50 DF,  p-value: 1.447e-15
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) 18.50265    7.83031   2.363   0.0221 *
# LeftArm      0.80205    0.17074   4.697 2.09e-05 ***
# LeftFoot     0.99730    0.16230   6.145 1.30e-07 ***
# HeadCirc     0.08052    0.14952   0.539   0.5926
# nose        -0.14740    0.49233  -0.299   0.7659

plot(fitted(model.1), residuals(model.1),
     panel.last = abline(h=0, lty=2))
```




```
anova(model.1)
```

```
## Analysis of Variance Table
##
## Response: Height
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## LeftArm    1 590.21  590.21 123.8106 3.917e-15 ***
## LeftFoot   1 224.35  224.35  47.0621 9.931e-09 ***
## HeadCirc   1   1.40    1.40   0.2940    0.5901
## nose       1   0.43    0.43   0.0896    0.7659
## Residuals 50 238.35    4.77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq  F value    Pr(>F)
# LeftArm    1 590.21  590.21 123.8106 3.917e-15 ***
# LeftFoot   1 224.35  224.35  47.0621 9.931e-09 ***
# HeadCirc   1   1.40    1.40   0.2940    0.5901
# nose       1   0.43    0.43   0.0896    0.7659
# Residuals 50 238.35    4.77

model.2 <- lm(Height ~ LeftArm + LeftFoot)
summary(model.2)
```

```
##
## Call:
## lm(formula = Height ~ LeftArm + LeftFoot)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.075 -1.179 -0.099  1.248  5.196
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21.8572     3.5840   6.098 1.35e-07 ***
## LeftArm       0.7958     0.1652   4.816 1.31e-05 ***
## LeftFoot      1.0229     0.1468   6.969 5.54e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.149 on 52 degrees of freedom
## Multiple R-squared:  0.7723, Adjusted R-squared:  0.7635
## F-statistic: 88.18 on 2 and 52 DF,  p-value: < 2.2e-16
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  21.8572     3.5840   6.098 1.35e-07 ***
# LeftArm       0.7958     0.1652   4.816 1.31e-05 ***
# LeftFoot      1.0229     0.1468   6.969 5.54e-09 ***

anova(model.2)
```

```
## Analysis of Variance Table
##
## Response: Height
##           Df Sum Sq Mean Sq F value    Pr(>F)
## LeftArm    1 590.21  590.21 127.782 1.275e-15 ***
## LeftFoot   1 224.35  224.35  48.572 5.538e-09 ***
## Residuals 52 240.18    4.62
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value    Pr(>F)
# LeftArm    1 590.21  590.21 127.782 1.275e-15 ***
# LeftFoot   1 224.35  224.35  48.572 5.538e-09 ***
# Residuals 52 240.18    4.62

(240.18-238.35)/(52-50) / (238.35/50) # F = 0.1919446
```

```
## [1] 0.1919446
```

```
pf(0.1919446, 2, 50, lower.tail=F) # p-value = 0.8259579
```

```
## [1] 0.8259579
```

```
anova(model.2, model.1)
```

```
## Analysis of Variance Table
##
## Model 1: Height ~ LeftArm + LeftFoot
## Model 2: Height ~ LeftArm + LeftFoot + HeadCirc + nose
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     52 240.18
## 2     50 238.35  2    1.8289 0.1918 0.8261
```

```
#   Res.Df    RSS Df Sum of Sq      F Pr(>F)
# 1     52 240.18
# 2     50 238.35  2    1.8289 0.1918 0.8261

model.3 <- lm(Height ~ LeftFoot)
anova(model.3)
```

```
## Analysis of Variance Table
##
## Response: Height
##           Df Sum Sq Mean Sq F value    Pr(>F)
## LeftFoot   1 707.42  707.42  107.95 2.172e-14 ***
## Residuals 53 347.33    6.55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value    Pr(>F)
# LeftFoot   1 707.42  707.42  107.95 2.172e-14 ***
# Residuals 53 347.33    6.55

(347.33-240.18) / 347.33 # Partial R-squared (LeftArm | LeftFoot) = 0.3084962
```

```
## [1] 0.3084962
```

```
detach(physical)
```