Statistics One

Lecture 9

Multiple regression in R

Two segments

- Multiple regression (MR) analysis in R
- MR analysis, standardized

Lecture 9
Segment 1

Multiple regression (MR) analysis in R

Goal

- Write a script in R
  - Simple regression
    - 1 predictor
  - Multiple regression
    - 2 predictors

Example

- Fictive data
  - Outcome (Y)
    - Physical endurance (endurance)
  - Predictors (X1, X2)
    - Age (age)
    - Years engaged in active exercise (activeyears)

Endurance example

- Data are available in the following file:
  - STATS1.EX.04.txt

Write a script

- First line(s) of code should be comments

  # Statistics One, Lecture 9, example script

  # Multiple regression analysis

Write a script

- Read data into a dataframe called “endur”

```
endur <- read.table(“STATS1.EX.04.txt”, header = T)
```

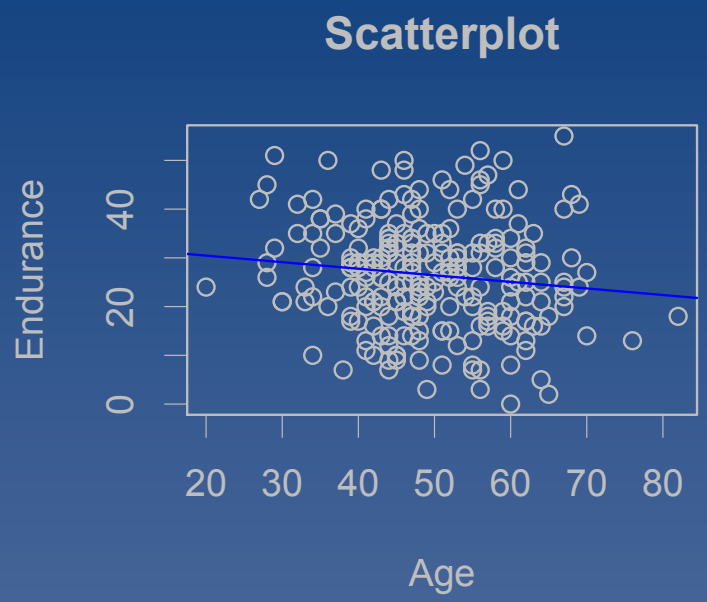
Write a script

- Examine a scatterplot

```
plot(endur$endurance ~ endur$age, main = "Scatterplot", ylab =
  "Endurance", xlab = "Age")

abline(lm(endur$endurance ~ endur$age), col="blue”)
```

endurance ~ age



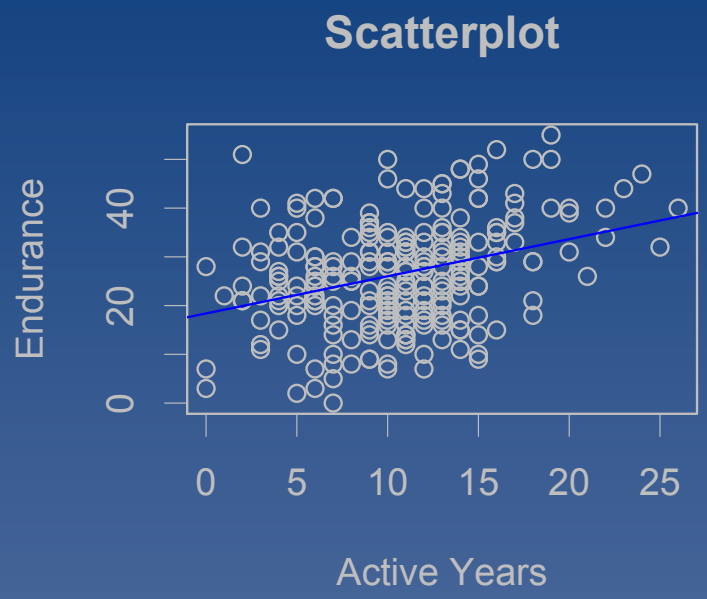
Write a script

- Examine another scatterplot

```
plot(endur$endurance ~ endur$activeyears, main = "Scatterplot",
  ylab = "Endurance", xlab = ”Active years")

abline(lm(endur$endurance ~ endur$activeyears), col="blue”)
```

endurance ~ activeyears



Write a script

- Simple regression (one predictor)

```
model1 = lm(endur$endurance ~ endur$age)

summary(model1)
```

endurance = 33.16 + -0.13(age)

```
Call:
lm(formula = endur$endurance ~ endur$age)

Residuals:
     Min       1Q   Median       3Q      Max
-25.0734  -7.6331   0.0974   6.7710  30.8696

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.15667    3.42033   9.694   <2e-16 ***
endur$age   -0.13472    0.06812  -1.978   0.0491 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.76 on 243 degrees of freedom
Multiple R-squared: 0.01584,  Adjusted R-squared: 0.01179
F-statistic: 3.911 on 1 and 243 DF,  p-value: 0.04911
```

Write a script

- Simple regression (one predictor)

```
model2 = lm(endur$endurance ~ endur$activeyears)

summary(model2)
```

endurance = 18.39 + 0.76(activeyears)

```
Call:
lm(formula = endur$endurance ~ endur$activeyears)

Residuals:
     Min       1Q   Median       3Q      Max
-23.7296  -7.0671   0.5579   5.7454  31.0829

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        18.3921     1.5998  11.496  < 2e-16 ***
endur$activeyears   0.7625     0.1369   5.571  6.7e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.21 on 243 degrees of freedom
Multiple R-squared: 0.1133, Adjusted R-squared: 0.1096
F-statistic: 31.04 on 1 and 243 DF,  p-value: 6.697e-08
```

Write a script

- Multiple regression (two predictors)

```
model3 = lm(endur$endurance ~ endur$age + endur$activeyears)

summary(model3)
```

endurance = 29.40 + -0.26(age) + 0.92(activeyears)

```
Call:
lm(formula = endur$endurance ~ endur$age + endur$activeyears)

Residuals:
     Min       1Q   Median       3Q      Max
-21.7994  -6.9040   0.5701   5.6326  27.2279

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        29.3952     3.2054   9.171  < 2e-16 ***
endur$age          -0.2571     0.0655  -3.925 0.000113 ***
endur$activeyears   0.9163     0.1386   6.610 2.44e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.919 on 242 degrees of freedom
Multiple R-squared: 0.1663, Adjusted R-squared: 0.1594
F-statistic: 24.14 on 2 and 242 DF,  p-value: 2.754e-10
```

Final script

```
# Statistics One, Lecture 9, example script
# Multiple regression analysis

library(psych)

endur <- read.table("STATS1.EX.04.txt", header = T)

# Change default settings for graphics
par(cex = 2, lwd = 2, col.axis = 200, col.lab = 200, col.main = 200, col.sub = 200, fg = 200)

#Scatterplots
plot(endur$endurance ~ endur$age, main = "Scatterplot", ylab = "Endurance", xlab = "Age")
abline(lm(endur$endurance ~ endur$age), col="blue")

plot(endur$endurance ~ endur$activeyears, main = "Scatterplot", ylab = "Endurance", xlab = "Active Years")
abline(lm(endur$endurance ~ endur$activeyears), col="blue")

# Regression analyses (unstandardized)
model1 = lm(endur$endurance~endur$age)
summary(model1)
model2 = lm(endur$endurance~endur$activeyears)
summary(model2)
model3 = lm(endur$endurance~endur$age + endur$activeyears)
summary(model3)
```