Lab9

2024-11-11

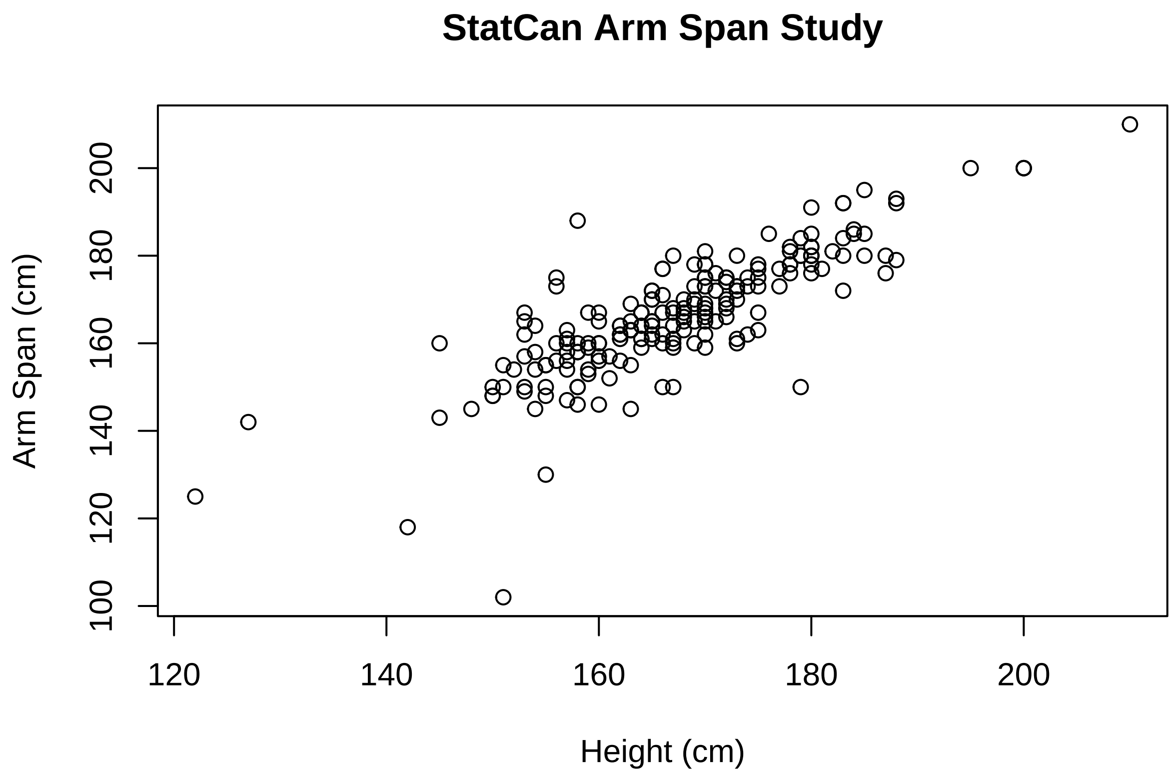
```
library(readr)
armspan <- read_csv("armspan.csv")
```

```
## Rows: 200 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (2): Height, ArmSpan
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

1.

Look at the scatterplot of the arm span and height values from the random sample of 200 students in Canada. What do you notice about the relationship between these two values?

```
plot(x = armspan$Height,
     y = armspan$ArmSpan,
     main="StatCan Arm Span Study",
     xlab="Height (cm)",
     ylab="Arm Span (cm)")
```




```
cor(armspan$Height, armspan$ArmSpan)
```

```
## [1] 0.8269918
```

The height and arm span appear to be positively linearly related to one another.

2.

Write the simple linear regression (SLR) model for this problem. Give the definition of the parameter values $\beta_0, \beta_1,$ and σ^2 in the context of the response and explanatory variables.

β_0:

β_1:

σ^2:

3.

Give the equation of the least squares regression line to predict the value of a student's arm span from their height.

```
slr <- lm(ArmSpan ~ Height, data=armspan)
summary(slr)
```

```
##
## Call:
## lm(formula = ArmSpan ~ Height, data = armspan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.596  -3.329   0.539   3.630  30.480
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.22306    8.01667   0.153    0.879
## Height       0.98922    0.04779  20.698   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.046 on 198 degrees of freedom
## Multiple R-squared:  0.6839, Adjusted R-squared:  0.6823
## F-statistic: 428.4 on 1 and 198 DF,  p-value: < 2.2e-16
```

$$\hat{y} = b_0 + b_1 x + \epsilon = 1.22306 + 0.98922x$$

4.

For the fitted SLR model, what is the interpretation of b1?

b1 is the model-estimated value of the parameter β_1

5.

Give the ANOVA Table for the SLR model. Use the ANOVA Table to conduct a test of significance for the SLR model.

```r
get.SS <- aov(ArmSpan ~ Height, data=armspan)
summary(get.SS)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Height        1  27738   27738   428.4 <2e-16 ***
## Residuals   198  12819      65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6.

Give the value of R^2 for the SLR model. Show this value is equal to the ratio of the SS_{Model} to SS_{Total} using the ANOVA Table. Give an interpretation of this value.

R^2: 0.6839

$\frac{SS_{Model}}{SS_{Total}} = \frac{SS_{Model}}{SS_{Model}+SS_{Residuals}}$ = 27738 / (27738 + 12819) = 0.6839263

"percent of variability in Y is captured using the X and Y model"

7.

Report the correlation coefficient between height and arm span. How does this value relate to the value of R^2 from the SLR model?

```
cor(armspan$Height, armspan$ArmSpan)
```

```
## [1] 0.8269918
```

Correlation Coefficient: 0.8269918

Correlation Coefficientˆ2 = R^2, as per: 0.8269918ˆ2 = 0.6839154

8.

Obtain the 95% confidence interval for the slope parameter in the SLR model. Give an interpretation of this interval.

```r
confint(slr)
```

```
##                   2.5 %    97.5 %
## (Intercept) -14.5859436 17.032065
## Height        0.8949754  1.083472
```

95% confident 0.8949754 to 1.083472

9.

Obtain a 95% confidence interval for the conditional mean arm span of all students in the population who are 170 cm tall. Give the interpretation of this interval.

```
predict.lm(slr, interval='confidence', newdata=data.frame(Height=170))
```

```
##        fit      lwr      upr
## 1 169.3911 168.2409 170.5413
```

95% confident 168.2409 to 170.5413

10.

Obtain a 95% prediction interval for the predicted arm span of a student in the population who is 188 cm tall. Give the interpretation of this interval.

```
predict.lm(slr, interval='prediction', newdata=data.frame(Height=188))
```

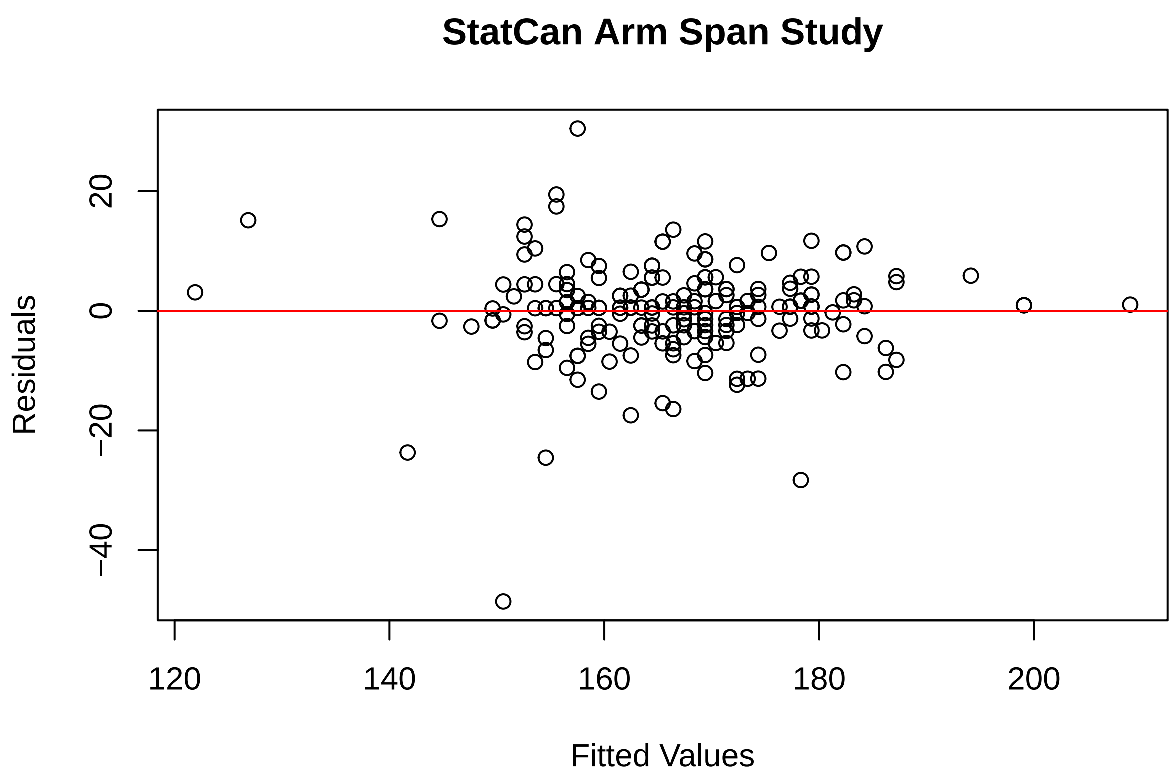
```
##        fit      lwr      upr
## 1 187.1971 171.1708 203.2234
```

95% confident 171.1708 to 203.2234

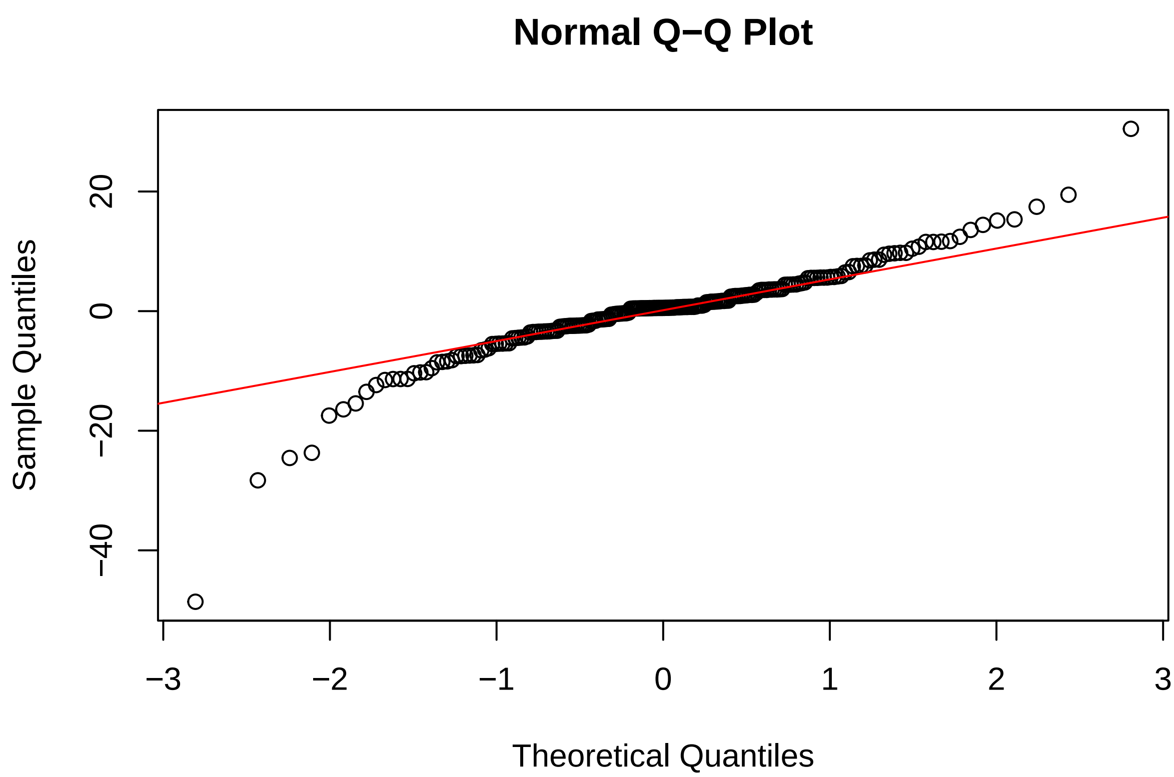
11.

Assuming that the independence and fixed-values-for-x assumptions are met, check the assumptions of linearity, constant variance, and normality. Summarize your findings.

```
plot(slr$fitted.values, slr$residuals, main="StatCan Arm Span Study",
xlab="Fitted Values", ylab="Residuals")
abline(h=0, col="red")
```

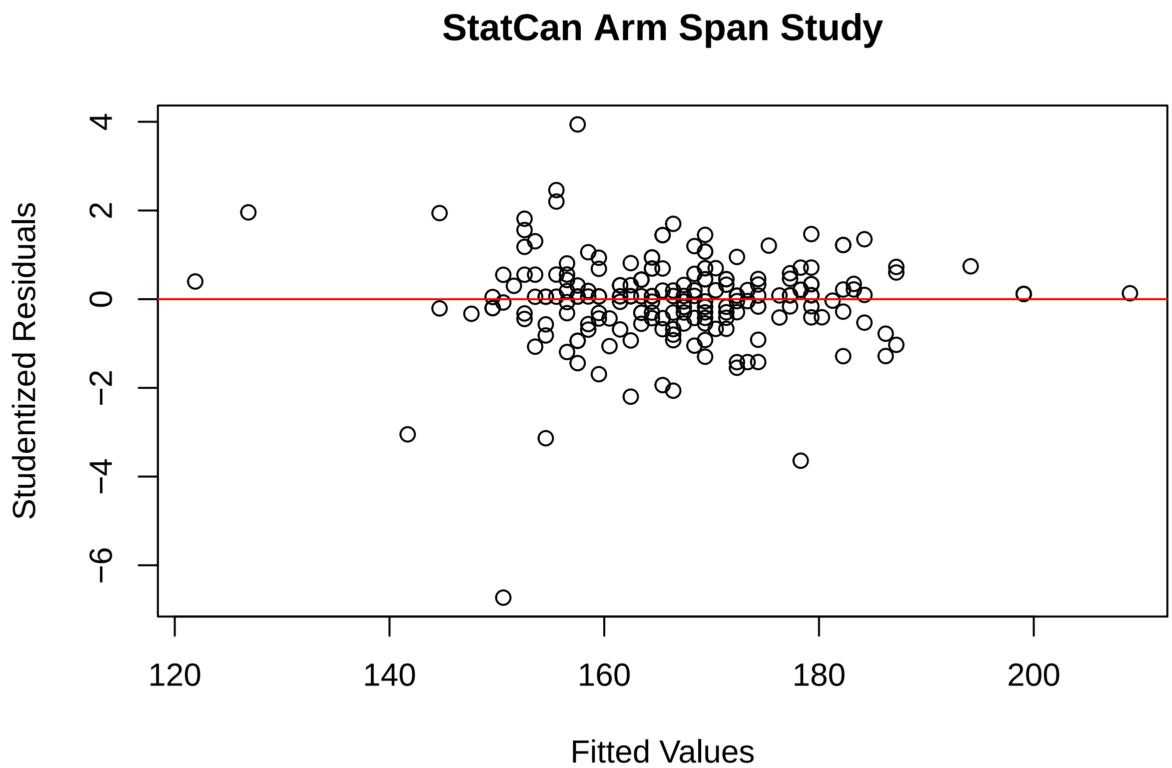



```
qqnorm(slr$residuals)
qqline(slr$residuals, col="red")
```

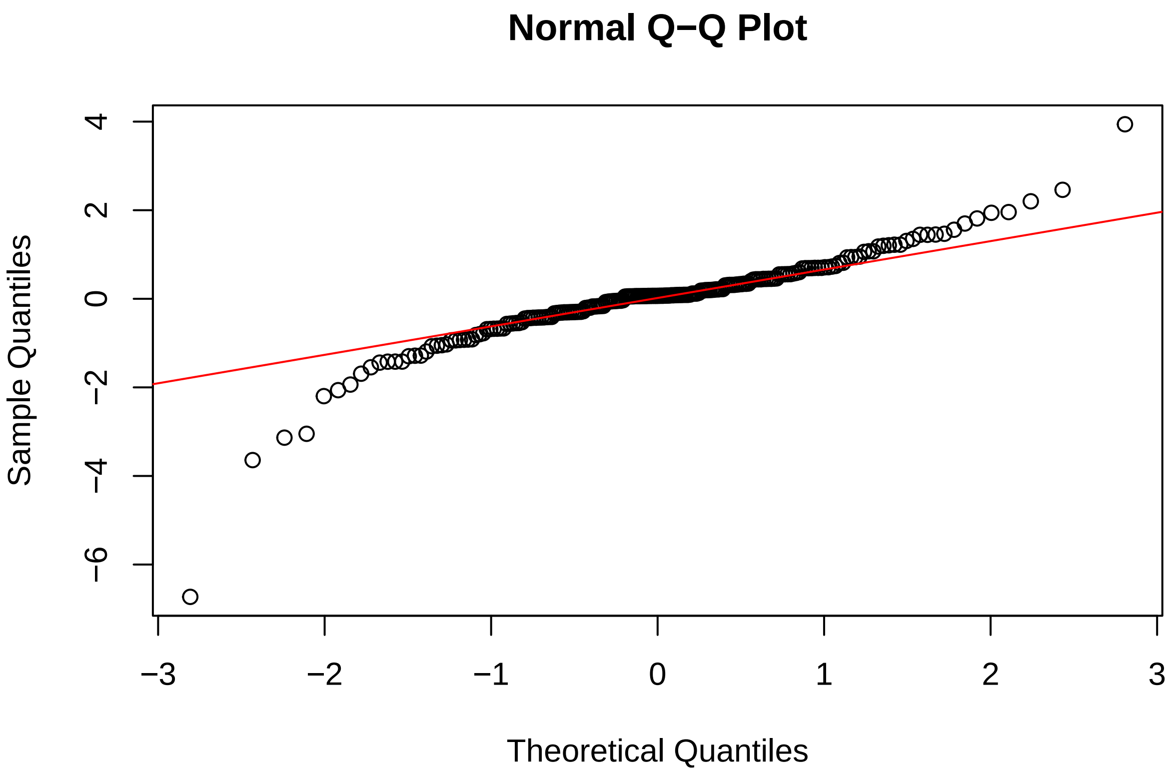



```r
library(MASS)
stdresids <- studres(slr)
```

```r
plot(slr$fitted.values, stdresids, main="StatCan Arm Span Study",
xlab="Fitted Values", ylab="Studentized Residuals")
abline(h=0, col="red")
```




```r
qqnorm(stdresids)
qqline(stdresids, col="red")
```



Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
4
2
0
−2
−4
−6
−3
−2
−1
0
1
2
3

12.

Conduct the F-test for lack-of-fit and report the results.

```
cell.means <- aov(ArmSpan ~ as.factor(Height), data=armspan)
summary(cell.means)
```

```
##                    Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(Height)  45  31054   690.1   11.18 <2e-16 ***
## Residuals         154   9503    61.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(slr, cell.means)
```

```
## Analysis of Variance Table
##
## Model 1: ArmSpan ~ Height
## Model 2: ArmSpan ~ as.factor(Height)
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1    198 12819.4
## 2    154  9503.4 44    3316.1 1.2213 0.1883
```