R08- Multiple Regression Examples

HCI/PSYCH 522
Iowa State University

April 7, 2022

Overview

- Children's heights depending on
  - mother's heights
  - + father's heights
  - + gender
- Wool breaks depending on
  - tension (L, M, H)
  - + type (A, B)

Multiple regression

Recall the simple linear regression model is

$$Y_i \overset{ind}{\sim} N(\mu_i, \sigma^2), \quad \mu_i = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}$$

where for observation i

- Y_i is the dependent variable and
- $X_{i,p}$ is the p^{th} independent variable.

Interpretation

- β_0 is mean of the dependent variable when all independent variables (X's) are 0
- β_p for $p \neq 0$, is the mean increase in the dependent variable for each unit increase in the associated independent variable
- R^2 is the proportion of variability in the dependent variable explained by the model

heights

```
ex0726
```

```
##     Gender Family Height Father Mother
## 1     male      1   73.2   78.5   67.0
## 2   female      1   69.2   78.5   67.0
## 3   female      1   69.0   78.5   67.0
## 4   female      1   69.0   78.5   67.0
## 5     male      2   73.5   75.5   66.5
## 6     male      2   72.5   75.5   66.5
## 7   female      2   65.5   75.5   66.5
## 8   female      2   65.5   75.5   66.5
## 9     male      3   71.0   75.0   64.0
## 10  female      3   68.0   75.0   64.0
## 11    male      4   70.5   75.0   64.0
## 12    male      4   68.5   75.0   64.0
## 13  female      4   67.0   75.0   64.0
## 14  female      4   64.5   75.0   64.0
## 15  female      4   63.0   75.0   64.0
## 16    male      5   72.0   75.0   58.5
## 17    male      5   69.0   75.0   58.5
```

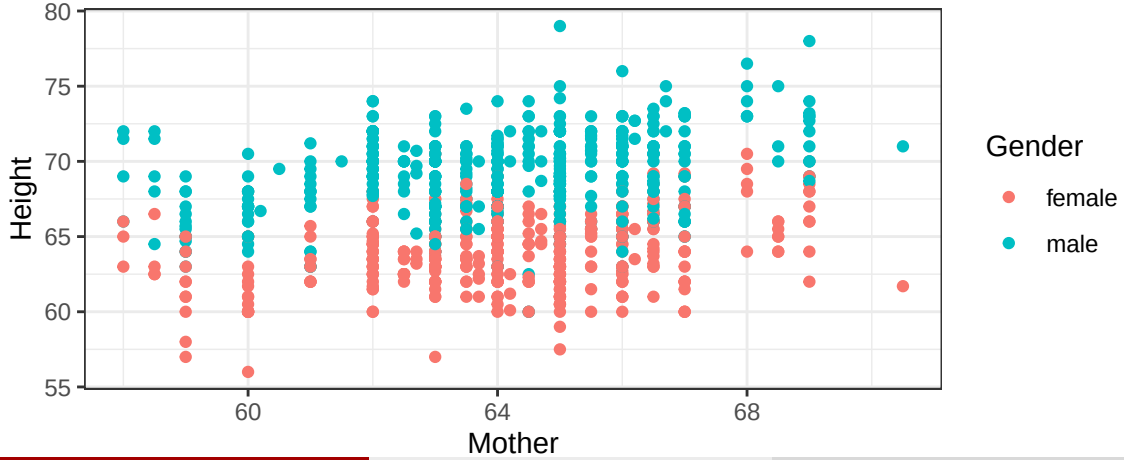
heights

```
summary(ex0726)
```

```
##     Gender        Family          Height          Father         Mother
##  female:453   Min.   :  1.0   Min.   :56.00   Min.   :62.0   Min.   :58.00
##  male  :480   1st Qu.: 59.0   1st Qu.:64.00   1st Qu.:68.0   1st Qu.:63.00
##               Median :106.0   Median :66.50   Median :69.0   Median :64.00
##               Mean   :106.2   Mean   :66.74   Mean   :69.2   Mean   :64.09
##               3rd Qu.:158.0   3rd Qu.:69.70   3rd Qu.:71.0   3rd Qu.:65.50
##               Max.   :205.0   Max.   :79.00   Max.   :78.5   Max.   :70.50
```

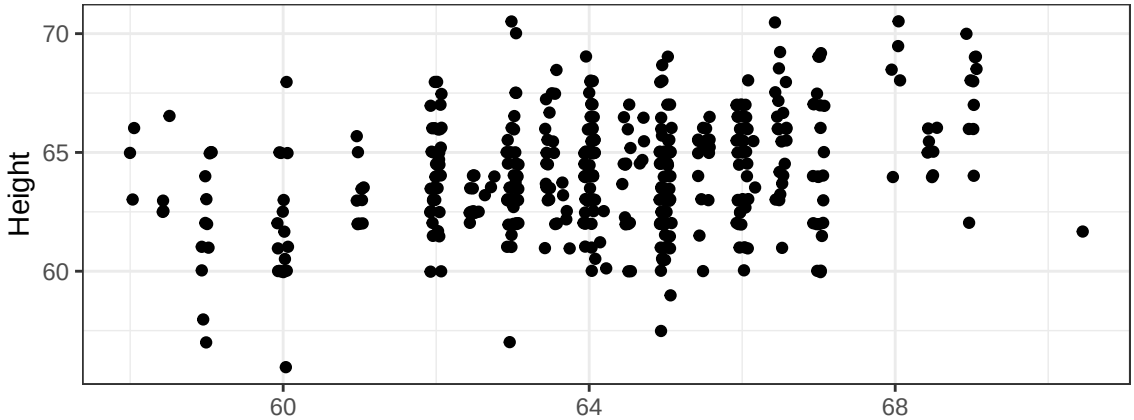
heights

```
ggplot(ex0726, aes(x=Mother, y=Height, color = Gender)) + geom_point()
```



heights

```
ggplot(ex0726 %>% filter(Gender == "female"),
       aes(x=Mother, y=Height)) + geom_jitter()
```

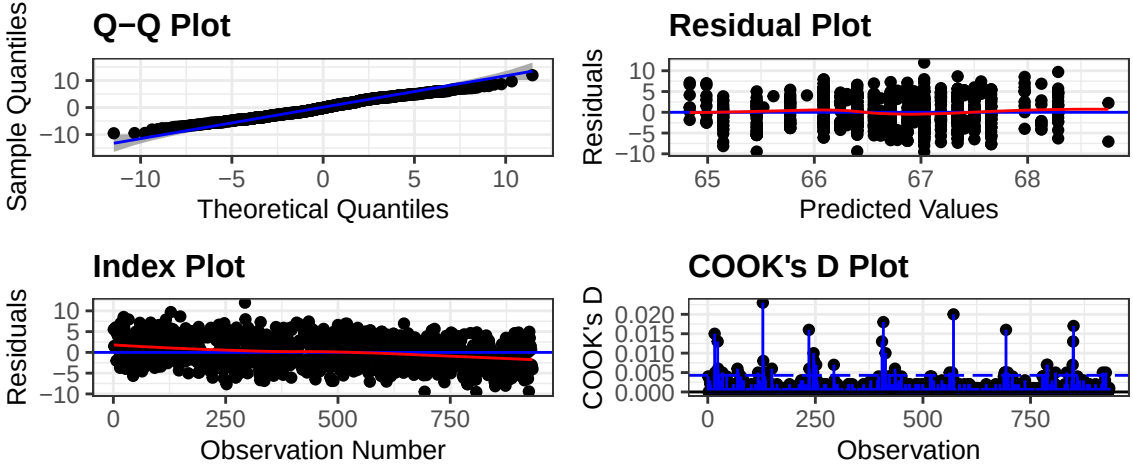


```
m <- lm(Height ~ Mother, data = ex0726 %>% filter(Gender == "female"),)
summary(m)

##
## Call:
## lm(formula = Height ~ Mother, data = ex0726 %>% filter(Gender ==
##     "female"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8739 -1.5331  0.0813  1.4445  6.7629
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 43.67884    3.00214  14.549  < 2e-16 ***
## Mother       0.31839    0.04677   6.807 3.18e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.246 on 451 degrees of freedom
## Multiple R-squared:  0.09318,Adjusted R-squared:  0.09117
## F-statistic: 46.34 on 1 and 451 DF,  p-value: 3.181e-11
```

heights

```
m <- lm(Height ~ I(Mother-64), data = ex0726)
```

```
resid_panel(m, plots = c("qq","resid","index","cookd"), smooth = TRUE, qqbands = TRUE)
```



```
summary(m)

##
## Call:
## lm(formula = Height ~ I(Mother - 64), data = ex0726)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.530 -2.629 -0.146  2.871 11.970
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    66.71614    0.11499 580.185  < 2e-16 ***
## I(Mother - 64)  0.31403    0.05019   6.256 6.01e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.51 on 931 degrees of freedom
## Multiple R-squared:  0.04034,Adjusted R-squared:  0.03931
## F-statistic: 39.14 on 1 and 931 DF,  p-value: 6.006e-10
```

heights

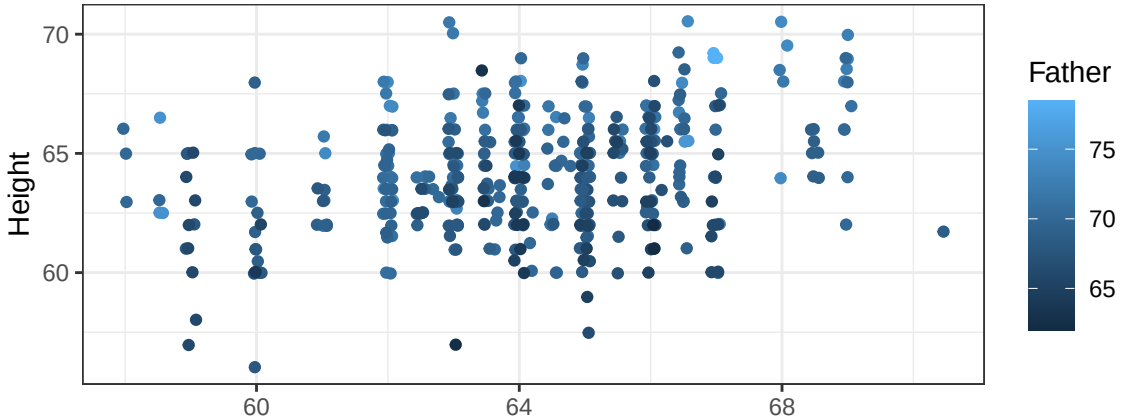
```
co <- coef(m)
co

##    (Intercept) I(Mother - 64)
##     66.7161395      0.3140259

ci <- confint(m)
ci

##                    2.5 %     97.5 %
## (Intercept)    66.490468 66.9418112
## I(Mother - 64)  0.215519  0.4125328
```

heights

- When the mother's height is 64 in, the mean daughter's height is 66.7 in (66.5,66.9).
- For each inch increase in mother's height, the mean daughter's height increases by 0.31 (0.22,0.41) inches.
- The model with mother's height explains 4% of the variability in daughter's height.

heights

```
ggplot(ex0726 %>% filter(Gender == "female"),
       aes(x=Mother, y=Height, color = Father)) + geom_jitter()
```

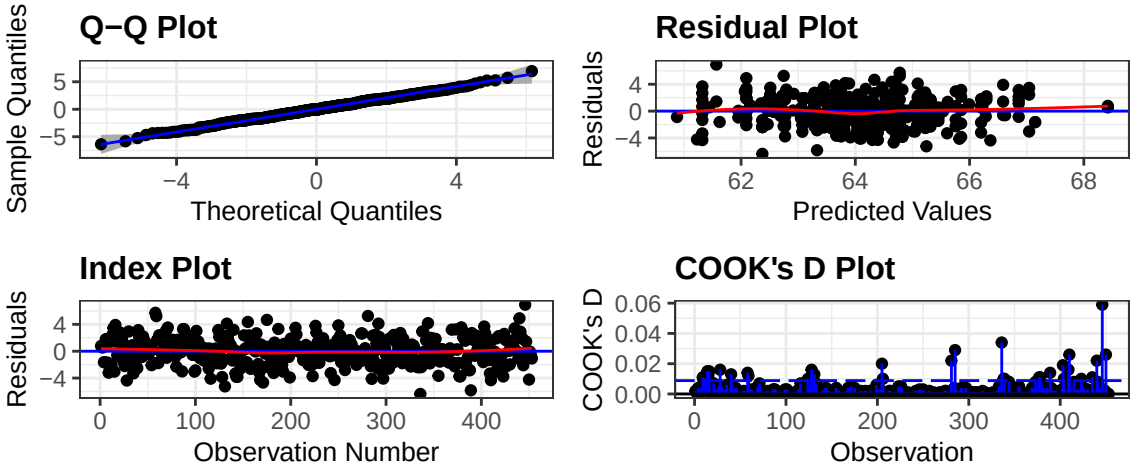


```
m <- lm(Height ~ Mother + Father, data = ex0726 %>% filter(Gender == "female"))
summary(m)

##
## Call:
## lm(formula = Height ~ Mother + Father, data = ex0726 %>% filter(Gender ==
##     "female"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3726 -1.4040 -0.0423  1.4130  6.9325
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.75770    3.60782   5.199 3.04e-07 ***
## Mother       0.30358    0.04206   7.218 2.27e-12 ***
## Father       0.37353    0.03590  10.406  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.019 on 450 degrees of freedom
## Multiple R-squared:  0.2691,Adjusted R-squared:  0.2658
## F-statistic: 82.82 on 2 and 450 DF,  p-value: < 2.2e-16
```

heights

```
m <- lm(Height ~ I(Mother-64) + I(Father-69), data = ex0726 %>% filter(Gender == "female"))
```

```
resid_panel(m, plots = c("qq","resid","index","cookd"), smooth = TRUE, qqbands = TRUE)
```



```
summary(m)

##
## Call:
## lm(formula = Height ~ I(Mother - 64) + I(Father - 69), data = ex0726 %>%
##     filter(Gender == "female"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3726 -1.4040 -0.0423  1.4130  6.9325
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    63.96048    0.09549 669.786  < 2e-16 ***
## I(Mother - 64)  0.30358    0.04206   7.218 2.27e-12 ***
## I(Father - 69)  0.37353    0.03590  10.406  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.019 on 450 degrees of freedom
## Multiple R-squared:  0.2691,Adjusted R-squared:  0.2658
## F-statistic: 82.82 on 2 and 450 DF,  p-value: < 2.2e-16
```

heights

```
co <- coef(m)
co
```

```
##    (Intercept) I(Mother - 64) I(Father - 69)
##     63.9604846      0.3035837      0.3735279
```

```
ci <- confint(m)
ci
```

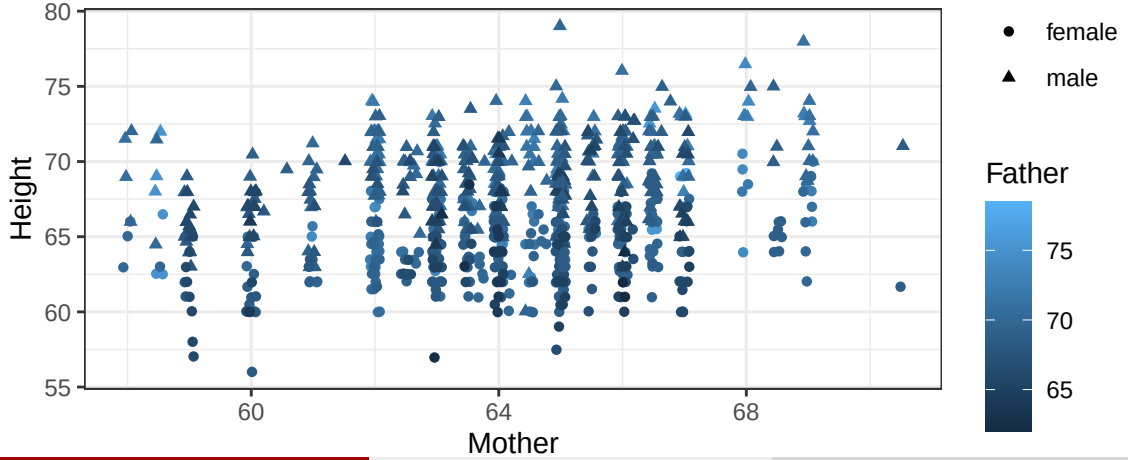
```
##                     2.5 %     97.5 %
## (Intercept)    63.7728152 64.1481540
## I(Mother - 64)  0.2209226  0.3862448
## I(Father - 69)  0.3029841  0.4440717
```

heights

- When mother's height is 64 in and father's height is 69 in, the mean daughter's height is 64 in (63.8,64.1).
- For each inch increase in mother's height, the mean children's height increases by 0.3 (0.22,0.39) inches while holding father's height constant.
- For each inch increase in father's height, the mean daughter's height increases by 0.37 (0.3,0.44) inches while holding mother's height constant.
- The model with mother's and father's height explains 27% of the variability in daughter's height.

heights

```
ggplot(ex0726, aes(x=Mother, y=Height, color = Father, shape=Gender)) + geom_jitter()
```

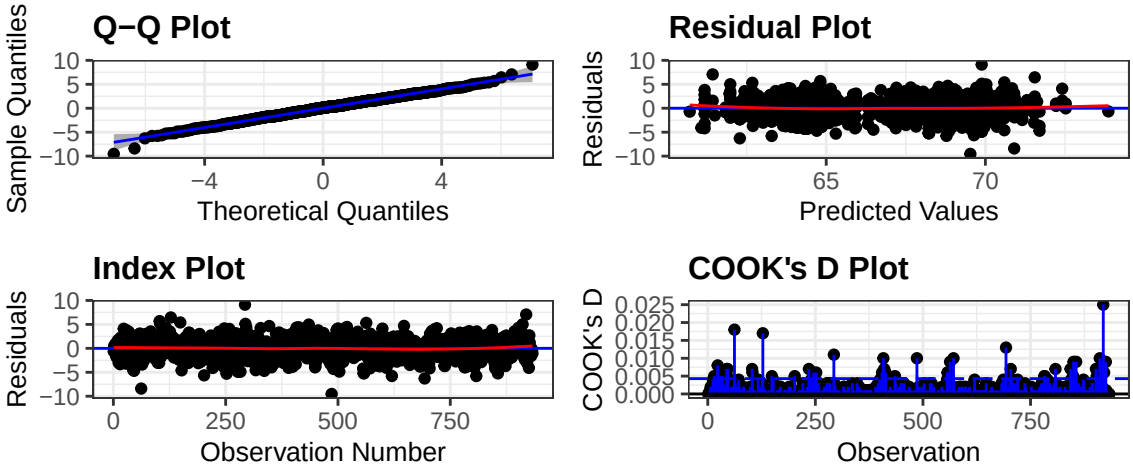


```
m <- lm(Height ~ Mother + Father + Gender, data = ex0726)
summary(m)

##
## Call:
## lm(formula = Height ~ Mother + Father + Gender, data = ex0726)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5280 -1.4604  0.0996  1.4783  9.1161
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.43221    2.72802   6.023 2.46e-09 ***
## Mother       0.31840    0.03102  10.263  < 2e-16 ***
## Father       0.39339    0.02868  13.718  < 2e-16 ***
## Gendermale   5.21902    0.14188  36.784  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.165 on 929 degrees of freedom
## Multiple R-squared:  0.6358,Adjusted R-squared:  0.6346
## F-statistic: 540.5 on 3 and 929 DF,  p-value: < 2.2e-16
```

heights

```
m <- lm(Height ~ I(Mother-64) + I(Father-69) + Gender, data = ex0726)
```

```
resid_panel(m, plots = c("qq","resid","index","cookd"), smooth = TRUE, qqbands = TRUE)
```



```
summary(m)

##
## Call:
## lm(formula = Height ~ I(Mother - 64) + I(Father - 69) + Gender,
##     data = ex0726)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5280 -1.4604  0.0996  1.4783  9.1161
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    63.95311    0.10206  626.61   <2e-16 ***
## I(Mother - 64)  0.31840    0.03102   10.26   <2e-16 ***
## I(Father - 69)  0.39339    0.02868   13.72   <2e-16 ***
## Gendermale      5.21902    0.14188   36.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.165 on 929 degrees of freedom
## Multiple R-squared:  0.6358,Adjusted R-squared:  0.6346
## F-statistic: 540.5 on 3 and 929 DF,  p-value: < 2.2e-16
```

heights

```
co <- coef(m)
co

##    (Intercept) I(Mother - 64) I(Father - 69)     Gendermale
##     63.9531126      0.3183957      0.3933851      5.2190191

ci <- confint(m)
ci

##                     2.5 %     97.5 %
## (Intercept)    63.7528127 64.1534126
## I(Mother - 64)  0.2575128  0.3792787
## I(Father - 69)  0.3371065  0.4496638
## Gendermale      4.9405709  5.4974674
```

heights

- When the mother's height is 64 in and father's height is 69 in, the mean daughter's height is 64 in (63.8,64.2).
- For each inch increase in mother's height, the mean daughter's height increases by 0.32 (0.26,0.38) inches while holding father's height and gender constant.
- For each inch increase in father's height, the mean daughter's height increases by 0.39 (0.34,0.45) inches while holding mother's height and gender constant.
- Male children are, on average, 0.39 (0.34,0.45) inches taller than female children while holding mother's and father's height constant.
- The model with mother's height, father's height, and gender explains 64% of the variability in children's height.

warpbreaks

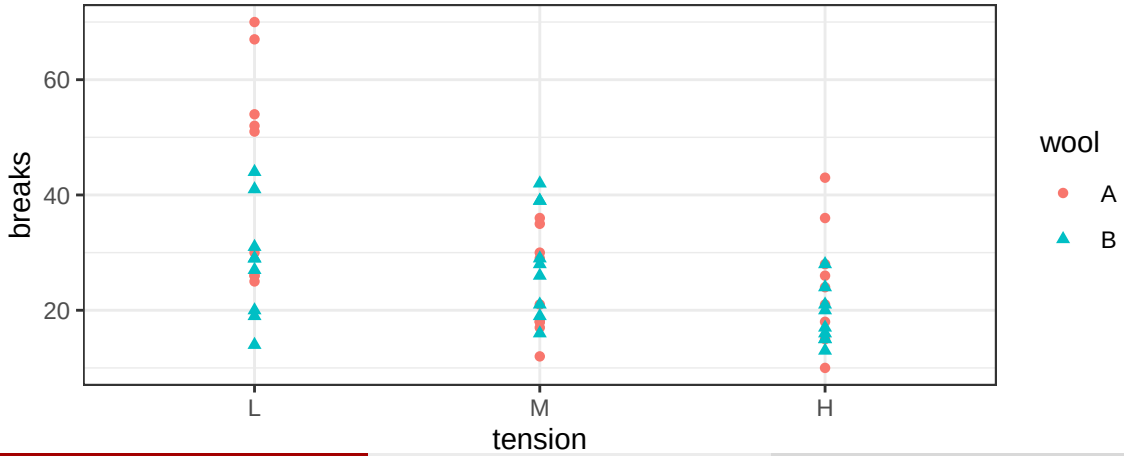
```
warpbreaks

##    breaks wool tension
## 1      26    A       L
## 2      30    A       L
## 3      54    A       L
## 4      25    A       L
## 5      70    A       L
## 6      52    A       L
## 7      51    A       L
## 8      26    A       L
## 9      67    A       L
## 10     18    A       M
## 11     21    A       M
## 12     29    A       M
## 13     17    A       M
## 14     12    A       M
## 15     18    A       M
## 16     35    A       M
## 17     30    A       M
```

warpbreaks

```
summary(warpbreaks)

##      breaks       wool   tension
##  Min.   :10.00   A:27   L:18
##  1st Qu.:18.25   B:27   M:18
##  Median :26.00          H:18
##  Mean   :28.15
##  3rd Qu.:34.00
##  Max.   :70.00
```

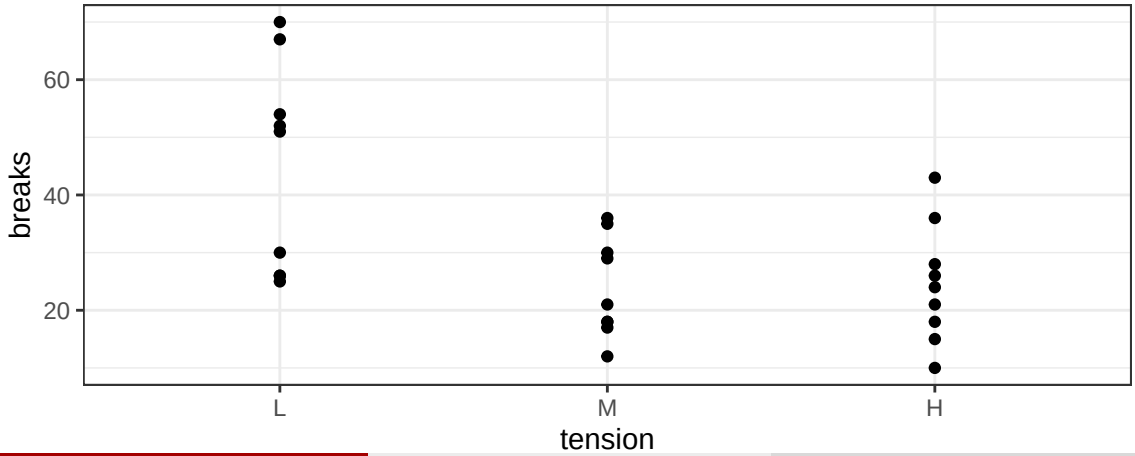
warpbreaks

```
ggplot(warpbreaks, aes(x=tension, y=breaks, color=wool, shape=wool)) + geom_point()
```



warpbreaks

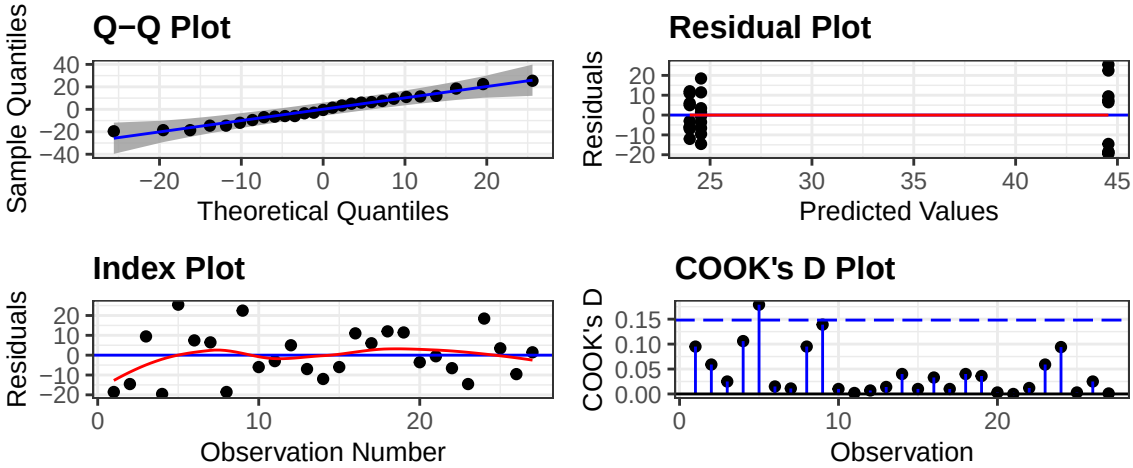
```
ggplot(warpbreaks %>% filter(wool == "A"), aes(x=tension, y=breaks)) + geom_point()
```



warpbreaks

```
m <- lm(breaks ~ tension, data = warpbreaks %>% filter(wool == "A"))
```

```
resid_panel(m, plots = c("qq","resid","index","cookd"), smooth = TRUE, qqbands = TRUE)
```



```
summary(m)

##
## Call:
## lm(formula = breaks ~ tension, data = warpbreaks %>% filter(wool ==
##     "A"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.5556  -8.2778  -0.5556   8.4444  25.4444
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   44.556      4.338  10.271 2.91e-10 ***
## tensionM     -20.556      6.135  -3.351  0.00266 **
## tensionH     -20.000      6.135  -3.260  0.00332 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.01 on 24 degrees of freedom
## Multiple R-squared:  0.3779,Adjusted R-squared:  0.326
## F-statistic: 7.288 on 2 and 24 DF,  p-value: 0.003363
```

warpbreaks

```
co <- coef(m)
co

## (Intercept)    tensionM    tensionH
##    44.55556   -20.55556   -20.00000

ci <- confint(m)
ci

##                 2.5 %    97.5 %
## (Intercept)  35.60269 53.508421
## tensionM    -33.21682 -7.894291
## tensionH    -32.66126 -7.338736
```

warpbreaks

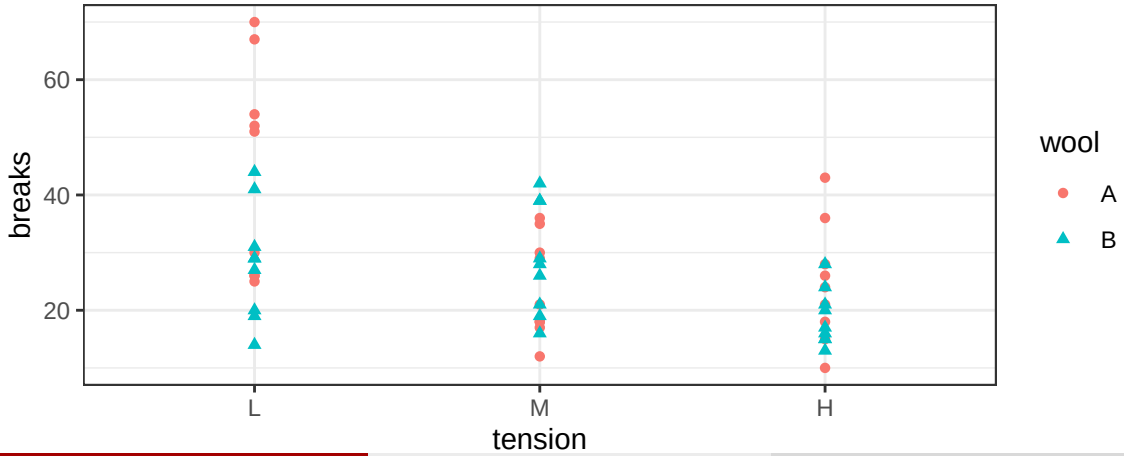
- For wool type A when tension is low, the mean number of breaks is 45 in (36,54).
- For wool type A when moving from tension low to tension medium, the mean number of breaks is 21 (8,33) lower.
- For wool type A when moving from tension low to tension high, the mean number of breaks is 20 (7,33) lower.
- The model with tension 38% of the variability in number of breaks for Wool type A.

warpbreaks

```
em <- emmeans(m, pairwise ~ tension, adjust = "none")
confint(em)

## $emmeans
##  tension emmean   SE df lower.CL upper.CL
##  L         44.6 4.34 24     35.6     53.5
##  M         24.0 4.34 24     15.0     33.0
##  H         24.6 4.34 24     15.6     33.5
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate   SE df lower.CL upper.CL
##  L - M      20.556 6.13 24     7.89     33.2
##  L - H      20.000 6.13 24     7.34     32.7
##  M - H      -0.556 6.13 24   -13.22     12.1
##
## Confidence level used: 0.95
```

warpbreaks

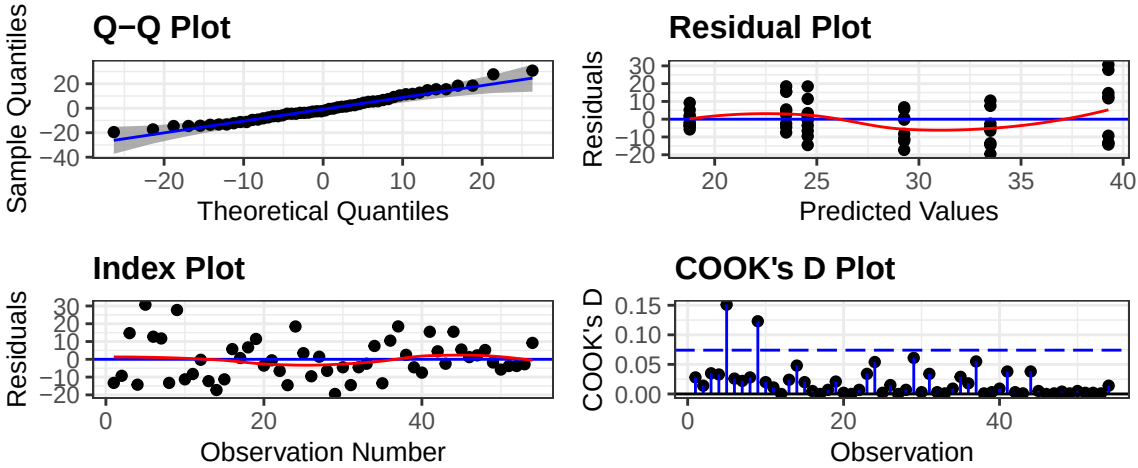
```
ggplot(warpbreaks, aes(x=tension, y=breaks, color=wool, shape=wool)) + geom_point()
```



warpbreaks

```
m <- lm(breaks ~ tension + wool, data = warpbreaks)
```

```
resid_panel(m, plots = c("qq","resid","index","cookd"), smooth = TRUE, qqbands = TRUE)
```



```
summary(m)

##
## Call:
## lm(formula = breaks ~ tension + wool, data = warpbreaks)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.500  -8.083  -2.139   6.472  30.722
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   39.278      3.162  12.423  < 2e-16 ***
## tensionM     -10.000      3.872  -2.582 0.012787 *
## tensionH     -14.722      3.872  -3.802 0.000391 ***
## woolB         -5.778      3.162  -1.827 0.073614 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.62 on 50 degrees of freedom
## Multiple R-squared:  0.2691,Adjusted R-squared:  0.2253
## F-statistic: 6.138 on 3 and 50 DF,  p-value: 0.00123
```

warpbreaks

```
co <- coef(m)
co

## (Intercept)    tensionM    tensionH       woolB
##   39.277778  -10.000000  -14.722222   -5.777778

ci <- confint(m)
ci

##                 2.5 %     97.5 %
## (Intercept)  32.92715 45.6284061
## tensionM    -17.77790 -2.2221006
## tensionH    -22.50012 -6.9443228
## woolB       -12.12841  0.5728505
```

warpbreaks

- For wool type A when tension is low, the mean number of breaks is 39 in (33,46).
- When moving from tension low to tension medium, the mean number of breaks is 10 (2,18) lower while holding wool type constant.
- When moving from tension low to tension high, the mean number of breaks is 15 (7,23) lower while holding wool type constant.
- On average, wool type B has 6 (-1,12) fewer breaks than wool type A.
- The model with tension and wool type explains 27% of the variability in number of breaks.

warpbreaks

```
em <- emmeans(m, pairwise ~ tension, adjust = "none")
confint(em)

## $emmeans
##  tension emmean   SE df lower.CL upper.CL
##  L         36.4 2.74 50     30.9     41.9
##  M         26.4 2.74 50     20.9     31.9
##  H         21.7 2.74 50     16.2     27.2
##
## Results are averaged over the levels of: wool
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate   SE df lower.CL upper.CL
##  L - M       10.00 3.87 50     2.22     17.8
##  L - H       14.72 3.87 50     6.94     22.5
##  M - H        4.72 3.87 50    -3.06     12.5
##
## Results are averaged over the levels of: wool
## Confidence level used: 0.95
```

warpbreaks

```
em <- emmeans(m, pairwise ~ wool, adjust = "none")
confint(em)

## $emmeans
##  wool emmean   SE df lower.CL upper.CL
##  A      31.0 2.24 50     26.5     35.5
##  B      25.3 2.24 50     20.8     29.7
##
## Results are averaged over the levels of: tension
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate   SE df lower.CL upper.CL
##  A - B        5.78 3.16 50   -0.573     12.1
##
## Results are averaged over the levels of: tension
## Confidence level used: 0.95
```