THE GRADES OF CLIMBING ROUTES

Suraaj Shrestha

CONTENTS

ABOUT DATASET

- 10,927 rows and 16 columns
- Stratified Sampling grouped by frequency of sex to 500 samples.
- Check for null values and drop irrelevant columns
- Columns: grades_mean, grades_first, years_cl, sex, age, height, weight
- Retrieved from Kaggle

RESEARCH SCENARIO

- Find the best model to predict grades_mean:
  - Linear Regression
  - Multiple Regression
- Run tests to check significance of the model and inference of the parameters.

LINEAR REGRESSION

- grades_mean ~ grades_first

```
Call:
lm(formula = grades_mean ~ grades_first, data = data1)

Residuals:
     Min       1Q   Median       3Q      Max
-10.9454  -3.0890  -0.1785   2.6601  16.6071

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.30534    0.95758   14.94   <2e-16 ***
grades_first  0.68840    0.02036   33.81   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.314 on 498 degrees of freedom
Multiple R-squared:  0.6965,	Adjusted R-squared:  0.6959
F-statistic:  1143 on 1 and 498 DF,  p-value: < 2.2e-16
```

1. Set up the hypotheses and select the alpha level

$$H_0\text{: } \boldsymbol{\beta_n = 0}\ (\boldsymbol{no\ linear\ association})$$
$$H_1\text{: } \boldsymbol{\beta_n \neq 0}\ (\boldsymbol{there\ is\ linear\ association})$$
$$\alpha = 0.05$$

2. Select the appropriate test statistic

$$Degrees\ of\ Freedom\text{: } df1 = 4, df2 = 498$$
$$\mathrm{F-Statistics} = \frac{Mean\ SS\ of\ regression}{Mean\ SS\ of\ residual}$$

> qf(0.05, df1 = 1, df2 = 498, lower.tail = FALSE)
[1] 3.860199

3. State the decision rule

Reject H_0 if $F \geq 3.86$

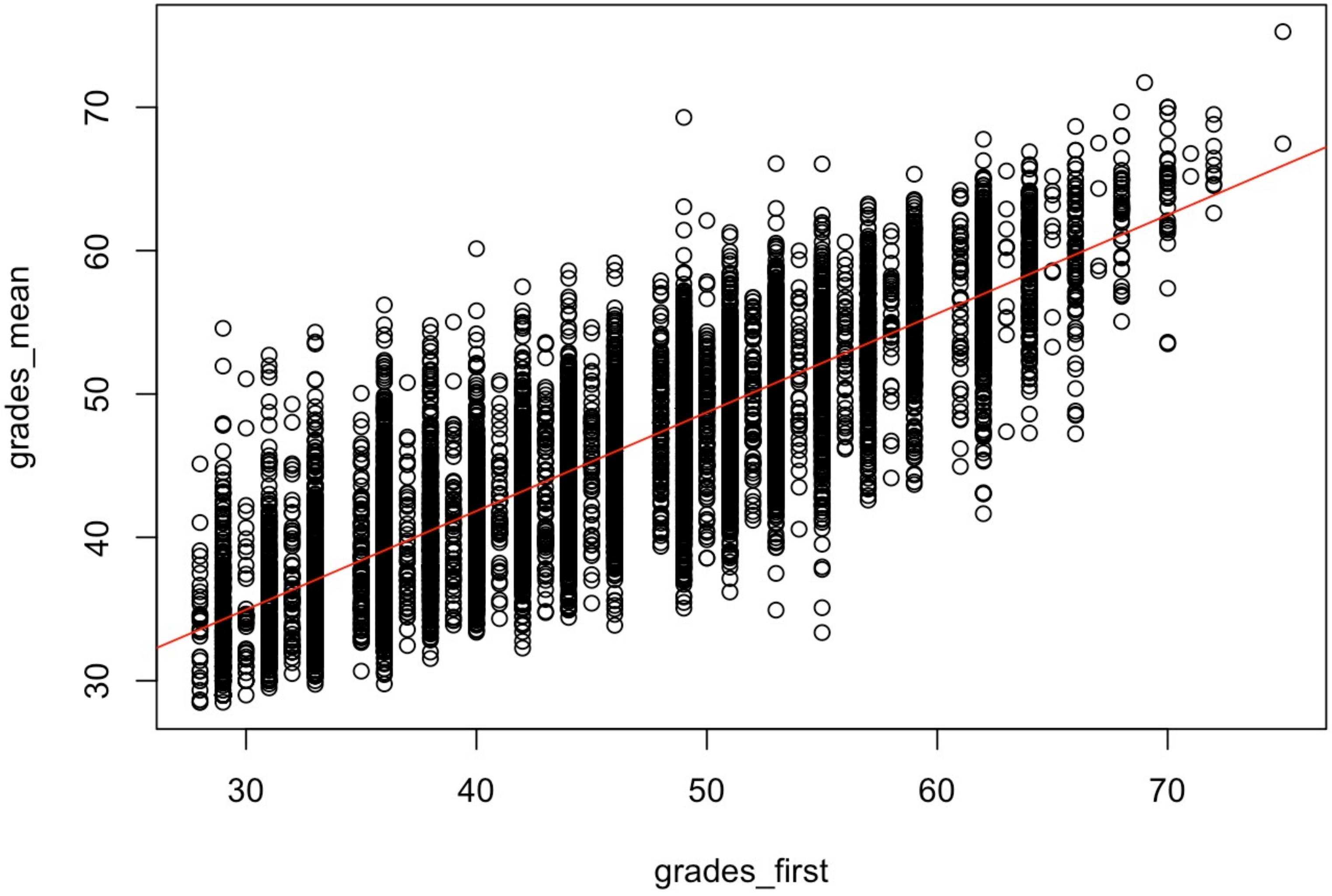
4. Compute the test statistic

$$\mathrm{F-Statistics} = \frac{Mean\ SS\ of\ regression}{Mean\ SS\ of\ residual} = 1143$$

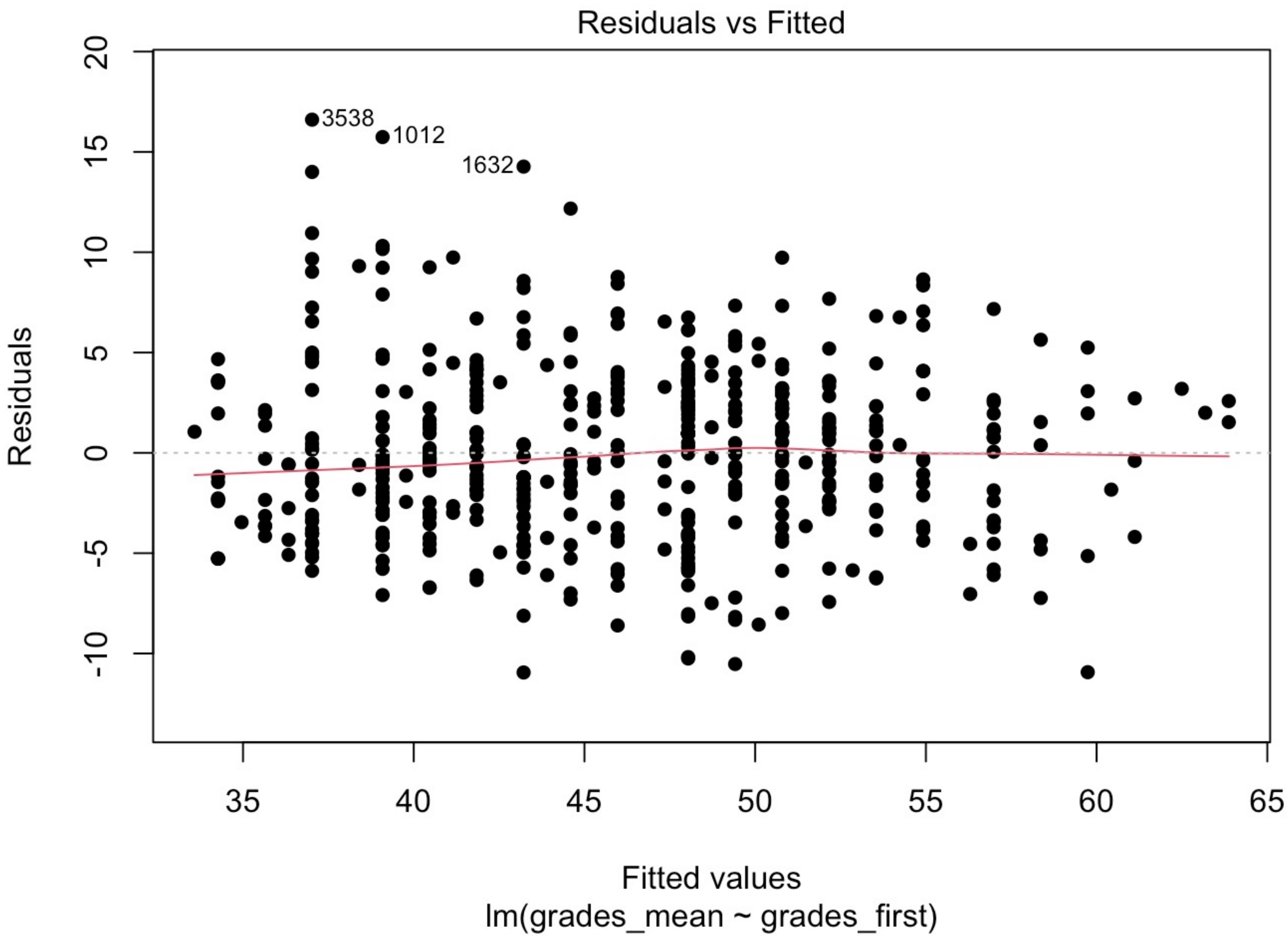
5. Conclusion

$1143 > 3.86$. We reject H_0 null hypothesis and conclude there is a linear relationship between grades_mean and grades_first.

LINEAR REGRESSION



LINEAR REGRESSION



MULTIPLE REGRESSION

grades_mean ~ grades_first + years_cl + sex + age + height + weight

```
lm(formula = grades_mean ~ grades_first + years_cl + sex + age +
    height + weight, data = data1)

Residuals:
     Min       1Q   Median       3Q      Max
-10.6642  -2.6359   0.0103   2.4644  14.8978

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.32373    4.71210   3.464 0.000578 ***
grades_first  0.63446    0.02219  28.593  < 2e-16 ***
years_cl      0.15853    0.04148   3.822 0.000149 ***
sex          -2.34752    0.68278  -3.438 0.000635 ***
age          -0.13967    0.03128  -4.465 9.92e-06 ***
height        0.06436    0.03152   2.042 0.041719 *
weight       -0.11785    0.02975  -3.961 8.56e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.097 on 493 degrees of freedom
Multiple R-squared:  0.7291,    Adjusted R-squared:  0.7258
F-statistic: 221.1 on 6 and 493 DF,  p-value: < 2.2e-16
```

grades_mean ~ grades_first + years_cl + sex + age + weight

```
lm(formula = grades_mean ~ grades_first + years_cl + sex + age +
    weight, data = data1)

Residuals:
     Min       1Q   Median       3Q      Max
-10.6736  -2.5987  -0.0984   2.4723  15.0598

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  25.14276    1.88890  13.311  < 2e-16 ***
grades_first  0.63360    0.02226  28.468  < 2e-16 ***
years_cl      0.15962    0.04161   3.836 0.000141 ***
sex          -2.67722    0.66553  -4.023 6.65e-05 ***
age          -0.14077    0.03137  -4.487 8.99e-06 ***
weight       -0.07896    0.02293  -3.444 0.000623 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.11 on 494 degrees of freedom
Multiple R-squared:  0.7268,    Adjusted R-squared:  0.724
F-statistic: 262.8 on 5 and 494 DF,  p-value: < 2.2e-16
```

MULTIPLE REGRESSION

1. Set up the hypotheses and select the alpha level

$$H_0: \ \boldsymbol{\beta_n = 0 \ (the \ model \ is \ not \ significant)}$$
$$H_1: \ \boldsymbol{\beta_n \neq 0 \ (the \ model \ is \ significant)}$$
$$\alpha = 0.05$$

2. Select the appropriate test statistic

$$Degrees \ of \ Freedom: \ df1 = 5, df2 = 494$$
$$\mathrm{F-Statistics} = \frac{Mean \ SS \ of \ regression}{Mean \ SS \ of \ residual}$$

> qf(0.05, df1 = 5, df2 = 494, lower.tail = FALSE)
[1] 2.232261

3. State the decision rule

Reject H_0 if $F \geq 2.23$

4. Compute the test statistic

$$\mathrm{F-Statistics} = \frac{Mean \ SS \ of \ regression}{Mean \ SS \ of \ residual} = 262.8$$

5. Conclusion

$262.8 > 3.86$. We reject H_0 null hypothesis and conclude the model is significant.

1. Set up the hypotheses and select the alpha level

$$H_0: \ \boldsymbol{\beta_{grades\_first} = 0}$$
$$H_1: \ \boldsymbol{\beta_{grades\_first} \neq 0}$$
$$\alpha = 0.05$$

2. Select the appropriate test statistic

$$Degrees \ of \ Freedom: \ df = 494$$

> qt(0.975, df = 494)
[1] 1.964778

3. State the decision rule

Reject H_0 if $|\mathrm{t}| \geq 1.96$ or $|\mathrm{t}| \leq -1.96$

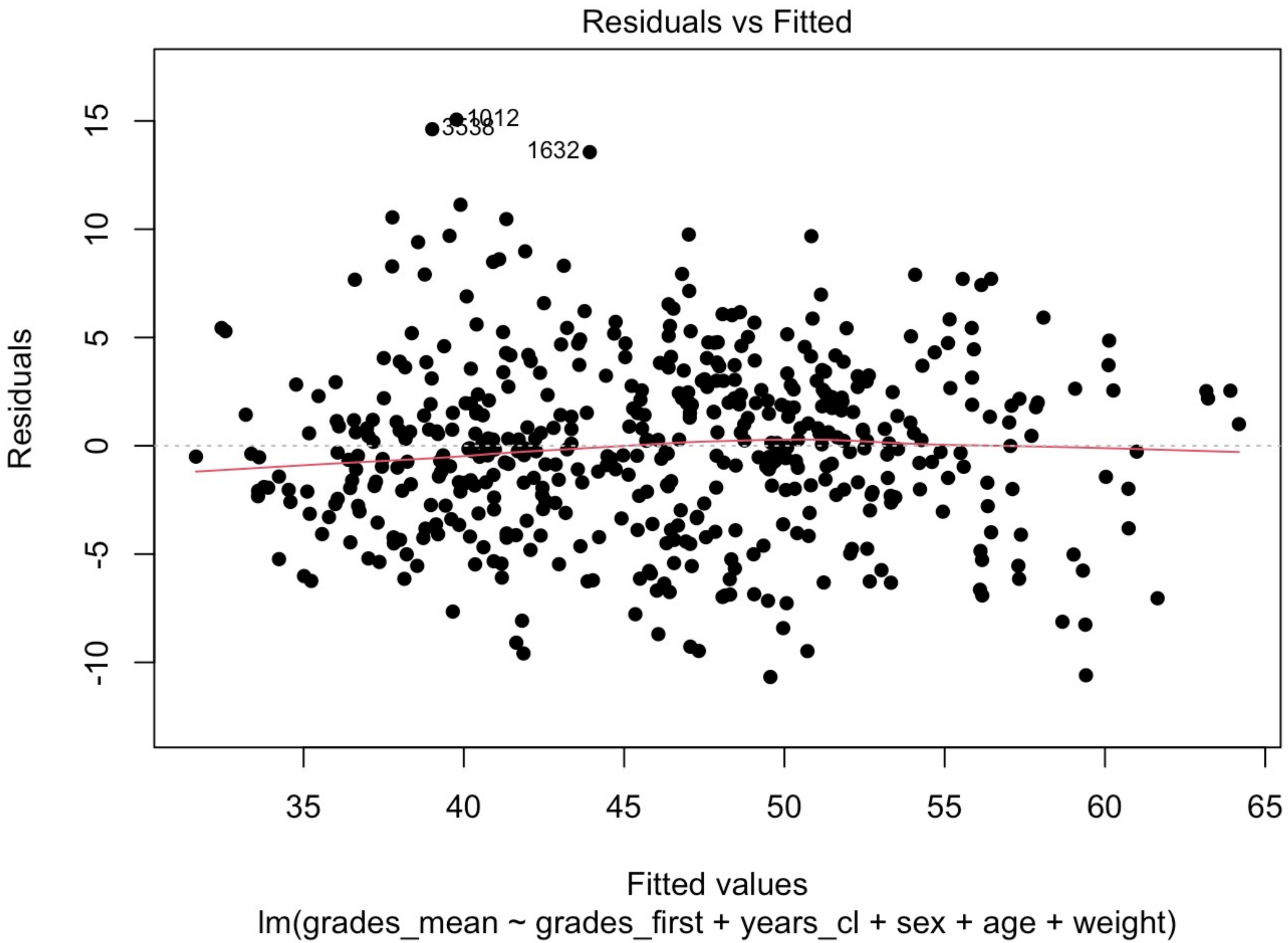
4. Compute the test statistic

$\mathrm{t} = 28.468$

5. Conclusion

$28.468 > 1.96$. We reject H_0 null hypothesis and have significant evidence at the $\alpha = 0.05$ after controlling for other parameters.

MULTIPLE REGRESSION



CONCLUSION

- As we reject Null Hypothesis, we have significant evidence that $\beta_n \neq 0$, and there is linear association between grades_mean~grades_first
- Suggested model: R-squared = 0.7268
  - grades_mean ~ grades_first + years_cl + sex + age + weight.
  - t-test provides significant evidence of contribution of main parameter while other parameters are included in the model.

THANK YOU !

CREDITS

- https://www.kaggle.com/datasets/jordizar/climb-dataset
- https://googleslides.org/free-templates