

Linear regression

Alizeh Azhar

9/20/2021

Problem 1: OLS Exercise 8.4

- a. The plot does not form a linear pattern, the pattern is nonlinear, linear model is not reasonable in this case.
- b. The plot also form a nonlinear pattern, linear model is not reasonable in this case.
- c. The plot suggests presence of strong linear relationship, because it forms a perfect linear pattern. Application of linear model is reasonable
- d. The pattern formed by the plot suggests weak linear relationship. Linear model is a reasonable
- e. The pattern linear pattern formed by the plot suggests presence of weak linear relationship. Linear model can be applied
- f. He pattern formed suggests moderate linear relationship. Linear model is reasonable because it forms linear pattern.

Problem 2: OLS Exercise 8.5

a.Exam 2 has the strongest correlation with the final exam. This is because the plot forms a better linear pattern as compared to Exam 1. Exam 1 has many outliers while majority of the points in Exam 2 linear on or closer to the fitted line as compared to exam 1. b. Exam 2 is close to final exam, the student is more prepare for Exam2 than Exam 1 because it's the predictor of final performance.

Problem 3:OLS Exercise 8.5

- a) There is strong linear relationship between husband and wife age. As husband's age increases woman's age rises.
- b) There is weak relationship between husband's and wife's height.
- c) Plot of husband and wife's age shows stronger correlation. There is a clear linear pattern with few outliers.
- d) Conversion of units for heights from centimeters to inches does not influence the correlation between husband's and wife's height.

Problem 4: OLS Exercise 8.20

0.5 shows under estimation of the incidences of skin cancer. This is because the residual is positive.

Problem 5: OLS Exercise 8.24

- a) Equation of the regression line

```
## [1] 0.00623042
```

```
## [1] 170.7321
```

```
# Equation height=170.73+0.00624 girth
```

b) Interpretation The model states that a change in girth by 1 cm increases height by 0.00623 cm.

c) Predicted height

```
## [1] 171.354
```

d) Residual

```
## [1] -11.5
```

The residual means that the height is over predicted by 11.5 cm.

Problem 6

a. Using `lm()`, fit a linear regression model predicting birthweight from gestation

```
## The following object is masked _by_ '.GlobalEnv':
##
##      height

##
## Call:
## lm(formula = bwt ~ gestation)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -49.394 -11.125   0.071  10.106  57.353
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.06418    8.32220  -1.209   0.227
## gestation     0.46426    0.02974  15.609 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.66 on 1221 degrees of freedom
## (13 observations deleted due to missingness)
## Multiple R-squared:  0.1663, Adjusted R-squared:  0.1657
## F-statistic: 243.6 on 1 and 1221 DF,  p-value: < 2.2e-16
```

b. Write the equation of the line using mathematical notation (2 pts)

$y = -10.06418 + 0.46426x$, Where y represent birth weight while x represent gestation ‘