Understanding Simple Linear Regression Metrics

2023-01-26

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
##
## Attaching package: 'janitor'
##
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

Suppose a simple linear model is fit to predict Y= weight in kilograms using X= height in centimeters of an adult. But say a new simple linear model is fit using Y = height in centimeters and X= weight in kilograms (that is to say Y and X have reversed). State whether each of the following would be the same for this new model as it was for the original model, or it would be different and explain in a sentence or two.

Generate dummy data for testing

```
# weights in kg, uniform distribution (100 units, a = 40, b = 120)
weights <- runif(n = 100, min = 40, max = 120)
# heights in cm, uniform distribution (100 units, a = 152, b = 214)
heights <- runif(n = 100, min = 152, max = 214)
```

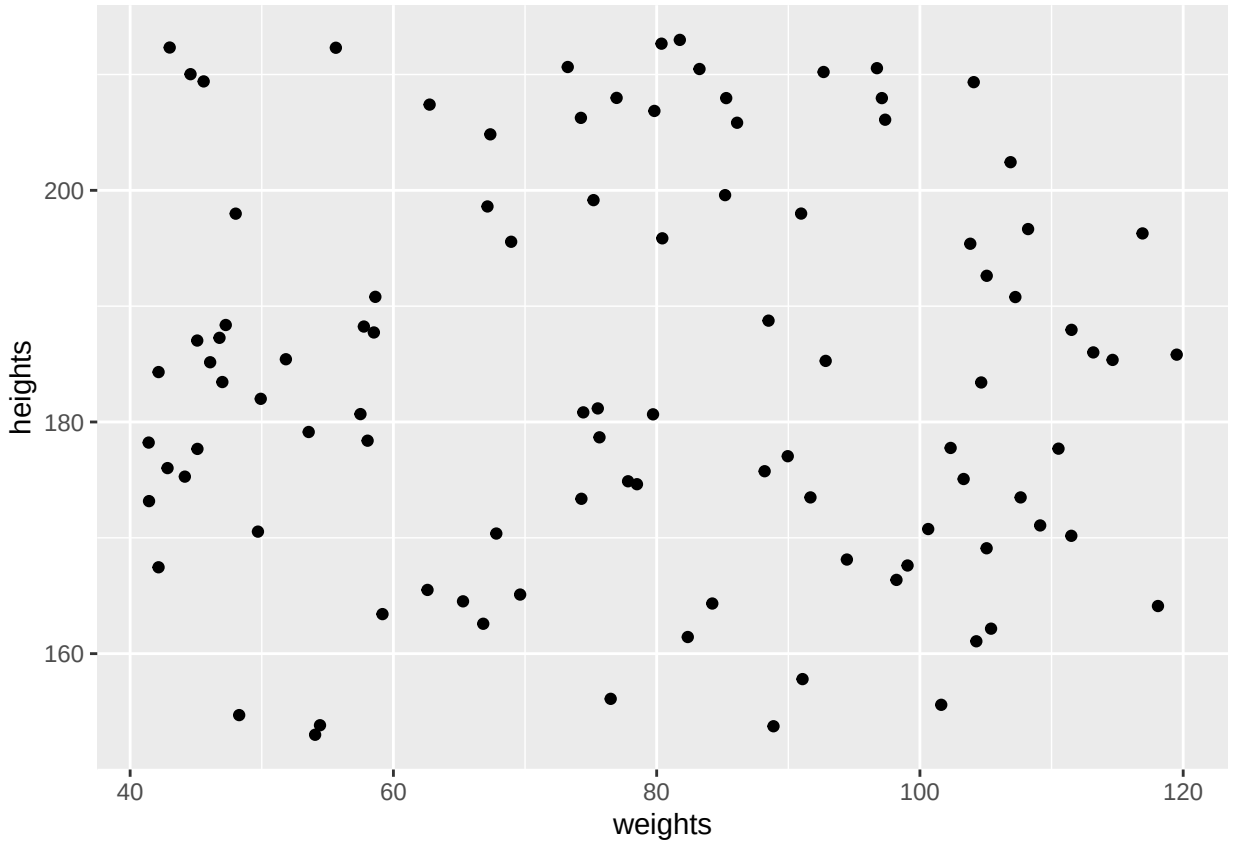
Put data into dataframe

```
df <- data.frame(weights, heights)
```

Plot data

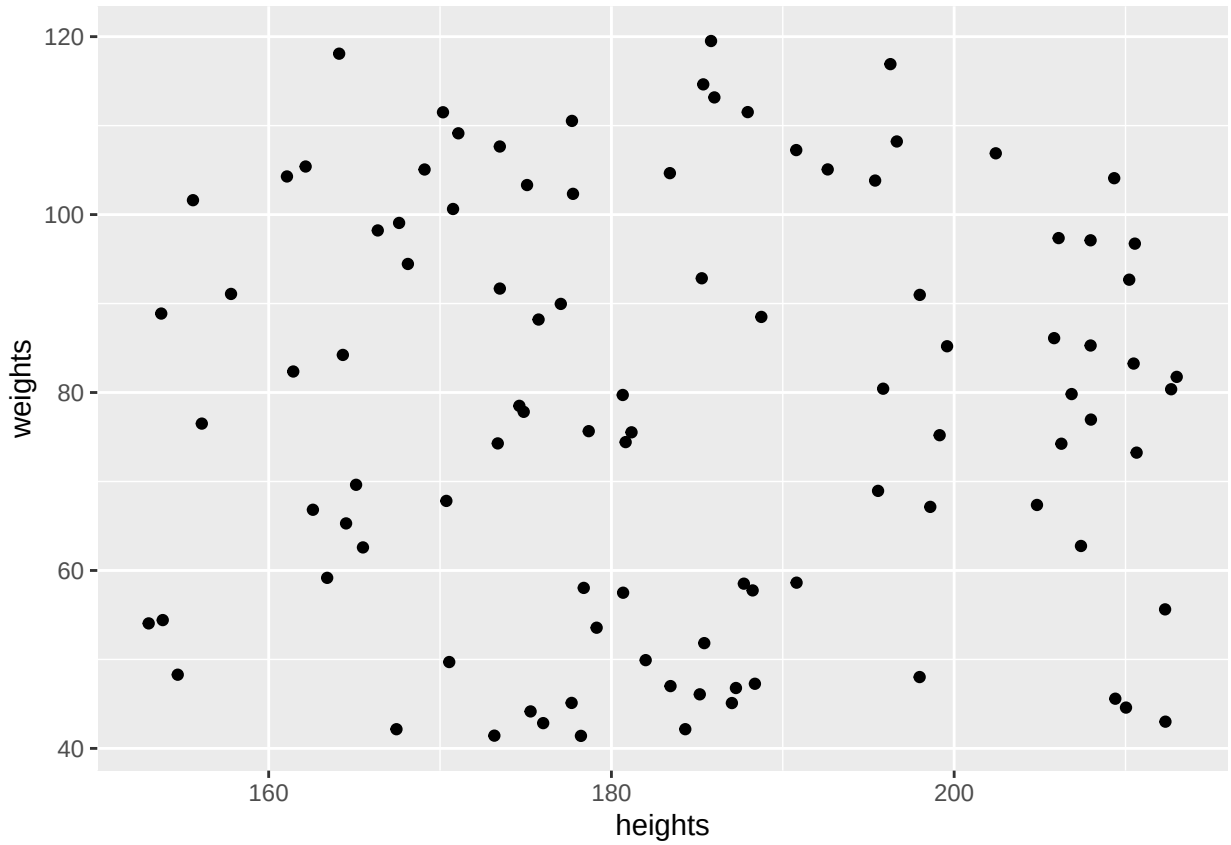
weight (x-axis), height (y-axis)

```
ggplot(
  data=df,
  aes(x=weights,
      y=heights
      )
) +
geom_point()
```



height (x-axis), weight (y-axis)

```
ggplot(
  data=df,
  aes(x=heights,
      y=weights
      )
) +
geom_point()
```


120
100
80
60
40
weights
160
180
200
heights

Model creation

weights (X), heights (Y)

```
model_a <- lm(heights ~ weights)
summary(model_a)
```

```
##
## Call:
## lm(formula = heights ~ weights)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.100 -13.214  -1.404  13.915  29.037
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 184.386660   6.040965  30.523   <2e-16 ***
## weights      -0.005336   0.074152  -0.072    0.943
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.14 on 98 degrees of freedom
## Multiple R-squared:  5.283e-05,  Adjusted R-squared:  -0.01015
## F-statistic: 0.005178 on 1 and 98 DF,  p-value: 0.9428
```

heights (X), weights (Y)

```
model_b <- lm(weights ~ heights)
summary(model_b)
```

```
##
## Call:
## lm(formula = weights ~ heights)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.792 -20.404   0.931  20.136  41.406
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 79.941904  25.422262   3.145   0.0022 **
## heights     -0.009901   0.137603  -0.072   0.9428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.35 on 98 degrees of freedom
## Multiple R-squared:  5.283e-05,  Adjusted R-squared:  -0.01015
## F-statistic: 0.005178 on 1 and 98 DF,  p-value: 0.9428
```

Standardizing data

```
z_heights <- scale(heights)
z_weights <- scale(weights)
```

Model creation w/ standardized data

z_weights (X), z_heights (Y)

```
z_model_a <- lm(z_heights ~ z_weights)
summary(z_model_a)
```

```
##
## Call:
## lm(formula = z_heights ~ z_weights)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.82387 -0.77497 -0.08234  0.81606  1.70290
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.815e-16  1.005e-01   0.000    1.000
## z_weights   -7.268e-03  1.010e-01  -0.072    0.943
##
## Residual standard error: 1.005 on 98 degrees of freedom
## Multiple R-squared:  5.283e-05,  Adjusted R-squared:  -0.01015
## F-statistic: 0.005178 on 1 and 98 DF,  p-value: 0.9428
```

z_heights (X), z_weights (Y)

```
z_model_b <- lm(z_weights ~ z_heights)
summary(z_model_b)
```

```
##
## Call:
## lm(formula = z_weights ~ z_heights)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.58396 -0.87842  0.04008  0.86690  1.78258
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.309e-17  1.005e-01   0.000    1.000
## z_heights   -7.268e-03  1.010e-01  -0.072    0.943
##
## Residual standard error: 1.005 on 98 degrees of freedom
## Multiple R-squared:  5.283e-05,  Adjusted R-squared:  -0.01015
## F-statistic: 0.005178 on 1 and 98 DF,  p-value: 0.9428
```