

Regression Computer Lab

Ben Goldstone

2/10/23

Import Data

```
library(readr)
Spring2023Dataset <- read_csv("~/CSVs/Spring2023Dataset.csv")

## Rows: 69 Columns: 18

## -- Column specification -----
## Delimiter: ","
## chr (10): ToppingPref, GiveUp, TextOften, Active, Generous, Gender, CoffeeFr...
## dbl (8): SpendOnYou, LongestRun, PizzaToppings, TextsSent, TextsReceived, G...

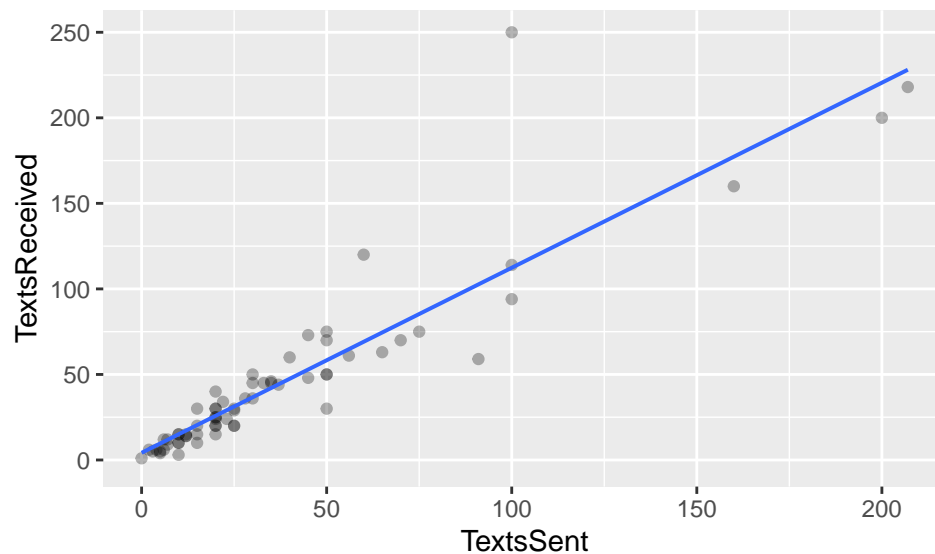
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
head(Spring2023Dataset)

## # A tibble: 6 x 18
##   SpendOnYou LongestRun PizzaToppings ToppingPref GiveUp TextsSent TextsReceived
##   <dbl>      <dbl>      <dbl> <chr>      <chr>      <dbl>      <dbl>
## 1      400        1          3 Mostly veg~ Meat          50          50
## 2      850        7          4 Mostly veg~ Caffe~         20          25
## 3      550       2.5          4 Mostly mea~ <NA>          23          24
## 4     1000       0.5          2 Mostly veg~ Meat          20          20
## 5      400       10          2 Mostly mea~ Desse~         15          20
## 6      300        1          2 Mostly mea~ Caffe~          0           1
## # ... with 11 more variables: GroupText <dbl>, TextOften <chr>, Active <chr>,
## #   Generous <chr>, Gender <chr>, HighwayAlone <dbl>, HighwayPassenger <dbl>,
## #   CoffeeFreq <chr>, StudyLocationOriginal <chr>, Personality <chr>,
## #   StudyLocation <chr>
```

Texts Sent and Texts Received

```
gf_point(TextsReceived~TextsSent,data=Spring2023Dataset,alpha=0.3) %>%
  gf_lm()

## Warning: Removed 4 rows containing non-finite values (stat_lm).
## Warning: Removed 4 rows containing missing values (geom_point).
```



Observation

The relationship between texts received and texts sent seem to fit the line with a positive strong correlation.

R Calculations

```
# r
cor(TextsReceived~TextsSent,data=na.omit(Spring2023Dataset))

## [1] 0.9111572

# r^2 as a percentage
(cor(TextsReceived~TextsSent,data=na.omit(Spring2023Dataset))^2)*100

## [1] 83.02074
```

Text sent accounted for 83% of the variability in the amount of texts sent.

Linear Model

```
model=lm(TextsReceived~TextsSent,data=na.omit(Spring2023Dataset))
summary(model)

##
## Call:
## lm(formula = TextsReceived ~ TextsSent, data = na.omit(Spring2023Dataset))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -42.315  -7.420  -2.172   3.049 139.049
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.88943    3.60406   1.079   0.285
## TextsSent     1.07061    0.06303  16.985 <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.92 on 59 degrees of freedom
## Multiple R-squared:  0.8302, Adjusted R-squared:  0.8273
## F-statistic: 288.5 on 1 and 59 DF,  p-value: < 2.2e-16
```

Linear Model Context

For every additional 1 text sent, we can expect an additional 1.07 text received. For 0 texts sent, we can expect 3.89 texts received.

```
slope = 1.07061
intercept = 3.88943
response = 300*slope + intercept

response
```

```
## [1] 325.0724
```

Prediction

Since sending 300 texts is outside of the range of 0 to 220, we cannot be confident in the result we receive.

Residual Analysis

For the one person who sent 100 texts and received 250 texts, the residual would be high compared to the majority of the other points since it is about 125 texts above the linear regression line. Since the residual is above the line, it should be positive.