Lab Assignment #6

1. Speed and Stopping Distance of a Car **(Data: CarStopDistance)**

The American Automobile Association looked at the relationship between the average stopping distance and the speed of a car.

Variable	Description
StopDist	Average stopping distance in feet (y)
Speed	Speed of a car in miles per hour (x)

(a) Draw a scatterplot and add a regression line on the scatterplot.

```
> plot(StopDist ~ Speed)
> abline( lm(StopDist ~ Speed) )
```

- Paste the R commands inside a text box. Paste the plot (no text box needed).

(b) Fit a linear model to the data. That is, obtain the regression analysis results.

- Paste the following command in RStudio.

```
> summary( lm(StopDist ~ Speed) )
```

- Paste the R command and the result inside a text box.

(c) Using the results from (b), write the equation of the regression line, $\hat{y} = b_0 + b_1 x$.

(d) Interpret the slope of the regression line in the context of the problem. **Note**: You need to use the value b_1 in your answer to receive full credit.

(e) Interpret the R-squared value. i.e. Does car speed explain a large portion (how much) of the variability in the average stopping distance? **Note**: You need to use the R-squared value in your answer to receive full credit.

(f) Predict the average stopping distance of a car when its speed is 32 miles per hour.

(g) Find the correlation coefficient between the average stopping distance and the car's speed.

- Paste the R command and their result inside a text box.

(h) Convert the stopping distance in feet to inches. (1 feet = 12 inches)

- Paste the following command in RStudio.

```
> StopDist.inch = StopDist * 12
```

The above command creates a new variable, named `StopDist.inch`, and this variable contains the average stopping distance in inches for the cars in the data (i.e. linear transformation).

- Paste the following command in RStudio.

```
> head( cbind(StopDist, StopDist.inch) )
```

  This command will display the first few values of two variables `StopDist` and `StopDist.inch` together.
- Paste the commands and its result inside a text box.

(i) Find the correlation coefficient between the average stopping distance (in inches) and the car's speed. **Note**: You'll notice that any linear transformation of x and/or y won't affect the correlation coefficient.

2. Duration and Waiting Time **(Data: Eruption)**

Millions of people from around the world flock to Yellowstone Park in order to watch eruptions of the Old Faithful geyser. It is believed that one can predict the time until the next eruption using the length of time of the last eruption. The data set contains 21 consecutive eruptions of Old Faithful geyser.

Variable	Description
`Duration`	Length of time in minutes of the last eruption (x)
`WaitingTime`	Time in minutes until the next eruption (y)

(a) Draw a scatterplot and add a regression line on the scatterplot.

- Paste the R commands inside a text box.
- Paste the plot (no text box needed). Resize the plot if it is too large or too small.

(b) Fit a linear model to the data. That is, obtain the regression analysis results.

- Paste the R command and the result inside a text box. (See Problem 1)

(c) Using the results from (b), write the equation of the regression line, $\hat{y} = b_0 + b_1 x$.

(d) Interpret the slope of the regression line in the context of the problem. **Note**: You need to use the value b_1 in your answer to receive full credit.

(e) Is the y-intercept of the regression line meaningful in this case?

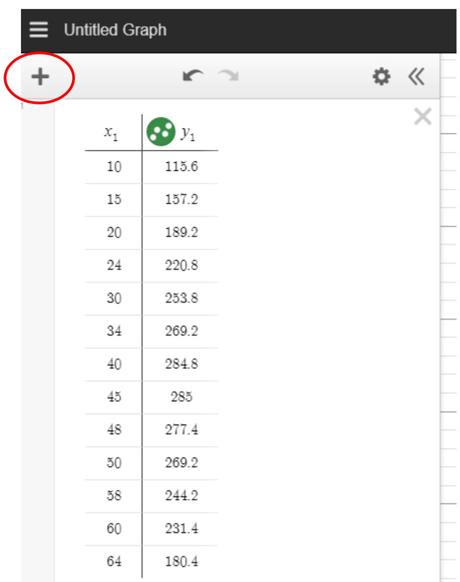
(f) Interpret the R-squared value. i.e. Does the length of time of the last eruption explain a large portion (how much) of the variability in the time until the next eruption?

(g) You just missed an eruption, but you know it lasted 3.3 minutes. In how many minutes do you expect to see the next eruption? **Note**: Make your prediction using the regression model.

(h) One of the eruptions in the data set lasted 3.3 minutes and the next eruption occurred 62 minutes later. Find the residual of this eruption and interpret the residual.

3. Angle and Distance **(Data: Baseball)**

The data set contains the horizontal distance (in feet) traveled by a baseball hit at various angles. The initial speed of the ball at the bat is constant.

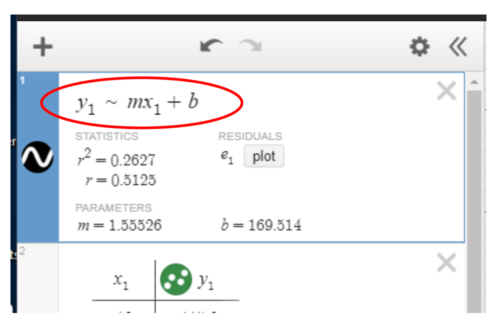


Variable	**Description**
`Angle`	Angle in degrees (x)
`Distance`	Distance in feet from home plate (y)

For this problem, we will use **Desmos** (Graphing Calculator).

- Open a browser. Any browser will do (as far as I know)
- Click this link www.desmos.com/calculator
- Open the data file, copy both columns (Ctrl+C) and paste them (Ctrl+V) into Desmos. Then, it will show the data as a table and the scatterplot

1) Let's fit a linear model (straight line) to the data.



- Click on the **+** button and select "$f(x)$ expression"
- Inside the box, type the following.

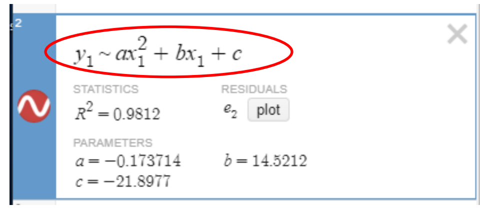
$$\boxed{y_1 \sim mx_1 + b}$$

- Then, you'll see the regression equation. It also adds a regression line (straight line) on the scatterplot.

(a) Using the results, write the regression equation.

(b) Report the correlation coefficient (r) and the R-squared value (r^2).

2) Let's fit a quadratic model (parabola) to the data.



- Click on the **+** button and select "$f(x)$ expression"
- Inside the box, type the following.

$$\boxed{y_1 \sim ax_1^2 + bx_1 + c}$$

- Then, you'll see the regression equation. It also adds a regression curve (parabola) on the scatterplot.

(a) Using the results, write the regression equation.

(b) Report the R-squared value (R^2).

(c) The left field fence is 280 feet from home plate. At what angles, to the nearest degree, will the ball be hit past the left field fence? **Note**: Look at the data set or the scatterplot to answer this.