Homework exercise

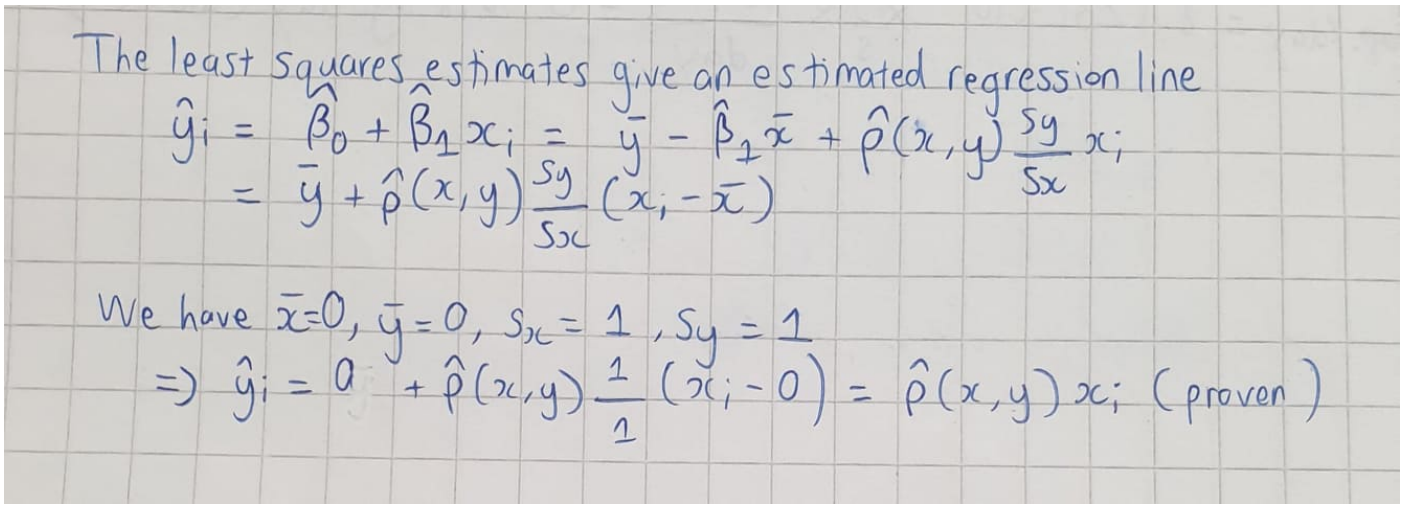
To be solved at home before the exercise session.

---

1. a. Show that if in simple linear regression both the explanatory variable x and the response y have been marginally standardized such that $\bar{x} = 0$, $s_x = 1$ and $\bar{y} = 0$, $s_y = 1$, then the estimated least squares regression model is simply,

$$\hat{y}_i = \hat{\rho}(x, y) x_i.$$

That is, the regression coefficient of x equals the sample correlation between x and y.

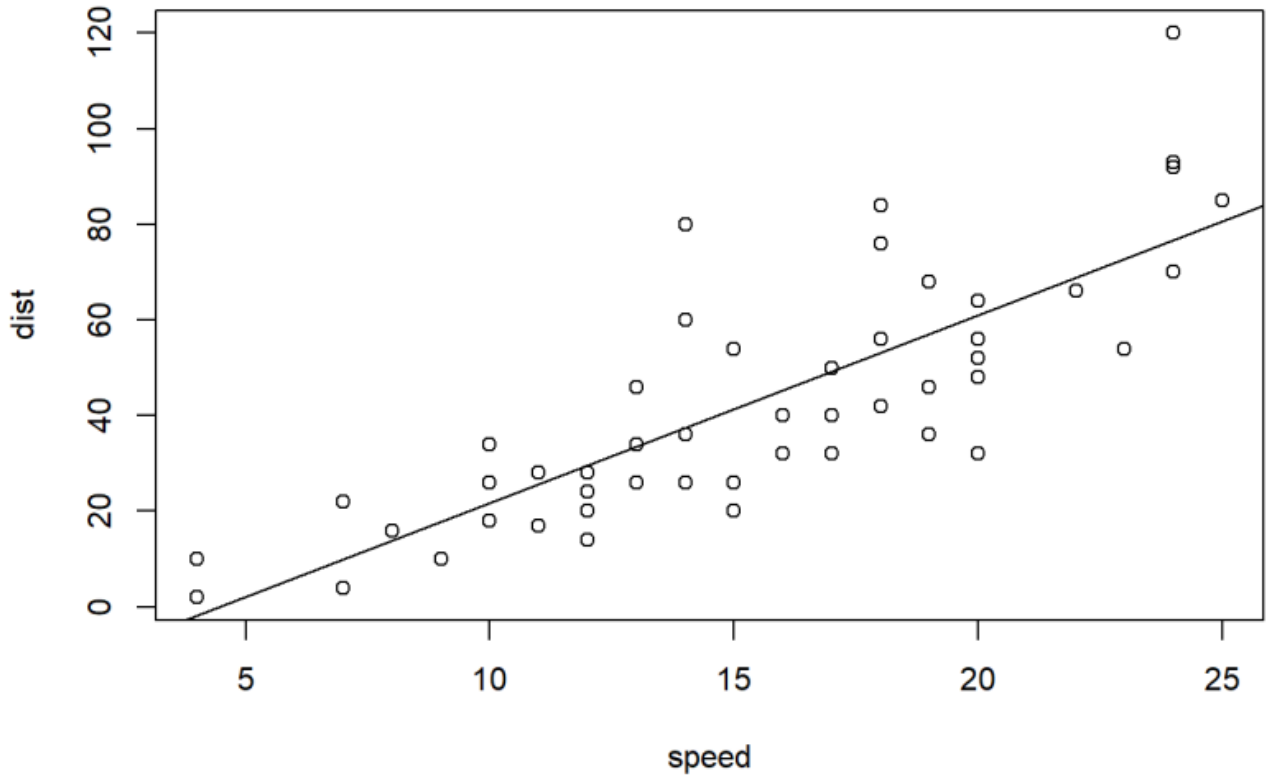


b. The `cars` data give the speeds of cars (`speed` , in mph) and the corresponding distances taken to stop (`dist` , in feet). The below shows the model summary of a simple linear regression model fit using `speed` as an explanatory variable and `dist` as a response. Interpret the model results.

```
cars_lm <- lm(dist ~ speed, data = cars)
summary(cars_lm)
```

```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## speed         3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

```
plot(dist ~ speed, data = cars)
abline(cars_lm)
```



Residuals: This is the residual calculated from the difference between the true distance and the estimated distance from the linear regression: min, max, first quarter, third quarter and mean residuals are reported

Coefficients: Estimated intercept and speed coefficient are b0 and b1 in the model y = b1 * x + b0, where x is the speed and y is the distance

The R-squared is 0.6438, suggesting that there are a lot of variations in the data. The higher R-squared, the more correlated between the speed and the distance