SOLUTIONS

Homework 4

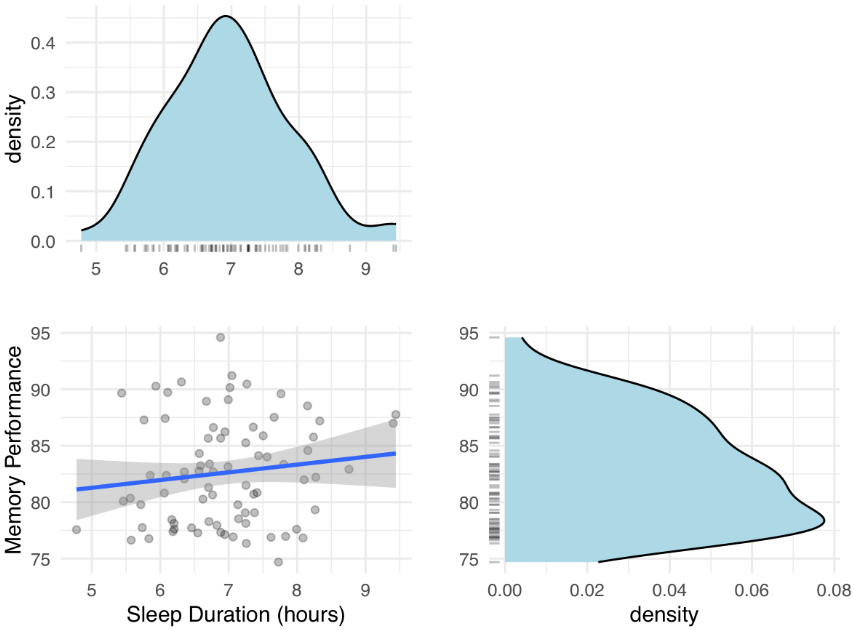
Author: Matthew J. Crossley

Last update: 23 May, 2024

Q1

Consider an investigation into the relationship between the duration of sleep and performance on a memory task. Below is a plot of the data collected, along with the predictions from the best fitting linear model.

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
##
## Call:
## lm(formula = memory_performance ~ sleep_duration, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.4508 -4.0698 -0.5393  3.2803 12.0390
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      77.8549     4.0697   19.13    <2e-16 ***
## sleep_duration    0.6841     0.5800    1.18     0.242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.637 on 78 degrees of freedom
## Multiple R-squared:  0.01752,    Adjusted R-squared:  0.004929
## F-statistic: 1.391 on 1 and 78 DF,  p-value: 0.2418
```

- Please write the equation of the best fitting linear model.

$$y \sim \beta_0 + \beta_1 x$$

y = MEMORY PERFORMANCE

x = SLEEP DURATION

TO BE VERY EXPLICIT ABOUT "BEST FITTING" YOU COULD WRITE: $\hat{y} \sim \hat{\beta}_0 + \hat{\beta}_1 x$

- Please list all random variables in the best fitting linear model and state how they are distributed.

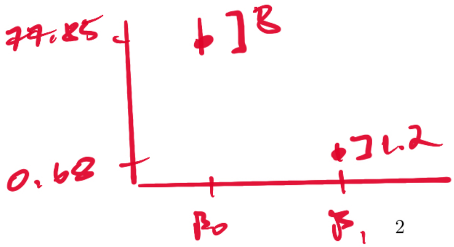
THE TERM "BEST FITTING" MIGHT BE SEEN AS TAKING ALL RANDOMNESS OUT BECAUSE THE SAMPLE x AND ESTIMATES $\hat{\beta}_0$ & $\hat{\beta}_1$ ARE FIXED - NOT RANDOM - VARIABLES.

HOWEVER, THE SPIRIT OF THE QUESTION IS:

$$y \sim N(\mu_y, \sigma_y) \quad \mu_y = \beta_0 + \beta_1 x$$

$$\varepsilon \sim N(0, \sigma_\varepsilon)$$

- Please draw a pointrange plot illustrating the best fittings β coefficients with error bars showing SEM.



- What is Pearson's correlation coefficient between sleep duration and memory performance?

$\hat{\beta}_1 = 0.6841$

$\hat{\beta}_1 = r \frac{s_y}{s_x} \rightarrow r = \hat{\beta}_1 \frac{s_x}{s_y}$

SINCE WE DON'T KNOW $\frac{s_x}{s_y}$

THIS IS AS FAR AS WE CAN GO.

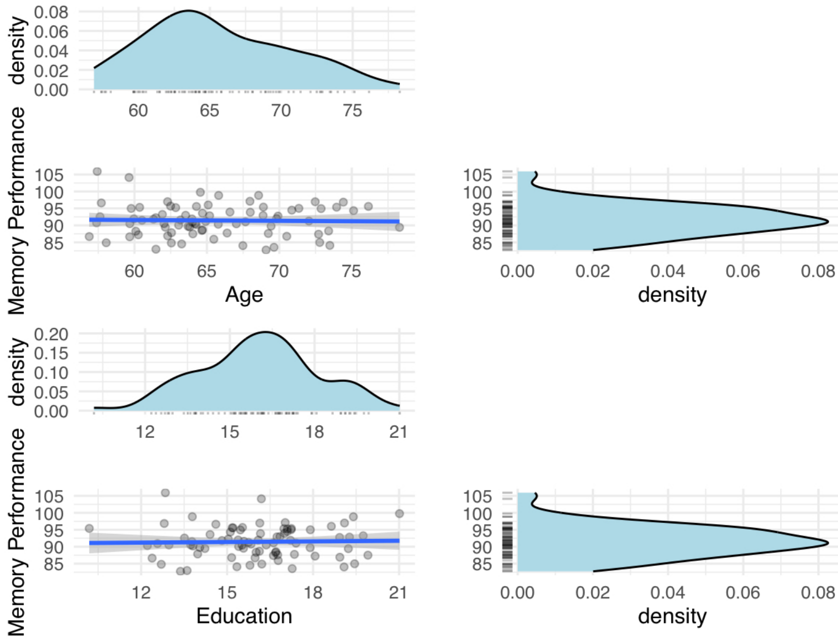
- Please write a few sentences reporting the results of this analysis.

WE PERFORMED A SIMPLE LINEAR REGRESSION TO EXAMINE MEMORY PERFORMANCE AS A LINEAR FUNCTION OF SLEEP DURATION. WE DID NOT FIND SIGNIFICANT EVIDENCE THAT THERE WAS A LINEAR RELATIONSHIP [$t(78) = 1.18$, $p = 0.24$]

Q2

Consider a study in which researchers collected data on participants' age and education level and examined how these variables predict performance on a memory task. Below is a plot of the data along with the predictions from the best-fitting multiple regression model.

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```




```
##
## Call:
## lm(formula = memory_performance ~ age + education, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.4433 -3.4024 -0.1586  3.3605 14.4783
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 92.10810    7.64862  12.042   <2e-16 ***
## age         -0.02659    0.10573  -0.251    0.802
## education    0.06764    0.25206   0.268    0.789
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.677 on 77 degrees of freedom
## Multiple R-squared:  0.001583,   Adjusted R-squared:  -0.02435
## F-statistic: 0.06102 on 2 and 77 DF,  p-value: 0.9408
```

- Please write the equation of the best fitting linear model.

SAME CAVEAT AS Q1 BUT ESSENTIALLY:

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

y = MEMORY PERFORMANCE

x_1 = AGE

x_2 = EDUCATION

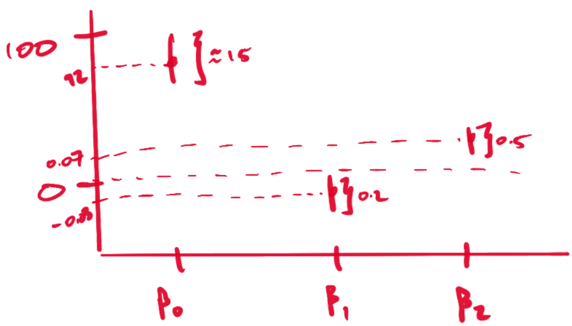
- Please list all random variables in the best fitting linear model and state how they are distributed.

SAME CAVEAT AS Q1 BUT ESSENTIALLY:

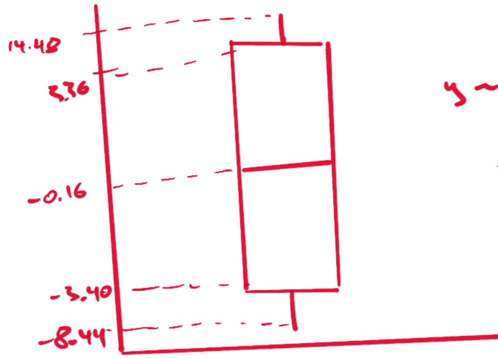
$y \sim N(\mu_y, \sigma_y)$ $\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$\varepsilon \sim N(0, \sigma_\varepsilon)$

- Please draw a pointrange plot illustrating the best fittings β coefficients with error bars showing SEM.



- Please draw a boxplot illustrating the residuals of the best fitting linear model. Please set the whiskers to extend to the min and max values. What random variable in your linear model does this plot tell you about?



$y \sim \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

TELLS US ABOUT

- What is Pearson's correlation coefficient between (1) age and memory performance, and (2) education level and memory performance?

AGE & MEMORY: $\beta_1 \rightarrow r_1 = \frac{s_{x_1}}{s_y} \hat{\beta}_1$

AGE & EDUCATION: $\beta_2 \rightarrow r_2 = \frac{s_{x_2}}{s_y} \hat{\beta}_2$

- What is the partial correlation between age and memory performance when controlling for education level?

y = MEMORY

x_1 = AGE

x_2 = EDUCATION

$$r_{x_1 y \cdot x_2} = \frac{r_{x_1 y} - r_{x_1 x_2} r_{y x_2}}{\sqrt{(1 - r_{x_1 x_2}^2)(1 - r_{y x_2}^2)}}$$

- Please write a few sentences reporting the results of this analysis for an academic journal.

A MULTIPLE REGRESSION USING AGE AND EDUCATION AS REGRESSORS FOUND NO SIGNIFICANT EFFECT OF EITHER AGE [t(77)=-.25, P=.80] OR EDUCATION [t(77)=.27, P=.79]. THE OVERALL MODEL DID NOT FIT BETTER THAN A MODEL BASED ONLY ON THE MEAN [F(2,77)=.06, P=.94].

- Please write R code (using your pen or pencil) that generates the regression analysis reported above.

```
fm <- lm(y ~ age + education, data=d)
```

$$d = \begin{array}{c} \begin{array}{ccc} \text{AGE} & \text{EDUCATION} & \text{SUBJECT} \end{array} \\ \left[\begin{array}{ccc} x_{11} & x_{21} & 1 \\ x_{12} & x_{22} & 2 \\ x_{13} & x_{23} & 3 \\ \vdots & \vdots & \vdots \end{array} \right] \end{array}$$