7. How tall will your child be? A researcher has collected a random sample of heights of parents and their female children (all heights are in inches). The heights of the mother, father, and daughter are recorded in the following table.

Table for Exercise 7 – Heights of Parents and Daughters (Inches)

Mother	64	66	62	70	70	58	66	66	64	67	65	66	68
Father	73	70	72	72	72	63	75	75	72	69	77	70	74
Daughter	65	65	61	69	67	59	69	70	68	70	70	65	70

a. Create two scatterplots using the mother with the daughter and the father with the daughter. Does there appear to be a linear relationship in either of the plots?

b. Using statistical software, estimate the parameters of the following regression model.

$$\text{Daughter Height} = \beta_0 + \beta_1(\text{Mother Height}) + \beta_2(\text{Father Height}) + \varepsilon_i$$

c. Is the overall model useful in explaining the variation in daughter height? Test at the 0.05 level.

d. Is the father's height useful in explaining the daughter's height? Test at the 0.05 level.

e. Is the mother's height useful in explaining the daughter's height? Test at the 0.01 level.

f. Interpret each of the regression coefficients.

g. Construct and interpret 95% confidence intervals for β_1 and β_2. Interpret these intervals.

h. Predict the height of a daughter whose father is six feet two inches tall and whose mother is five feet four inches tall.

i. Find a 95% prediction interval for the height of a daughter whose father is six feet two inches tall and whose mother is five feet four inches tall. Interpret this interval.

j. Find a 95% confidence interval for the average height of a daughter whose father is six feet two inches tall and whose mother is five feet four inches tall.

Question 7a

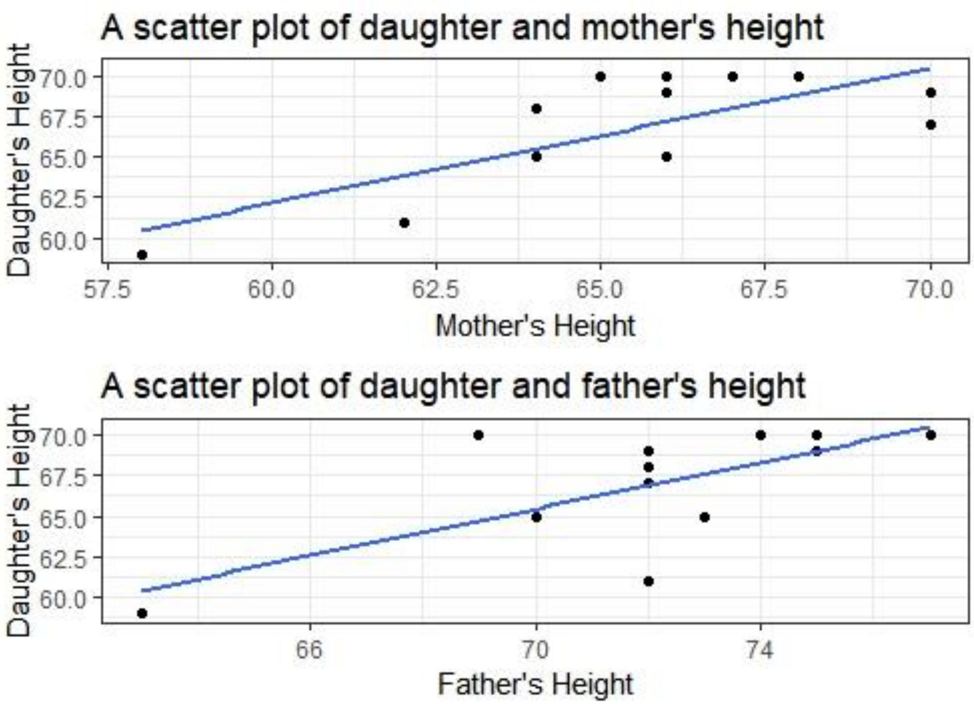


Figure 1: Scatterplots of the relationship with daughter's height

The exploratory data analysis (EDA) with scatterplots show that there is a linear relationship in the two plots

Question 7b

Table 1: Model statistics

Parameter (s)	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	-4.6456	15.0956	-0.308	0.7646
Mother height	0.5939	0.2260	2.628	0.0253
father height	0.4523	0.2078	2.176	0.0546

Daughter height = $\beta_0 + \beta_1(Mother\ height) + \beta_2(Father\ height) + \varepsilon_i$

Using R programming software, the model is

$\widehat{\text{Daughter height}} = -4.6456 + 0.5939(Mother\ height) + 0.4523(Father\ height)$

Question 7c

Table 2: Analysis of variance (ANOVA)table

Source of variance	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Predictors	2	108.219	54.110	11.252	0.002756
Residuals	10	48.088	4.809		

The overall model is significant (P-value<0.05) in explaining variation in daughter's height at $\alpha = 0.05$ level of significance.

Question 7d

The father's height is not significant (P-value>0.05) in explaining the daughter's height at $\alpha = 0.05$ level of significance.

Question 7e

The mother's height is not useful (P-value>0.01) in explaining the daughter's height at $\alpha = 0.01$ level of significance.

Question 7f

Without any knowledge of father and mother's height, the daughter's height will decrease by 4.646 inches.

An inch increase in mother's height will bring about 59% increase in daughter's height, other variables being held in the model.

An inch increase in father's height will bring about 45% increase in daughter's height, other variables being held in the model.

Question 7g

Table 3: Confidence interval for $\boldsymbol{\beta_1 and\ \beta_2}$

Parameter (s)	2.5 %	97.5 %
β_0	-38.28069452	28.9894820
β_1	0.09034753	1.0974020
β_2	-0.01082795	0.9153495

We are 95% confident that the true confidence interval for the regression slope or β_1 is :
0.09034753< β_1 < 1.0974020
We are 95% confident that the true confidence interval for the regression slope or β_2 is :
-0.0108< β_2 < 1.0974

Question 7h

$\widehat{\text{Daughter height}} = -4.6456 + 0.5939(Mother\ height) + 0.4523(Father\ height)$

If the father's height is six feet two inches tall and mother's height is five feet four inches tall, then daughter's height will be 66.82968
$\widehat{\text{Daughter height}} = -4.6456 + 0.5939(64) + 0.4523(74)$ =66.82968

Question 7i

We are 95% confident that the true prediction interval for the predicted daughter's height of 66.82968 is (61.5329 <predicted daughter's height < 72.12646)

Question 7j

We are 95% confident that the true confidence interval for the predicted daughter's height of 66.82968 is (64.7847 <average height of the daughter< 68.87465)

8. An economist is studying the relationship between income and IRA contributions. He has randomly selected eight subjects and obtained annual income and IRA contribution data from them. He wishes to predict the amount of money contributed to an IRA based on annual income.

Table for Exercise 8 – Income and IRA Contributions

Annual Income (Thousands of Dollars)	IRA Contribution (Thousands of Dollars)
28	0.3
25	0
34	1.0
43	1.3
48	3.3
39	2.2
74	8.5

a. Draw a scatterplot of the data. Describe the relationship that you observe between income and IRA contribution.

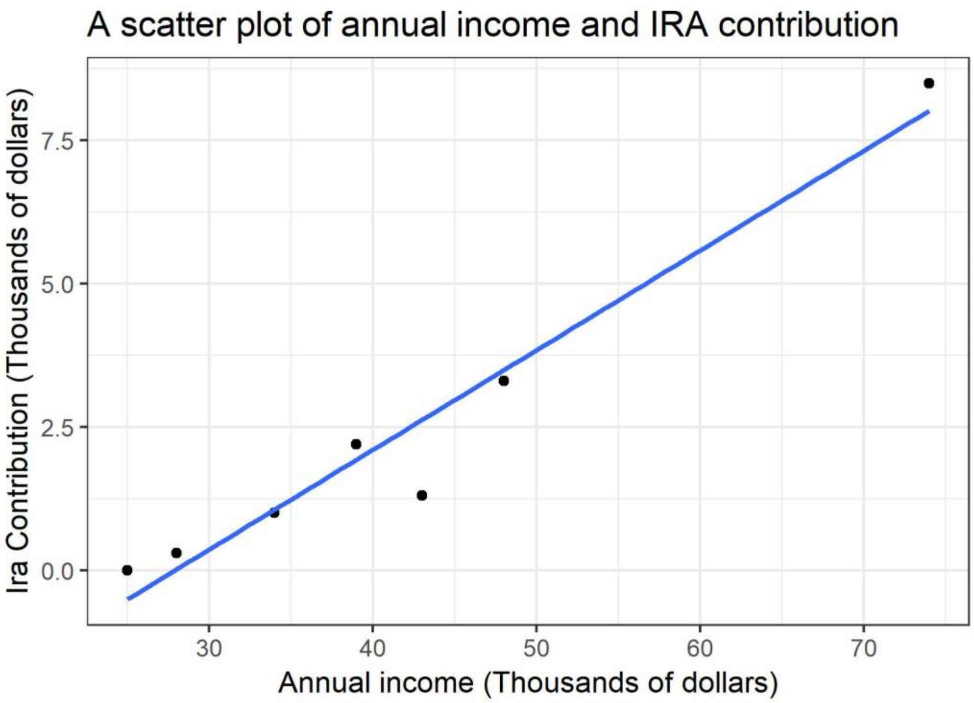
b. Estimate the parameters of the following model using statistical software.

$$\text{IRA Contribution} = \beta_0 + \beta_1(\text{Income}) + \varepsilon_i$$

c. Calculate and interpret a 95% confidence interval for β_1.

d. What assumptions are being made in the construction of the confidence interval for β_1?

Question 8a



The scatterplot shows a positive linear relationship between income and IRA contribution. The higher the annual income, the more amount of IRA contribution will be and vice versa.

Question 8b

Table 1: Model statistics

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-4.86025	0.76879	-6.322	0.00146
Annual Income	0.17396	0.01737	10.016	0.00017

IRA Contribution = $\beta_0 + \beta_1(Income) + \varepsilon_i$

Using R programming software, the model is

$\text{IRA } \widehat{\text{Contribution}} = -4.86025 + 0.17396\ (Income)$

Question 8c

Table 2: Confidence interval for $\boldsymbol{\beta_1}$

Parameter (s)	2.5 %	97.5 %
β_1	0.1293109	0.218605

We are 95% confident that the true confidence interval for the regression slope or β_1 is :
$0.1293109 < \beta_1 < 0.218605$

Question 8d

If we replicate the same study multiple times with different random samples and compute a confidence interval for each sample, we would expect 95% of the confidence intervals to contain the true slope of the regression line