7.4 Identify relationships, Part II. For each of the six plots, identify the strength of the relationship (e.g. weak, moderate, or strong) in the data and whether fitting a linear model would be reasonable.

a. strong relationship, a linear model would NOT be reasonable ✓
b. strong relationship, a linear model would NOT be reasonable ✓
c. strong relationship, a linear model would be reasonable ✓
d. weak relationship, a linear model would be reasonable ✓
e. weak relationship, a linear model would NOT be reasonable ✗
f. moderate relationship, a linear model would be reasonable ✓

7.6 Husbands and wives, Part I. The Great Britain Office of Population Census and Surveys once collected data on a random sample of 170 married couples in Britain, recording the age (in years) and heights (converted here to inches) of the husbands and wives.[16] The scatterplot on the left shows the wife's age plotted against her husband's age, and the plot on the right shows wife's height plotted against husband's height.

a) Describe the relationship between husbands' and wives' ages.

Based on the scatterplot, the age's of husbands and wives in this survey have a strong, positive correlation.

b) Describe the relationship between husbands' and wives' heights.

Based on the scatterplot, the heights of husbands and wives in this survey have a very weak correlation if any correlation at all.

c) Which plot shows a stronger correlation? Explain your reasoning.

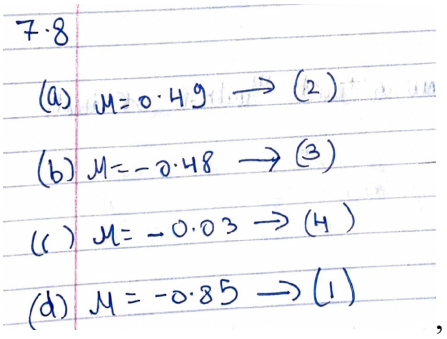
The age scatterplot shows a much stronger correlation as the points are much closer together, in a positive linear direction.

d)Data on heights were originally collected in centimeters, and then converted to inches. Does this conversion affect the correlation between husbands' and wives' heights?

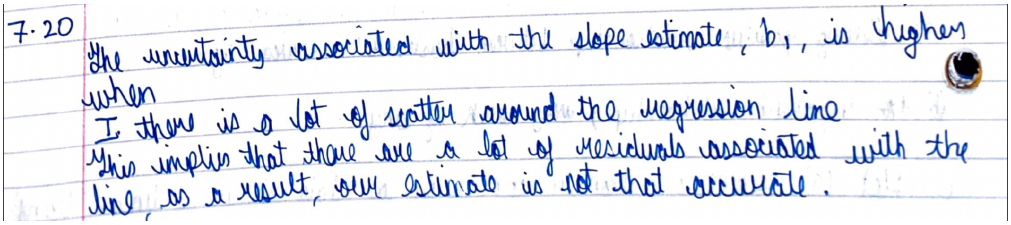
Yes. This conversion would affect the correlation between husbands' and wives' heights because it would create less variance in the points potentially making a stronger correlation.

7.7 Match the correlation, Part I. Match the calculated correlations to the corresponding scatterplot.
r= −0.7 → (4).
r=0.45 → (3).
r=0.06 → (1).
r=0.92 → (2).

7.8



7.20



7.23 Tourism spending. The Association of Turkish Travel Agencies reports the number of foreign tourists visiting Turkey and tourist spending by year.[20] Three plots are provided: scatter- plot showing the relationship between these two variables along with the least squares t, residuals plot, and histogram of residuals.

(a) The relatonship is linear and has a strong positive correlation.

(b) The explanatory varibale is Number of Tourists and response is Spending

(c) The relationship looks linear as seen from the scatter plot, therefore we might want to fit a regression line to this data.

(d) The conditions for least squared line are:

1. Linear trend: The first plot shows the linear trend, so this condition is met.
2. Residuals should be nearly normal: The histogram of the residuals is nearly normal so this condition is met.
3. Constant variability: From the second graph the variance is not constant, so this condition is not met.
4. Observation independent: Since these observation are not from time series, we assume they are independent.

Since the varinace is not constant, the data does not meet the required condition for fitting the least squared line (regression line)

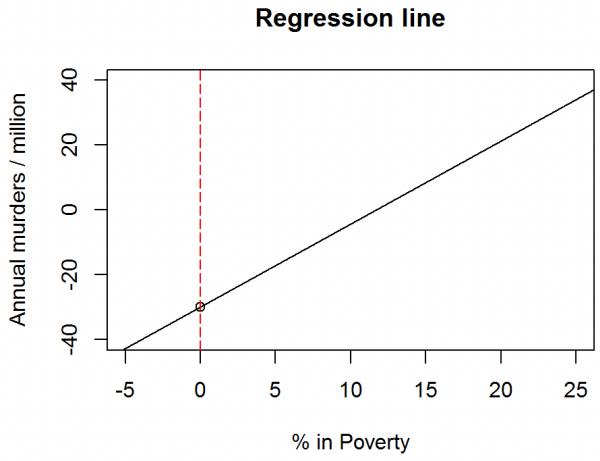
7.29 Murders and poverty, Part I. The following regression output is for predicting annual murders per million from percentage living in poverty in a random sample of 20 metropolitan areas.

a. Write the linear model.

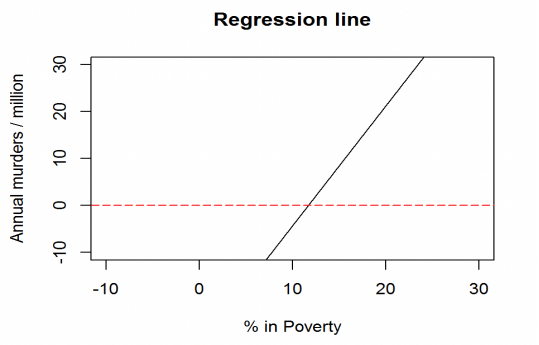
$y=\beta 0+\beta 1x$
Annual murders = -29.901 + 2.559 poverty%

b. Interpret the intercept.

Expected murder rate in metropolitan areas with no poverty (poverty = zero) is -29.901 million.
In this case there is no meaningful value.
The intercept serves to adjust the height of the regression line.



Interpret the slope.
The slope is responsible for the inclination of the regression line. The angle of the regression line is defined by the slope, θ=arctangβ1.**In this case 2.559 is equivalent to an angle of = 68º39'19.99".** **The practical interpretation is, for each percentage increase in poverty, we expect murders per million to be higher on average by 2.559**



d. Interpret R^2.
R^2 of linear model it is also known as the coefficient of determination is one measure of how close are fitted regression line to the data.
Describes the amount of variation in the response that is explained by the least squares lines (regression model).
R^2 = Explained variation/ Total variation
Poverty level explain 70.52% of the variability in murder rates in metropolitan areas.

e. Calculate the correlation coefficient.
It is simply the square root of R^2 $0.7052\text{——}\sqrt{}=0.8398$

7.36 Beer and blood alcohol content.

 Describe the relationship between the number of cans of beer and BAC.
Answer:
The increase in no. of cans consumed leads to increase in BAC. Also by looking at the scatterplot and positive reg coeff. I can deduce that y vs x plot is an upward sloping line indicating a moderate-strong positive linear relationship.

Write the equation of the regression line. Interpret the slope and intercept in context.
Answer:
$y=b0+b1*x$
$BAC=Intercept+b1*beers$
$BAC=-0.0127+0.0180*beers$
Do the data provide strong evidence that drinking more cans of beer is associated with an increase in blood alcohol? State the null and alternative hypotheses, report the p-value, and state your conclusion.
Answer:
p-value of the reg coeff for 'beers' = 0.0000 , since p-value<0.05 it is a statistically significant variable. *Null Hypothesis, Ho*: There is no significant association or b1=0.
Alternate Hypothesis, Ha: There is some significant association, b1!=0.
p-value of b1 = 0.0000, which makes us reject Ho or indicating that there is strong relationship between the response & explanatory variables. *This is a strong evidence that drinking more cans of beer is associated with an increase in blood alcohol.*

 The correlation coefficient for number of cans of beer and BAC is 0.89. Calculate R2 and interpret it in context.
Answer:

```
r=0.89
```

```
paste("R2: ", round(r^2,3))
```

```
## [1] "R2:  0.792"
```

Suppose we visit a bar, ask people how many drinks they have had, and also take their BAC. Do you think the relationship betwe en number of drinks and BAC would be as strong as the relationship found in the Ohio State study?
Answer:
Yes, the relationship will be as strong as that in the Ohio state study.

7.37 Husbands and wives, Part II.

a. Is there strong evidence that taller men marry taller women? State the hypotheses and include any information used to conduct the test. Answer

H0:H0: β1=0.β1=0. HA:HA: β1>0.β1>0. A two-sided test would also be acceptable for this application. The p-value, as reported in the table, is incredibly small. Thus, for a one-sided test, the p-value will also be incredibly small, and we reject H0.H0. The data provide convincing evidence that wives' and husbands' heights are positively correlated.

a. Write the equation of the regression line for predicting wife's height from husband's height. Answer
ˆheightW=43.5755+0.2863×heightH.

a. Interpret the slope and intercept in the context of the application. Answer
Slope: For each additional inch in husband's height, the average wife's height is expected to be an additional 0.2863 inches on average. Intercept: Men who are 0 inches tall are expected to have wives who are, on average, 43.5755 inches tall. The intercept here is meaningless, and it serves only to adjust the height of the line.

Given that R2=0.09,R2=0.09, what is the correlation of heights in this data set? Answer
The slope is positive, so rr must also be positive. r=√0.09=0.30.r=0.09=0.30.
You meet a married man from Britain who is 5'9" (69 inches). What would you predict his wife's height to be? How reliable is this prediction? Answer
63.2612. Since R2R2 is low, the prediction based on this regression model is not very reliable.
You meet another married man from Britain who is 6'7" (79 inches). Would it be wise to use the same linear model to predict his wife's height? Why or why not? Answer

No, we should avoid extrapolating.

7.38 Husbands and wives, Part III

(a) Is there strong evidence that taller men marry taller women? State the hypotheses and include any information used to conduct the test?

Using a one-sided test, the hypotheses are:

- Ho:β1=0
- HA:β1>0

The p-value from the table is practically 0, which gives strong evidence in rejecting Ho in favor HA. That is, the data do provide convincing evidence that a taller man will marry a taller woman.

(b) Write the equation of the regression line for predicting wife's height from husband's height.

$$\hat{height}_W = 43.5755 + 0.2863 \times \hat{height}_H$$

(c) Interpret the slope and intercept in the context of the application.

The slope in this context, 0.2863, means that for every inch increase in a husband's height, his wife's height is expected to increase by 0.2863".

The intercept, 43.5755", is the expected height of the wife when the husband's height is 0. This does not hold any meaning in this context; it is just an adjustment for the linear model.

(d) Given that R^2 **= 0.09, what is the correlation of heights in this data set?**
Since the slope is positive, the correlation, r, will also be positive. r is simply the square root of R^2:

$$R^2 = 0.09 \rightarrow r = \sqrt{R^2} = \sqrt{0.09} = 0.3$$

The correlation of 0.3 implies a weak-positive correlation between a wife and husband's height.

(e) You meet a married man from Britain who is 5'9" (69 inches). What would you predict his wife's height to be? How reliable is this prediction?

Using the equation derived from **(b)** and substituting 69" into $\hat{height}_H$:

$$\hat{height}_W = 43.5755 + 0.2863 \times 69" = 63.3302"$$

Since R2 is low, only 9% of the variability in the wife's height is explained by the husband's height, the prediction is not very reliable.

(f) You meet another married man from Britain who is 6'7" (79 inches). Would it be wise to use the same linear model to predict his wife's height? Why or why not?

The height of this man is outside the limits of what this model was based off on, which is approximately 61"<=heightH<=75", so it would not be wise to use the same model.
Using the least squares equation from **(b)**, it would predict the height of the wife as 66.1932". At the lower end and upper ends of the height spectrum, there are fewer and fewer"mates" in the pool that would be available to the respective partners to have this equation hold true. This model, at best, should only be used for the heights in the range of height distributions presented, and it should be used with some reservations, due to the low correlation.

7.40 Rate my professor.

a. Given that the average standardized beauty score is -0.0883 and average teaching evaluation score is 3.9983, calculate the slope. Alternatively, the slope may be computed using just the information provided in the model summary table.

```
b1 <- (3.9983 - 4.010) / -0.0883
b1
```

```
## [1] 0.1325028
```

We find our estimate of (_1 = 0.1325028).

b. Do these data provide convincing evidence that the slope of the relationship between teaching evaluation and beauty is positive? Explain your reasoning.

   Given the p-value for the slope is (0.0000), this provides strong evidence that the slope is not 0 on a two tail test. Looking at the positive t-value of 4.13 and half the two-tail p-value (still very close to zero), this provides strong evidence of a positive relationship between teaching evaluation and beauty.

c. List the conditions required for linear regression and check if each one is satisfied for this model based on the following diagnostic plots.

Linearity: There is a weak trend in the scatterplot. We are not provided the correlation coefficient nor the (R^2), therefore we will accept, with concerns, the linearity condition is satisfied.

Nearly normal residuals: As shown in the histogram and Q-Q plot, they are in fact nearly normal.

Constant variability: The scatterplot of the residuals does appear to have constant variability.

Independent observations: We do not have much information on how the data sample was collected beyond the fact that it was collected for 463 professors. We might assume independence of observations.

7.42 Babies

What is the predicted head circumference for a baby whose gestational age is 28 weeks?
Answer:
headCircumference=3.91+0.78∗gestationalAge headCircumference=3.91+0.78∗28 $headCircumference = 25.75 cm $
The standard error for the coefficient of gestational age is 0.35, which is associated with df = 23. Does the model provide strong evidence that gestational age is significantly associated with head circumference?
Answer:
As shown in the regression equation of the model the regression coefficient = 0.78 i.e. it is positive indicating there is a positive correlation between the 2 variables.

For lack of more information related to the p-values of the explanatory variables , R2 etc. We can perform a hypothesis test and check the evidence of significance association.
Null Hypothesis, Ho: There is no significant association.
Alternate Hypothesis, Ha: There is some significant association.
now that the n is small (n=23 or n<30), calculate t=(0.78-0)/0.35 = 2.229. for df=23 p-value from the t-table = 0.0178, which makes us reject Ho or indicating that there is strong relationship between the 2 variables in question.