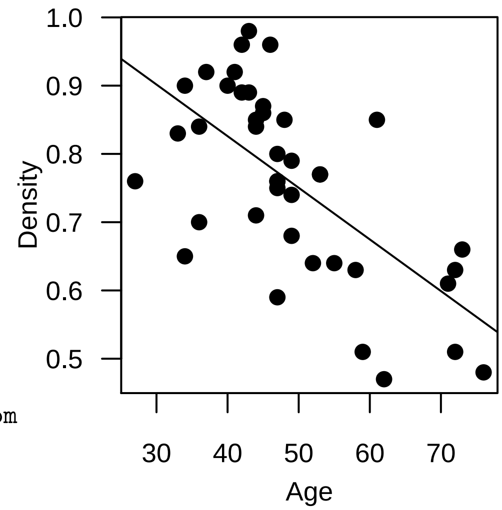
Researchers are interested in studying the change in bone density as adults age. To the right is a scatter plot of lumbar spine bone density (grams per cm^2) versus age (years) in a random sample of 41 American women smokers. Partial R output of fitting a regression line to predict bone density from age is below.




```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.128578   0.071565  15.770  < 2e-16 ***
age         -0.007565   0.001425  -5.308 4.72e-06 ***
---

Residual standard error: 0.1061 on 39 degrees of freedom
Multiple R-Squared: 0.4194,Adjusted R-squared: 0.4045
F-statistic: 28.18 on 1 and 39 DF,  p-value: 4.719e-06
```

(a) Circle the number closest to the correlation coefficient: -0.95 -0.65 -0.25 0 0.42 0.9 1.1 28.2.

Solution: We have $r = \pm\sqrt{4.194} = \pm 0.648$. The slope is negative and the trend is obvious in the plot — $r \approx -0.65$.

(b) Write the regression line as a formula in the style of the example (height in inches) $= 30.25 + 0.25$(age in months)

Solution:

$$\text{(bone density in g/cm}^2) = 1.129 - 0.007565\text{(age in years)}$$

(c) Construct a 95% confidence interval for the slope. Interpret the interval in the context of the problem.

Solution: There are 39 degrees of freedom. With the table, round down to 30 df to be safe.

$$-0.007565 \pm (2.042)(0.001425) \quad \text{or} \quad -0.0076 \pm 0.0029 \quad \text{or} \quad (-0.0105, -0.0047)$$

We are 95% confident that the mean bone density in this population of American, adult, women smokers decreases by between 0.005 and 0.01 grams per cm^2 for each increase in age by one year.

(Note, the units do not make sense to me either, but that is what it said in the book I used. Other books have a similar example where bone density is measured on thin slices by counting the number of osteocytes per unit area.)

(d) Use the regression line to predict the lumbar spine bone density of a 60-year-old woman from the population. Comment on the validity of this prediction.

Solution: $1.129 - 0.007565(60) = 0.675$ g/cm^2. The plot shows a somewhat weak linear relationship is reasonable. The prediction is within the range of the data and a good guess, although there is a fair amount of individual variability around the line.

(e) Use the regression line to predict the lumbar spine density of a 16-year-old girl. Comment on the validity of this prediction.

Solution: $1.129 - 0.007565(16) = 1.008$ g/cm^2. This prediction is dodgy. There is reason to think that a linear relationship from women aged 25 or so to 75 may not extend down to 16. This estimate is a potentially dangerous extrapolation.

The following questions do not relate to the data set. If the answer is FALSE, either make a small change so that the statement is TRUE or briefly explain why it is FALSE.

(f) Circle TRUE or FALSE: If the correlation coefficient of X and Y is close to zero, then X and Y cannot have a strong nonlinear relationship.

Solution: False. The statement is true about linear relationships. X and Y can have a strong nonlinear relationship with $r \approx 0$.

(g) Circle True or False: In simple linear regression, if the X value is one standard deviation above the mean, then the predicted Y value could be two standard deviations above the mean if the correlation coefficient was large enough.

Solution: False. The predicted value of Y is rz standard deviations from the mean if X is z stnadard deviations from the mean. Here, $z = 1$ and we would need $r = 2$. This is impossible.

(h) Circle True or False: In simple linear regression, if a plot of residuals versus fitted values shows a pattern of positive residuals at the ends and negative residuals in the middle, this indicates that the linear model fits the data well.

Solution: False. The pattern in the residual plot indicates potential non-linearity and that the model does not fit well.