Name: Solutions

Math 130
Linear Regression Analysis Quiz (Not Given)

1. In order to study the relationship between how much people smoke and how long it takes to develop lung cancer, 10 smokers who already have lung cancer were asked how many cigarettes they smoke in a day and how old they were when they first developed lung cancer. The data is summarized below.

No, of cigarettes smoked per day (x)	3	10	16	24	5	18	8	10	12	12
Age at which they first developed lung cancer (y)	68	61	48	44	68	49	59	59	54	52

In order to facilitate the rest of the calculations, here are some of the calculations already done for you:
$\sum x = 118$ $\sum x^2 = 1{,}742$ $\sum y = 562$ $\sum y^2 = 32{,}192$ $\sum xy = 6{,}194$

a) Find r. What does the sign of r tell you about the data?

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}} = \frac{10(6194) - (118)(562)}{\sqrt{10(1742) - (118)^2}\sqrt{10(32192) - (562)^2}}$$

$$\boxed{r = -0.9494732174}$$

r is negative, so when x increases, y decreases

If you smoke more cigarettes per day, the age at which you get lung cancer goes down.

b) Find the equation of the least squares regression line for this data

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{10(6194) - (118)(562)}{10(1742) - (118)^2} = -1.251716247$$

$$b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{(562)(1742) - (118)(6194)}{10(1742) - (118)^2} = 70.97025172$$

$$\hat{y} = b_1 x + b_0 \Rightarrow \boxed{\hat{y} = -1.251716247x + 70.97025172}$$

c) Find the best point estimate for the age at which a person who smokes 20 cigarettes per day is likely to develop lung cancer.

plug in $x_0 = 20$ into the least squares regression line equation from part b.

$\hat{y} = -1.251716247(20) + 70.97025172 \Rightarrow \boxed{\hat{y} \approx 45.9 \text{ years old}}$

d) Find a 95% prediction interval for the age at which a person who smokes 20 cigarettes per day is likely to develop lung cancer.

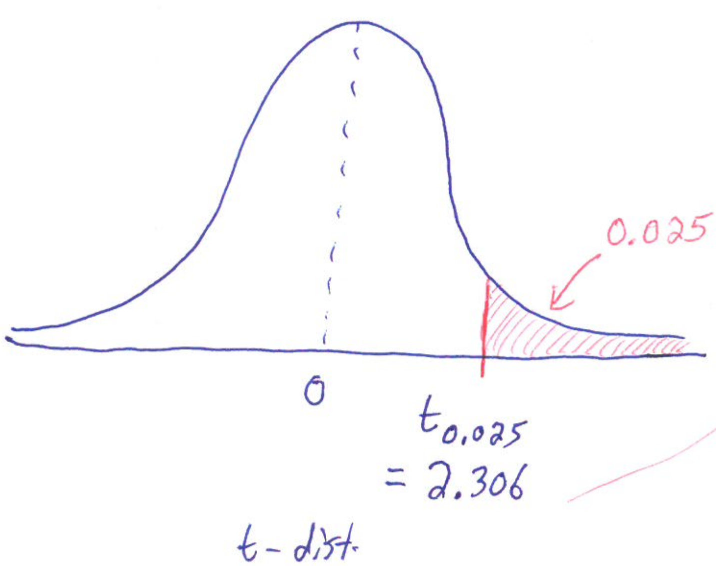
$t_{\alpha/2} = ?$

$\alpha = 1 - \text{conf. level} = 1 - 0.95 = 0.05$

$\alpha/2 = 0.05/2 = 0.025$

$df = n-2 = 10-2 = 8$

$t_{\alpha/2} = t_{0.025} = 2.306$



(part d continued...)

$s_e = ?$

$$s_e = \sqrt{\frac{\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy}{n-2}} = \sqrt{\frac{32,192 - (70.97025172)(562) - (-1.251716247)(6194)}{10-2}}$$

$$= 2.735163781$$

$E = ?$

$$\left(\bar{x} = \frac{3+10+\cdots+12}{10} = \frac{118}{10} = 11.8 \right)$$

$$E = t_{\alpha/2} \, s_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\Sigma x^2) - (\Sigma x)^2}}$$

$$= (2.306)(2.735163781)\sqrt{1 + \frac{1}{10} + \frac{10(20 - 11.8)^2}{10(1742) - (118)^2}}$$

$$= 7.170179717$$

Interval

$$\hat{y} - E < y < \hat{y} + E$$

$$45.9 - 7.17 < y < 45.9 + 7.17$$

$$\boxed{38.73 \text{ yrs old} < y < 53.07 \text{ yrs old}}$$

Some formulas you may need:

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

$$\hat{y} = b_1 x + b_0$$

$$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$b_0 = \frac{\left(\sum y\right)\left(\sum x^2\right) - \left(\sum x\right)\left(\sum xy\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$df = n - 2$$

$$s_e = \sqrt{\frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n-2}}$$

$$E = t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{n\left(x_0 - \bar{x}\right)^2}{n\left(\sum x^2\right) - \left(\sum x\right)^2}}$$