

Name: _____ Math 130 Day 24 In Class Worksheet
Date: _____ Sections 4.1, 4.2, and 14.2: Linear Regression Analysis

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:

$$\sum x = 118 \quad \sum x^2 = 1,742 \quad \sum y = 562 \quad \sum y^2 = 32,192 \quad \sum xy = 6,194$$

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

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

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.

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.

Some formulas you may need:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}} \quad \hat{y} = b_1x + b_0$$

$$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad b_0 = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad df = n - 2$$

$$s_e = \sqrt{\frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n - 2}} \quad E = t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\sum x^2) - (\sum x)^2}}$$