

Name _____

Please answer each of the following on these test pages in the blanks or spaces provided. Show any significant work (except arithmetic) for part credit. A calculator is recommended. Continue answers on the backs of these test pages if necessary.

1) "My understanding is ..."

(1a) Why is the "standard deviation" $\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$ defined in such a complicated way?
(Explain all the parts you can; continue on the back of this page if necessary.)

(1b) How can I remember which is greater in a skewed distribution, the mean or the median?
Why does it matter?

2) The data given at the right show 1975 average consumption (in grams per year) of tobacco per person over 15 years of age in ten European countries and 1990 lung cancer deaths (in deaths per 100,000). The question the data raise is whether there is an association between tobacco consumption in a given year and lung cancer deaths fifteen years later.

(2a) Do a stem-and-leaf plot of the data representing lung cancer deaths per 100,000 in 1990, and find the median. [Use 3, 4, 5, ... as the stems and round off the leaves.]

Stem-and-leaf:

Country	Tobacco consumption	Lung cancer deaths
Portugal	2255	30.1
Greece	3117	57.5
Ireland	3521	67.8
Germany	2804	57.9
France	2357	53.4
Belgium	3667	88.4
Italy	2517	67.5
Netherlands	3294	83.9
Denmark	2512	77.4
United Kingdom	2579	82.1

Median = _____

(2b) Produce the five number summary (minimum, lower quartile, median, upper quartile, maximum) for the rounded off lung cancer deaths data, and then sketch a box-and-whiskers plot for this data.

Five-number summary:

Box-and-whiskers:

(3) The least squares regression line for the Cancer/Tobacco data in (2) has the equation

$$y = 0.0194 x + 10.992,$$

with an $R^2 = 0.3078$, where y = Lung Cancer Deaths per 100,000 and x = Tobacco Consumption.

(3a) Use this regression equation to compute the number of deaths from lung cancer that would be predicted by this regression equation in a country that had an average tobacco consumption of 2700 grams per person per year.

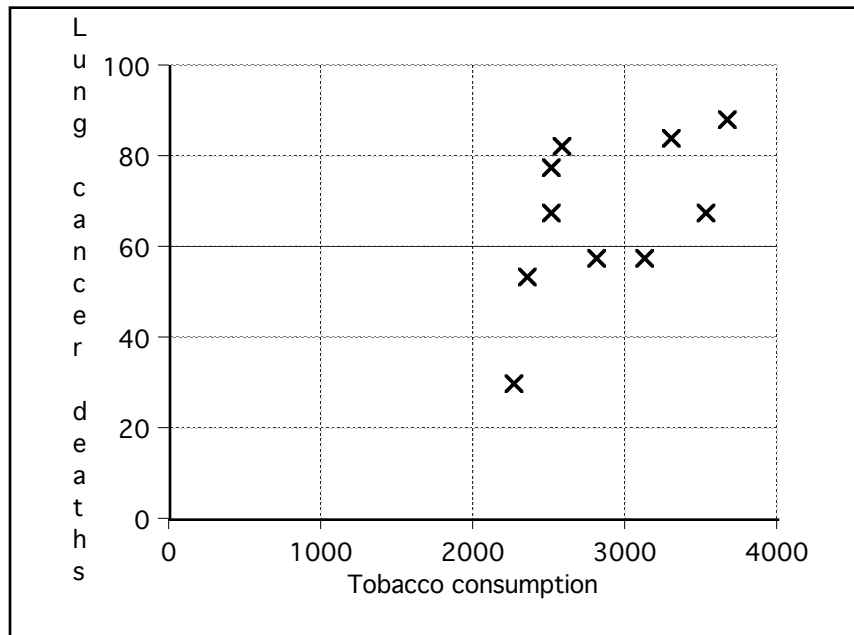
Lung cancer deaths per 100,000 = _____

(3b) Use information about the regression above to determine the correlation between Tobacco Consumption and Lung Cancer Deaths.

Correlation = _____

(3c) Draw accurately the regression line in the scatterplot at the right. [The equation of the line is given above.]

(3d) Circle an observation in the scatter plot you think is likely to have a large influence on the *slope* of the regression line, and explain what influence you think it has. What would happen to the line if your circled point were removed from the data?



(4) The table at the right shows data about the socioeconomic status of various categories of smokers and nonsmokers.

Smoking	Socioeconomic Status			Totals
	High	Middle	Low	
Current	51	22	43	116
Former	92	21	28	141
Never	68	9	22	99
Totals	211	52	93	356

Brenda claims that in this data people of high socioeconomic status are more likely than others to give up smoking.

Draw a bar graph (and write a little explanatory text to go with it) to support or refute *vividly* Brenda's claim. [What data you choose to show and what type of bar graph you chose will make a difference here. Be sure to label your scales.]