

D. Goodness of Fit and Correlation

D1. You are given 30 pairs of observations (X_i, Y_i) that are to be represented by the following model:

$$Y = \alpha + \beta X + \varepsilon,$$

where ε is a random error term with mean 0 and variance σ^2 . You have determined:

$$R = .5 \quad \overset{\text{variance}}{\bar{s}_X} = 7 \quad \bar{s}_Y = 6$$

Calculate the least-squares estimate of β . (Answer to nearest .1.)

A. .4 B. .5 C. .6 D. .7 E. .8 (85F-3-5)

D2. You are representing ten observations (X_i, Y_i) by the following model:

$$Y = \alpha + \beta X + \varepsilon,$$

where ε is a random error term with mean 0 and variance σ^2 . You have determined:

$$R^2 = .6 \quad \sum_{i=1}^{10} Y_i = 30 \quad \sum_{i=1}^{10} Y_i^2 = 690$$

Calculate s^2 . (Answer to the nearest integer.)

A. 27 B. 30 C. 36 D. 45 E. 50 (85F-3-10)

D3. You are to use simple linear regression to assess the impact of the unemployment rate on the duration of disability for injured workers. You have collected data on the unemployment rate (X_i) and the duration of disability expressed in days (Y_i) for 1967 through 1987. You have determined:

$$\sum_{i=1967}^{1987} X_i = 1.5095 \quad \sum_{i=1967}^{1987} Y_i = 1.599.84 \quad \sum_{i=1967}^{1987} (X_i - \bar{X})^2 = .01491$$

$$\sum_{i=1967}^{1987} (Y_i - \bar{Y})^2 = 761.43 \quad \sum_{i=1967}^{1987} (X_i - \bar{X})(Y_i - \bar{Y}) = 2.8396$$

Determine the sample correlation coefficient.

A. .25 B. .35 C. .50 D. .71 E. .84 (88S-120-6)

D4. You are given the following information concerning a simple linear regression:

$$\sum_{i=1}^8 (Y_i - \bar{Y})^2 = 435 \quad R^2 = .65$$

Calculate s^2 .

A. 19.0 B. 21.8 C. 25.4 D. 40.4 E. 47.1 (88F-120-6)

- D5. You are interested in the relationship between the price movements of XYZ Corporation and the market during the fourth quarter of 1987. You have used the least-squares criterion to fit the following line to fourteen weekly closing values of XYZ stock (Y_t) and the Dow Jones Industrial Average (X_t) during the period of interest:

$$\hat{Y}_t = -116.607 + .195X_t$$

You are given:

$$\sum_{t=1}^{14} (X_t - \bar{X})^2 = 17,807.080 \quad \sum_{t=1}^{14} (Y_t - \bar{Y})^2 = 949.388$$

Determine the percentage of variation in the value of XYZ stock that was explained by variations of the Dow.

- A. 50 B. 60 C. 70 D. 80 E. 90 (89S-120-7)

- D6. If a regression equation is not significant, \bar{Y} is the best point estimate of Y_i . (89S-120-10-MC) ~~9/10~~

- D7. You are given:

X	6.8	7.0	7.1	7.2	7.4
Y	.8	1.2	.9	.9	1.5

Determine R^2 , the coefficient of determination for the regression of Y on X.

- A. .3 B. .4 C. .5 D. .6 E. .7 (89F-120-8)

- D8. The coefficient of determination (R^2) is the fraction of the variation in Y that is explained by the linear relationship of Y with X. (90S-120-14-MC)

- D9. The coefficient of determination (R^2) is the ratio of the regression sum of squares to the total sum of squares. (90S-120-14-MC)

- D10. A bank is examining the relationship between income (X) and savings (Y). A survey of six randomly selected depositors yielded the following sample means, sample variances and sample covariance:

$$\bar{X} = 27.5 \quad \bar{Y} = 3.0 \quad \bar{s}_X^2 = 87.5 \quad \bar{s}_Y^2 = 3.5 \quad \bar{s}_{XY} = 17.0$$

Determine the proportion of the variation in Y about \bar{Y} that is explained by the linear relationship of Y with X.

- A. .1 B. .2 C. .3 D. .7 E. .9 (91S-120-6)

- D11. According to Pindyck and Rubinfeld, an $R^2 \geq .95$ is the generally accepted objective standard of quality. (91-10-21-MC)

- D12. You use simple linear regression and have observed the following five values of the dependent variable (Y): 1, 2, 3, 4, 5. You determine that $s^2 = 1$. Calculate R^2 .

- A. .1 B. .3 C. .5 D. .6 E. .7 (Sample-120-2)

- D13. You fit a simple linear regression model to five pairs of observations. The residuals of the first four observations are .4, -.3, 0, -.7, and the estimated variance of the dependent variable Y is

$$\frac{\sum (Y_i - \bar{Y})^2}{N - 1} = 1.5$$

Calculate R^2 .

- A. .82 B. .84 C. .86 D. .88 E. .90 (96S-120-3)
- D14. You wish to determine the nature of the relationship between sales (Y) and the number of radio advertisements broadcast (X). Data collected on four consecutive days is shown below.

Day	Sales	Number of Radio Advertisements
1	10	2
2	20	2
3	30	3
4	40	3

Using the method of least squares, you determine the estimated regression line:

$$\hat{Y} = -25 + 20X$$

Determine the value of R^2 for this model. (Sample-4-29)

- D15. In order to simplify an actuarial analysis, actuary A uses an aggregate distribution $S = X_1 + \dots + X_N$, where N has a Poisson distribution with mean 10 and $X_i = 1.5$ for all i. Actuary A's work is criticized because the actual severity distribution is given by:

$$\Pr(Y_i = 1) = \Pr(Y_i = 2) = .5, \text{ for all } i, \text{ where the } Y_i\text{'s are independent.}$$

Actuary A counters this criticism by claiming that the correlation coefficient between S and $S^* = Y_1 + \dots + Y_N$ is high. Calculate the correlation coefficient between S and S^* .

- A. .75 B. .80 C. .85 D. .90 E. .95 (01F-4-29)
- D16. You fit the following model to eight observations:
- $$Y = \alpha + \beta X + \epsilon$$
- You are given:
- $$\hat{\beta} = 2.065 \quad \sum (X_i - \bar{X})^2 = 43 \quad \sum (Y_i - \bar{Y})^2 = 182$$
- Determine R^2 .
- A. .48 B. .62 C. .83 D. .91 E. .98 (02F-4-5)
- D17. You fit a two-variable linear regression model to 20 pairs of observations. You are given:
- The sample mean of the independent variable is 100.
 - The sum of squared deviations from the mean of the independent variable is 2,266.
 - The ordinary least-squares estimate of the intercept parameter is 68.73.
 - The error sum of squares (ESS) is 5,348.

Determine the lower limit of the symmetric 95% confidence interval for the intercept parameter.

- A. -273 B. -132 C. -70 D. -8 E. -3 (02F-4-38)