

Linear Regression

Sample Exam 1 (Chapters 1 -2)

Show your work. Be neat. Be clear. Be precise.

1. (70 points). When anthropologists analyze human skeletal remains, an important piece of information is living stature (the height of a person when alive). Since skeletons are usually quite incomplete, inferences about stature are commonly based on statistical methods that utilize measurements on small bones.

The paper "The Estimation of Adult Stature from Metacarpal Bone Length" (Amer. J. of Phys. Anthro.(1978):113-120) presented data to validate one such method. Consider the following representative data, where x = metacarpal bone I length (cm) and y = stature (cm).

x	45	51	39	41	52	48	49	46	43	47
y	171	178	157	163	183	172	183	172	175	173

The summary statistics are

$$n = 10 \qquad \sum x = 461 \qquad \sum y = 1727$$

$$\sum xy = 79,886 \qquad \sum x^2 = 21,411 \qquad \sum y^2 = 298,843$$

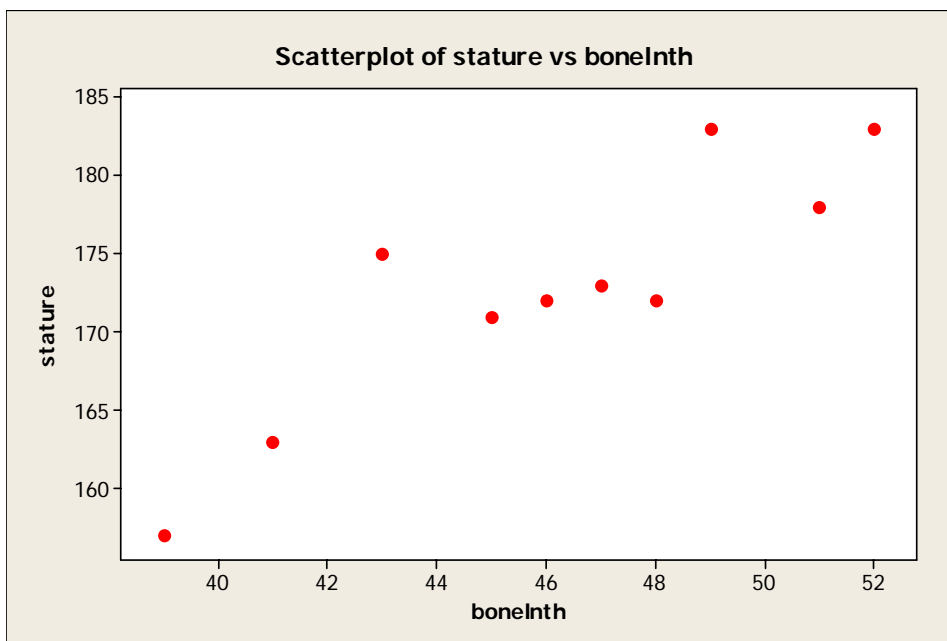
$$SSx = 158.9 \qquad SSy = 590.1 \qquad SSxy = 732.3$$

(a) Fit a simple linear regression (least squares) line to the data. (Find b_1 and b_0 , then write the estimated regression function.)

$$b_1 = \underline{\hspace{2cm}} \qquad \text{and} \qquad b_0 = \underline{\hspace{2cm}}$$

$$\text{estimated regression function: } \hat{Y} = \underline{\hspace{4cm}}.$$

(b) Below is a scatter plot of Y vs X. Draw your fitted line on the same plot. Comment on the fit.



(c) Fill in the following ANOVA table:

SOURCE	DF	SS	MS	F*
regression	___	_____	_____	_____
error	___	_____	_____	
total	___	_____		

(d) Calculate the coefficient of determination r^2 and interpret its value.

(e) Conduct a test to decide whether or not there is a linear association between stature and metacarpal bone length. Use a level of significance $\alpha = .05$. State the alternatives, test statistic, decision rule, and conclusion. Use either a t -test or an F -test.

Hypotheses: H_0 :

Test statistic and its value:

H_A :

Decision rule:

Conclusion:

Bounds on the p-value: _____ \leq p-value \leq _____

(f) Find a 95% confidence interval for the mean stature when the bone length equals 50 cm.

(g) A 95% prediction interval for stature when $x = 50$ would be wider than the corresponding confidence interval for mean stature when $x = 50$. Briefly explain why this occurs.

2. (30 points.) Fill in the blank(s) with an appropriate answer or circle the most appropriate answer in the problems below.

(a) For the simple linear model with normal errors ϵ_i , the errors are assumed to have mean $E(\epsilon_i) = \underline{\hspace{1cm}}$ and σ^2 $\{ \text{decreasing..... constant..... increasing} \}$ variance for all $i = 1, \dots, n$. Furthermore, the errors ϵ_i are assumed to be $\{ \text{dependent..... independent} \}$.

(b) In hypothesis testing, rejecting the null hypothesis when it is actually true is called a $\underline{\hspace{2cm}}$ error. The probability of this type of error is called the level of $\underline{\hspace{2cm}}$.

(c) In the general linear test, the null hypothesis specifies the $\{ \text{full.....reduced} \}$ model and the test statistic used is

$$F^* = \frac{\frac{\text{SSE}(\underline{\hspace{1cm}}) - \text{SSE}(\underline{\hspace{1cm}})}{\text{df}(\underline{\hspace{1cm}}) - \text{df}(\underline{\hspace{1cm}})}}{\frac{\text{SSE}(\underline{\hspace{1cm}})}{\text{df}(\underline{\hspace{1cm}})}} .$$

(d) The least square estimator b_1 is "best" in the sense that, among all linear unbiased estimators of β_1 , b_1 has minimum $\{ \text{expected value..... variance..... slope} \}$.

This result is part of the Gauss- $\underline{\hspace{1cm}}$ Theorem.

(e) The method of least squares regression involves finding a regression line that has the least $\{ \text{slope..... error sum of squares..... regression sum of squares} \}$.

(f) For the simple normal error regression model, $\frac{\text{SSE}}{\sigma^2}$ has a $\underline{\hspace{1cm}}$ distribution with $\underline{\hspace{1cm}}$ degrees of freedom. If $\beta_1 = 0$, then $\frac{\text{SSR}}{\sigma^2}$ also has a $\underline{\hspace{1cm}}$ distribution with $\underline{\hspace{1cm}}$ df, due to Cochran's Theorem.

Hence, under the null hypothesis that $\beta_1 = 0$, $\frac{\text{MSR}}{\text{MSE}}$ has an $\underline{\hspace{1cm}}$ distribution.

*3. Prove that, for a simple linear regression model, the least squares estimator b_1 is unbiased.