

Exam 1 is Wed., Sept 21. Quiz 4 Sept 28 covers Quiz 3 and HW4 material. **3 pages: Problems A)-H).**

Output for Problem 2.3

Label	Estimate	Std. Error	t-value	p-value
Constant	-5.07459	1.85124	-2.741	0.0076
log[H]	1.12399	0.498937	2.253	0.0270
log[S]	0.573167	0.116455	4.922	0.0000

R Squared: 0.895655 Sigma hat: 0.223658, n = 82

(log[H] log[S]) (4 5)

Prediction = 2.2872, s(pred) = 0.467664,

Estimated population mean value = 2.287, s = 0.410715

A) 2.3. The above output was produced from the file *mussels.lsp* in *Arc*. See Cook and Weisberg (1999a). Let $Y = \log(M)$ where M is the muscle mass of a mussel. Let $X_1 \equiv 1$, $X_2 = \log(H)$ where H is the height of the shell, and let $X_3 = \log(S)$ where S is the shell mass. Suppose that it is desired to predict Y_f if $\log(H) = 4$ and $\log(S) = 5$, so that $\mathbf{x}_f^T = (1, 4, 5)$. Assume that $se(\hat{Y}_f) = 0.410715$ and that $se(\text{pred}) = 0.467664$.

- If $\mathbf{x}_f^T = (1, 4, 5)$ find a 99% confidence interval for $E(Y_f)$.
- If $\mathbf{x}_f^T = (1, 4, 5)$ find a 99% prediction interval for Y_f .

B) 2.23 ijk (See HW3 C for the other parts of this problem.)

In *Arc* open the file *cyp.lsp*.

The response variable Y is *height*, and the explanatory variables are a constant, $X_2 = \text{sternal height}$ (probably height at shoulder), and $X_3 = \text{finger to ground}$.

Enter the menu commands “Graph&Fit>Fit linear LS” and fit the model: enter *sternal height* and *finger to ground* in the “Terms/Predictors” box, *height* in the “Response” box and click on *OK*.

i) The *Arc* menu “L1” should have been created for the regression. Use the menu commands “L1>Prediction” to open a dialog window. Enter 1400 650 in the box and click on *OK*. Include the resulting output in *Word*.

j) Let $X_{f,2} = 1400$ and $X_{f,3} = 650$ and use the output from i) to find a 95% CI for $E(Y_f)$. Use the last line of the output, that is, $se = S(\hat{Y}_f)$.

k) Use the output from i) to find a 95% PI for Y_f . Now $se(\text{pred}) = s(\text{pred})$.

C) 2.8 Suppose it is desired to predict the yearly return from the stock market from the return in January. Assume that the correlation $\hat{\rho} = 0.496$. Using the table below, find the least squares line $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$. Use $r = \hat{\rho}$.

variable	mean \bar{X} or \bar{Y}	standard deviation s
January return	1.75	5.36
yearly return	9.07	15.35

D) 2.9 Suppose that $\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 70690.0$, $\sum (X_i - \bar{X})^2 = 19800.0$, $\bar{X} = 70.0$, and $\bar{Y} = 312.28$.

- Find the least squares slope $\hat{\beta}_2$.
- Find the least squares intercept $\hat{\beta}_1$.
- Predict Y if $X = 80$.

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
38	41				
56	63				
59	70				
64	72				
74	84				

E) 2.10. In the above table, x_i is the length of the femur and y_i is the length of the humerus taken from five dinosaur fossils (*Archaeopteryx*) that preserved both bones. See Moore (2000, p. 99).

- Complete the table and find the least squares estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.
- Predict the humerus length if the femur length is 60.

F) 2.11 Suppose that the regression model is $Y_i = 7 + \beta X_i + e_i$ for $i = 1, \dots, n$ where the e_i are iid $N(0, \sigma^2)$ random variables. The least squares criterion is $Q(\eta) = \sum_{i=1}^n (Y_i - 7 - \eta X_i)^2$.

- What is $E(Y_i)$?
- Find the least squares estimator $\hat{\beta}$ of β by setting the first derivative $\frac{d}{d\eta} Q(\eta)$ equal to zero.
- Show that your $\hat{\beta}$ is the global minimizer of the least squares criterion Q by showing that the second derivative $\frac{d^2}{d\eta^2} Q(\eta) > 0$ for all values of η .

G) 2.26 : a) Activate the *cbrain.lsp* data set in *ARC* (as done in Problem 2.20 on HW1). Fit least squares with *age*, *sex*, $size^{1/3}$, and *headht* as terms and *brnweight* as the response. Assume that the multiple linear regression model is appropriate (this may be a reasonable assumption, 5 infants appear as outliers but the data set has hardly any cases that are babies. If *age* was uniformly represented, the babies might not be outliers anymore). Assuming that *ARC* makes the menu “L1” for this regression, select “AVP-All 2D.” A window will appear. Move the OLS slider bar to 1 and click on the “zero line box”. The window will show the added variable plots for *age*, *sex*, $size^{1/3}$, and *headht* as you move along the slider bar that is below “case deletions”. Include all 4 added variable plots in *Word*. b) What information do the 4 plots give? For example, which variables do not seem to be needed?

H) 2.29.

To download *lregpack.txt* and *lregdata.txt* (use the source commands as in HW 3).

a) Download the *R* functions `piplot` and `pisim` from *lregpack.txt*. (Done with the source commands.)

b) The command `pisim(n=100, type = 1)` will produce the mean length of the classical, semiparametric, conservative and asymptotically optimal PIs when the errors are normal, as well as the coverage proportions. Give the simulated lengths and coverages. Note: `pimenlen` gives the four lengths (classical, semi, ac, aopt) Make table with headers
classical semi ac aopt

c) Repeat b) using the command `pisim(n=100, type = 3)`. Now the errors are $EXP(1) - 1$.

d) Download *lregdata.txt* and type the command `piplot(cbrainx,cbrainy)`. This command gives the semiparametric PI limits for the Gladstone data. Include the plot in *Word*.

e) The infants are in the lower left corner of the plot. Do the PIs seem to be better for the infants or the bulk of the data. Explain briefly.