

[PRE] FINAL EXAM S2008 ANSWERS

FIRST NAME _____ LAST NAME _____ ANSWERS _____
(PRINT IN LARGE CAPITALS) (PRINT IN LARGER CAPITALS)

DATE: 2 MAY 2008

ANSWERS SHOULD BE CORRECT TO THREE SIGNIFICANT DIGITS

NOTATION: We use $SD(X)$ to denote the standard deviation of the unknown or random variable X , so $SD(X) = \sigma_X$.

Suppose that U, V, W, X, Y are all unknowns with $E(U)=11$, $E(V)=12$, $E(W)=13$, $E(X)=14$, $E(Y)=15$. Further, suppose that $SD(U)=3$, $SD(V)=4$, $SD(W)=5$, $SD(X)=6$, $SD(Y)=7$. Finally, suppose that the correlation coefficient of U with V is $.3$, the correlation coefficient of U with W is $.4$, the correlation coefficient of U with X is $.5$, and the correlation coefficient of X with Y is $\rho = .8$.

1. What is the expected value of $U+2V-Y$? $E(U+2V-Y) = 11+2*12-15=20$ _____
2. What is the variance of U ? $Var(U) = 3^2 = 9$ _____
3. What is the covariance of U with W ? $Cov(U,W) = .4*3*5 = (2/5)*3*5 = 6$ _____
4. What is the covariance of U with $V+2W$? $Cov(U,V+2W) = .3*3*4 + 2*6 = 3.6+12 = 15.6$ _____
5. If we perform regression of Y on X , getting the optimal linear equation of the form $y=a+bx$ then the optimal regression slope, β , which minimizes the expected squared error for predicting Y from values of X is
$$\beta = .8*(7/6) = 56/60 = 14/15 = .933333... \text{_____}$$
6. If we perform regression of Y on X , getting the optimal linear equation of the form $y=a+bx$ then what is the optimal regression intercept, α , which minimizes the expected squared error for predicting values of Y from values of X ?
$$\alpha = 15 - (14/15)*14 = 15 - (1-1/15)*14 = 1 + 14/15 = 29/15 = 1.9333... \text{_____}$$
7. What is the coefficient of determination for regression of Y on X ? $.8^2 = .64$ _____

We are trying to use weight to predict blood pressure in a population of adult people. In a sample of 20 people the mean weight was 150 pounds with a standard deviation of 15 pounds whereas the mean blood pressure was 130 with a standard deviation of 10. The sample correlation coefficient was .7. John Doe was not in the sample, but he weighs 162 pounds. His actual blood pressure, y , is unknown to us. We are going to use linear regression with the information about the sample data to make an optimal guess for y and estimate our margin of error.

9. What is the coefficient of determination for the sample data? .49

$$r^2 = .7^2 = .49$$

10. What is the sample estimate of the regression slope, b , based on our data? .4666...

$$b = .7 * 10 / 15 = 7 / 15 = .4666\dots$$

11. What is the estimate of the standard deviation of the sample correlation coefficient based on this sample data?

$$\sqrt{(1 - .49) / (20 - 2)} = \sqrt{17 / 600} = .1683250823$$

12. What is the estimate of the standard deviation of the error divided by the square root of the sample size?

$$10 * (\text{ANSWER TO \#11}) = \sqrt{17 / 6} = 1.683250823$$

13. What is the estimate of the standard deviation of b , the regression slope?

$$(\text{ANSWER TO 12}) / 15 = \sqrt{17 / 1350} = .1122167215$$

14. What is the z-score of John Doe's weight? $Z = (162 - 150) / 15 = .8$

15. What should our optimal guess, w , be for John Doe's blood pressure? 135.6

$$W = 130 + (.7) * 10 * Z = 135.6$$

OR, calculate the regression intercept $a = 130 - (7 / 15) * 150 = 60$, then

$$W = 60 + (7 / 15) * 162 = 135.6$$

16. What should be our estimate of the standard deviation in w , our optimal guess for John Doe's blood pressure?

$$(\text{ANSWER TO 12}) * \sqrt{1 + Z^2} = 2.155612829$$

17. What should be our estimate of the standard deviation in the difference, $y-w$ between John Doe's actual blood pressure, y , and our optimal guess, w , for his blood pressure?

$$\underline{\underline{(\text{ANSWER TO 12}) \cdot \text{SQRT}(20+1+Z^2) = \text{SQRT}(697/150) = 7.830283094 \underline{\underline{\quad}}}}$$

18. If we use a level of confidence so that $t=2.5$, then what should be the margin of error in our confidence interval for using our optimal guess, w , to guess John Doe's actual blood pressure, y ?

$$\underline{\underline{2.5 \cdot (\text{ANSWER TO 17}) = \text{SQRT}(5748125/150000) = 19.57570774 \underline{\underline{\quad}}}}$$

We are using height, weight, shoe size, waist measurement, and hat size to predict blood pressure in a population of adult people. In a sample of 20 people, we find that our data has a coefficient of determination equal to .73 whereas if we ignore the shoe size and hat size, the coefficient of determination drops down to .65.

19. What is the F-score for evaluating the full model for predicting blood pressure with all 5 variables, and what is the p-value as evidence that at least one of these variables has non-zero regression coefficient?

$$\underline{\underline{F\text{-score} = (.73/5) / (.27/14) = 1022/135 = 7.57037037 = F \underline{\underline{\quad}}}}$$

$$\underline{\underline{p\text{-value} = 1 - Fcdf(0, F, 5, 14) = .0012438404 \underline{\underline{\quad}}}}$$

20. What is the F-score for comparing the full linear regression model with the model obtained by eliminating the shoe size and hat size variables, and what is the p-value of the data as evidence that at least one of the regression coefficients for the eliminated variables is non-zero?

$$\underline{\underline{F\text{-score} = (.08/2) / (.27/14) = 56/27 = 2.074074074 \underline{\underline{\quad}}}}$$

$$\underline{\underline{p\text{-value} = 1 - Fcdf(0, 56/27, 2, 14) = .1625810929 \underline{\underline{\quad}}}}$$

21. **EXTRA CREDIT (5 points).** If W is the resulting variable we get from optimal linear regression which gives blood pressure in terms of height, weight, shoe size, waist measurement, and hat size using the optimal regression coefficients $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$, then what is the best estimate of the correlation coefficient between Y and W based on the data?

$$\underline{\underline{\text{SQRT}(.73) = .8544003745 \underline{\underline{\quad}}}}$$



*Your complimentary
use period has ended.
Thank you for using
PDF Complete.*

[*Click Here to upgrade to
Unlimited Pages and Expanded Features*](#)