

## Module 8: Simple linear regression practice problems

(The attached PDF file has better formatting.)

### LINEAR REGRESSION: PRACTICE EXAM PROBLEMS

This posting illustrates linear regression exam problems covering the basic formulas. On the final exam, expect a scenario with five pairs of points similar to the exercise below. The problem derives the ordinary least squares estimators, their standard errors, t-values, levels of significance, and F-statistic. Some statistical items are taught in later modules; these practice problems covers many items in basic regression analysis.

An actuary fits a two-variable regression model ( $Y_i = \alpha + \beta \times X_i + \epsilon_i$ ) to the relation between the incurred loss ratio ( $x$ ) and the retrospective ratio ( $y$ ), using the data below:

Policy Year	( $x$ )	( $y$ )	( $x - \bar{x}$ )	( $x - \bar{x}$ ) <sup>2</sup>	( $y - \bar{y}$ )	( $y - \bar{y}$ ) <sup>2</sup>	( $x - \bar{x}$ )( $y - \bar{y}$ )
20X1	61.00%	15.00%	0.00%	0.00%	0.32%	0.0010%	0.0000%
20X2	62.00%	13.20%	1.00%	0.01%	-1.48%	0.0219%	-0.0148%
20X3	63.00%	14.00%	2.00%	0.04%	-0.68%	0.0046%	-0.0136%
20X4	60.00%	15.20%	-1.00%	0.01%	0.52%	0.0027%	-0.0052%
20X5	59.00%	16.00%	-2.00%	0.04%	1.32%	0.0174%	-0.0264%
Average	61.00%	14.68%	0.00%	0.02%	0.00%	0.009536%	-0.01200%

The column captions use lower case  $x$  and  $y$  for the variables; the deviations are shown explicitly as  $(x - \bar{x})$  and  $(y - \bar{y})$ . Some statistician use upper case letter for the variables and lower case letters for the deviations.

*Take heed:* The notation in the John Fox regression analysis text differs slightly from the notation in some of the discussion forum postings.

- John Fox uses the symbols  $A$  and  $B$  as the least squares estimators for  $\alpha$  and  $\beta$ .
- He uses  $RSS$  for the residual sum of squares; other authors use  $ESS$ , the error sum of squares.
- He uses  $RegSS$  for the regression sum of squares; other authors use  $RSS$ .

The final exam problems use Fox's notation.

Question 8.1: Ordinary Least Squares Estimator of  $\beta$

What is the value of  $B$ , the ordinary least squares estimator of  $\beta$ ?

- A. -0.600
- B. -0.120
- C. -0.020
- D. -0.019
- E. -0.012

Answer 8.1: A

The table gives the sum of the cross-product terms and of the squared deviations of  $X$ .

$$B = \sum(x_i - \bar{x})(y_i - \bar{y}) / \sum(x_i - \bar{x})^2 = -0.012 / 0.020 = -0.600$$

*Note:* The last row of the table shows averages. The ratio of the averages is the ratio of the sums.

Question 8.2: Ordinary Least Squares Estimator of  $\alpha$

What is the value of  $A$ , the ordinary least squares estimator of  $\alpha$ ?

- A. -0.6100
- B. -0.1468
- C. +0.1468
- D. +0.5128
- E. +0.6100

Answer 8.2: D

Use the relation:  $A = \bar{y} - B \times \bar{x} = 14.68\% - (-0.60) \times 61.00\% = 0.5128$

Question 8.3: Total Sum of Squares (TSS)

What is the total sum of squares (TSS)?

- A. 0.0117%
- B. 0.0360%
- C. 0.0477%
- D. 0.0833%
- E. 0.1310%

Answer 8.3: C

The total sum of squares can be found two ways.

(1) We subtract the mean of Y from each observed value and square the deviations:

$$\sum (y_i - \bar{y})^2 = 0.32\%^2 + (-1.48\%)^2 + (-0.68\%)^2 + 0.52\%^2 + 1.32\%^2 = 0.04768\%$$

(2) We square the observed values of Y and subtract N times the square of the mean:

$$\sum Y^2 - N\bar{Y}^2 = \sum Y^2 - (\sum Y)^2 / N$$

$$= 15\% + 13.2\% + 14\% + 15.2\% + 16\% - 14.68\%^2 / 5 = 0.04768\%$$

The table in the exam problem gives the TSS as  $5 \times 0.009536\% = 0.04768\%$

Question 8.4: Regression Sum of Squares (RegSS)

What is the regression sum of squares (RegSS)?

- A. 0.0117%
- B. 0.0360%
- C. 0.0477%
- D. 0.0833%
- E. 0.1310%

Answer 8.4: B

Find the fitted Y value at each observation as  $A + B \times X$ . Subtract the mean of Y and square the result. The sum of these is the regression sum of squares.

Policy Year	(x)	(y)	$\hat{y}$	$(\hat{y} - \bar{y})$	$(\hat{y} - \bar{y})^2$
20X1	61.00%	15.00%	14.68%	0.00%	0.0000%
20X2	62.00%	13.20%	14.08%	-0.60%	0.0036%
20X3	63.00%	14.00%	13.48%	-1.20%	0.0144%
20X4	60.00%	15.20%	15.28%	0.60%	0.0036%
20X5	59.00%	16.00%	15.88%	1.20%	0.0144%
Average	61.00%	14.68%	14.68%	0.00%	0.007200%

$$5 \times 0.0072\% = 0.0360\%$$

$$\text{A quick formula: regression sum of squares (RegSS)} = B^2 \times \sum(x_i - \bar{x})^2$$

$$\sum(x_i - \bar{x})^2 = 1\%^2 + 2\%^2 + (-1\%)^2 + (-2\%)^2 = 0.10\%$$

$$\text{RegSS} = B^2 \times 0.10\% = 0.6^2 \times 0.10\% = 0.0360\%$$

$$F = t^2 = \text{RegSS} / s^2 \Rightarrow \text{RegSS} = t^2 \times s^2 = B^2 \times \sum(x_i - \bar{x})^2$$

Question 8.5: Error Sum of Squares (ESS) or Residual Sum of Squares (RSS)

What is the error sum of squares (ESS) or residual sum of squares (RSS)

- A. 0.0117%
- B. 0.0360%
- C. 0.0477%
- D. 0.0833%
- E. 0.1310%

Answer 8.5: A

We compute the residual sum of squares two ways.

(1) The ESS (RSS) is the TSS minus the RegSS.

$$0.04768\% - 0.0360\% = 0.01168\%$$

(2) We determine residuals as the observed Y minus the fitted Y. The sum of the squared residuals is the residual sum of squares (RSS).

<i>Policy Year</i>	(x)	(y)	$\hat{y}$	$(\hat{y} - y)$	$(\hat{y} - y)^2$
20X1	61.00%	15.00%	14.68%	-0.32%	0.0010%
20X2	62.00%	13.20%	14.08%	0.88%	0.0077%
20X3	63.00%	14.00%	13.48%	-0.52%	0.0027%
20X4	60.00%	15.20%	15.28%	0.08%	0.0001%
20X5	59.00%	16.00%	15.88%	-0.12%	0.0001%
Average	61.00%	14.68%	14.68%	0.00%	0.002336%

$$5 \times 0.002336\% = 0.011680\%$$

Question 8.6: Standard Error

What is  $s^2$ , the estimated variance of the regression?

- A. 0.0036%
- B. 0.0039%
- C. 0.0360%
- D. 0.0389%
- E. 0.0117%

Answer 8.6: B

The estimated variance of the regression is the residual sum of squares divided by the degrees of freedom (the number of observations minus the number of explanatory variables). The explanatory variables are the independent variables plus the constant term. For a simple linear regression, this is  $N-2$ :  $0.01168\% / 3 = 0.003893\%$ . This is an unbiased estimate of  $\sigma^2$ .

Question 8.7: Variance of Ordinary Least Squares Estimator of  $\beta$

What is the *variance* of the ordinary least squares estimator of  $\beta$ ? (This is the variance, not the standard error.)

- A. 0.36%
- B. 0.39%
- C. 3.60%
- D. 3.89%
- E. 1.17%

Answer 8.7: D

The variance of  $B$  is the  $\sigma^2$  (or its unbiased estimate) divided by  $\sum(x_i - \bar{x})^2$ :

$$0.00389\% / 0.10\% = 3.890\%$$



Question 8.8: *t* Statistic

What is the *t* statistic for testing the null hypothesis that  $\beta = 0$ ?

- A. -3
- B. -2
- C. -1
- D. +1
- E. +2

Answer 8.8: A

The *t* statistic is (the difference between *B* and the null hypothesis) divided by the standard deviation of *B*, which is the square root of the variance of *B*:

$$-0.6 / 3.890\%^{\frac{1}{2}} = -3.042$$

Question 8.9:  $p$ -value

The  $p$ -value for  $\beta$  for this regression equation is 0.0558. Which of the following is true, assuming the classical regression assumptions hold?

- A. The true  $\beta$  is within  $\pm 5.58\%$  (multiplicative) of the ordinary least squares estimator.
- B. The true  $\beta$  is within  $\pm 0.0558$  (additive) of the ordinary least squares estimator.
- C. The probability is 95% that the true  $\beta$  is within  $\pm 0.0558$  of the ordinary least squares estimator.
- D. If the true value of  $\beta$  is zero, the probability that the absolute value of the ordinary least squares estimator of  $\beta$  is at least as great as in this regression equation is 5.58%.
- E. If the true value of  $\beta$  is zero, the probability is 95% that the absolute value of the ordinary least squares estimator of  $\beta$  is no more than 0.0558.

Answer 8.9: D

To test hypotheses, we consider the probability that we would observe an ordinary least squares estimator as far from the null hypothesis (or farther) because of sampling error. The  $p$ -value gives this probability, as stated in Statement D.

Question 8.10: *F* Statistic

What is the *F* statistic for testing the null hypothesis that  $\beta = 0$ ?

- A. -4
- B. -1
- C. +1
- D. +4
- E. +9

Answer 8.10: E

We compute the *F* statistic two ways.

(1) For a two-variable regression model, the *F* statistic is the square of the *t* statistic:

$$-0.6^2 / 3.893\% = 9.247$$

(2) The *F* statistic is the ratio of the regression sum of squares divided by its degrees of freedom to the error sum of squares divided by its degrees of freedom:

$$[0.0360\% / 1] / [0.01168\% / 3] = 9.247$$

Question 8.11:  $R^2$

What is the value of  $R^2$ , the coefficient of determination?

- A. 55%
- B. 65%
- C. 75%
- D. 85%
- E. 95%

Answer 8.11: C

$$R^2 = 0.036\% / 0.04768\% = 75.50\%$$