

Week Three Homework Assignment - Linear Regression

The personnel director for a local manufacturing firm has received complaints from the employees in a certain shop regarding what they perceive to be inequities in the annual salary for employees who have similar performance ratings, years of service and relevant certifications. The personnel director believes that an employee's pay in this particular shop should be positively correlated to their prior performance rating, years of service and relevant certifications. The personnel director has collected the data shown in the following table pertaining to the employees within the shop.

Employee	Current Annual Salary (Thousands)	Average Performance Rating for Past 3 Years (5 point scale)	Years of Service	Number of Relevant Certifications
1	48.2	2.18	9	6
2	55.3	3.31	20	6
3	53.7	3.18	18	7
4	61.8	3.62	33	7
5	56.4	2.62	31	8
6	52.5	3.75	13	6
7	54.0	4.25	25	6
8	55.7	3.43	30	4
9	45.1	1.93	5	6
10	67.9	4.5	47	8
11	53.2	2.81	25	5
12	46.8	3.06	11	6
13	58.3	5	23	8
14	59.1	4.06	35	7
15	57.8	4.12	39	5
16	48.6	2.31	21	4
17	49.2	3.87	7	6
18	63.0	4.37	40	7
19	53.0	2.5	35	6
20	50.9	2.81	23	4
21	55.4	3.68	33	5
22	51.8	3.5	27	4
23	60.2	3	34	8
24	50.1	2.43	15	5

The personnel director is interested in creating a linear regression model that can be used to estimate the annual salary an employee might expect to receive based upon his or her past performance, years of service and/or number of relevant certifications. The regression model will be used as a basis for determining whether or not there is any validity to the employees' complaints regarding salary inequities.

Perform each of the following seven regression analyses using a 95% confidence level.

- Annual salary vs. average performance rating for the past 3 years
- Annual salary vs. years of service
- Annual salary vs. number of relevant certifications
- Annual salary vs. average performance rating for the past 3 years and years of service
- Annual salary vs. average performance rating for the past 3 years and number of relevant certifications
- Annual salary vs. years of service and number of relevant certifications
- Annual salary vs. average performance rating for the past 3 years, years of service and number of relevant certifications

Hint: Refer to the handouts posted on Blackboard pertaining to interpreting regression statistics in order to determine if a given regression model is acceptable. This same handout also provides guidance regarding how to select a preferred regression model from amongst multiple acceptable regression models, including models with differing numbers of independent variables.

Hint: For the purposes of this homework assignment, the minimum difference between the R^2 or Adjusted R^2 values for two acceptable models with differing numbers of independent variables that would favor selecting the model with the larger number of independent variables is 0.03. Please ensure that you fully understand the process for selecting a preferred model before attempting to apply this criterion.

Hint: Question 19 is intended to have you demonstrate that you understand how to determine which univariate models are acceptable, and then select a preferred univariate model from amongst the acceptable univariate models. Question 20 is intended to have you demonstrate that you understand how to determine which bivariate models are acceptable, and then select a preferred bivariate model from amongst the acceptable bivariate models. Question 21 is intended to have you demonstrate that you understand how to select a preferred model from amongst multiple acceptable models that have differing numbers of independent variables. Question 22 is intended to have you demonstrate that you understand how to determine if the trivariate model is acceptable, and then select a preferred model from amongst multiple acceptable models.

Hint: For questions 24 and 25, you need to use the regression equation associated with the preferred model selected for question 22 in order to calculate the predicted salary for each of the 24 employees. In order to answer questions 24 and 25 you need to keep in mind that the predicted salary value for each employee is only a point estimate (this

concept was discussed in week one relative to the mean). While a point estimate is a precise value, it is not necessarily an accurate value since the standard error value tells us there is some potential degree of error associated with using the preferred regression model to predict salary values. In order to answer questions 24 and 25 you will need to create an interval estimate (this concept was also discussed during week one relative to the mean) for the predicted salary for each of the 24 employees. To calculate the interval estimate for each employee, simply multiply the standard error value for the preferred regression model by 1.5 and then subtract this value from the predicted point estimate salary value to define the lower limit of the interval estimate and add this value to the predicted point estimate salary value to define the upper limit for the interval estimate. Once you have created an interval estimate for each employee, you will then need to compare each employee's current salary to their corresponding interval estimate in order to determine if each employee's current salary falls within their predicted interval estimate.

Use the results for the univariate regression analysis for annual salary vs. average performance rating for the past 3 years in order to answer questions 1 through 14.

1. What is the degree of correlation between the dependent variable and the independent variable?
 - ☐ 0.8198
 - ☐ 0.6672
 - ☐ 0.7862
 - ☐ 0.6523
2. Does the regression model confirm a positive correlation between the dependent variable and the independent variable as hypothesized?
 - ☐ Yes
 - ☐ No
3. What is the desired statistical significance for the regression model?
 - ☐ 0.00
 - ☐ 0.01
 - ☐ 0.05
 - ☐ 0.10
4. Is the statistical significance of the model as a whole less than the desired statistical significance for the regression model?
 - ☐ Yes
 - ☐ No
5. What is the actual confidence level for the regression model as a whole?
 - ☐ 99.97%
 - ☐ 95%
 - ☐ 90.5%
 - ☐ 80.2%

6. Is the statistical significance of the linear relationship between the dependent and independent variables less than the desired statistical significance for the regression model?
 - ☐ Yes
 - ☐ No
7. Should the coefficient of determination or adjusted coefficient of determination be used to evaluate this regression model?
 - ☐ Coefficient of determination
 - ☐ Adjusted coefficient of determination
8. What percentage of the observed variation between the actual values of the dependent variable and the mean value of the dependent variable in the sample data set is explained by the regression model?
 - ☐ 44.51%
 - ☐ 66.72%
 - ☐ 41.99%
 - ☐ 76.24%
9. What is the amount by which we will be off on average when predicting values for the dependent variable using the regression model?
 - ☐ \$12,287
 - ☐ \$32,966
 - ☐ \$4,169
 - ☐ \$25,896
10. What is the coefficient for the y-intercept for the regression model?
 - ☐ 39.38
 - ☐ 19.94
 - ☐ 12.29
 - ☐ 2.96
11. What is the coefficient for the independent variable for the regression model?
 - ☐ 63.08
 - ☐ 4.52
 - ☐ 12.29
 - ☐ 2.96
12. What is the point estimate for the predicted salary for an employee with an average performance rating of 3.9?
 - ☐ \$71,562
 - ☐ \$57,006
 - ☐ \$41,299
 - ☐ \$50,896

13. What is the interval estimate for the predicted salary for an employee with an average performance rating of 3.9 based upon taking into consideration the standard error?
- ☐ \$61,913 - \$71,562
 - ☐ \$57,006 - \$64,159
 - ☐ \$52,837 - \$61,175
 - ☐ \$54,896 - \$60,873

14. What is the 95% confidence level interval estimate for the salary for an employee with an average performance rating of 3.9?
- ☐ \$60,265 - \$85,789
 - ☐ \$75,412 - \$78,523
 - ☐ \$40,636 - \$73,376
 - ☐ \$65,141 - \$72,269

Perform a correlation analysis between the dependent variable and each of the three independent variables. Use the results of the correlation analysis to answer questions 15 and 16.

15. Which independent variables evidence a positive correlation with the dependent variable?
- ☐ Average performance rating for the past 3 years
 - ☐ Years of service
 - ☐ Number of relevant certifications
 - ☐ All of the above
 - ☐ None of the above
16. Which independent variable evidences the highest degree of correlation with the dependent variable?
- ☐ Average performance rating for the past 3 years
 - ☐ Years of service
 - ☐ Number of relevant certifications
 - ☐ All of the above
 - ☐ None of the above

Perform a correlation analysis between each of the three pairs of independent variables. Use the results of the correlation analyses to answer question 17.

17. Which pair of independent variable evidences a degree of collinearity that should be cause for concern when performing multivariate linear regression (i.e., evidences a degree of correlation in excess of 0.5)?
- ☐ Average performance rating for the past 3 years vs. years of service
 - ☐ Average performance rating for the past 3 years vs. number of relevant certifications
 - ☐ Years of service vs. number of relevant certifications
 - ☐ All of the above
 - ☐ None of the above

Use the regression statistics pertaining to all seven regression analyses in order to answer questions 15 through 23.

18. Of the seven regression models, which model both accounts for the lowest percentage of the observed variation between the actual values of the dependent variable and the mean value of the dependent variable in the sample data set and evidences the highest degree of error for predicting values for the dependent variable?
- ☐ Annual salary vs. average performance rating for the past 3 years
 - ☐ Annual salary vs. years of service
 - ☐ Annual salary vs. number of relevant certifications
 - ☐ Annual salary vs. average performance rating for the past 3 years and years of service
 - ☐ Annual salary vs. average performance rating for the past 3 years and number of relevant certifications
 - ☐ Annual salary vs. years of service and number of relevant certifications
 - ☐ Annual salary vs. average performance rating for the past 3 years, years of service and number of relevant certifications
19. If you were to consider only the three regression models that are based upon a single independent variable, which of the following models would be your preferred model?
- ☐ Annual salary vs. average performance rating for the past 3 years
 - ☐ Annual salary vs. years of service
 - ☐ Annual salary vs. number of relevant certifications
20. If you were to consider only the three regression models that are based upon two independent variables, which of the following models would be your preferred model?
- ☐ Annual salary vs. average performance rating for the past 3 years and years of service
 - ☐ Annual salary vs. average performance rating for the past 3 years and number of relevant certifications
 - ☐ Annual salary vs. years of service and number of relevant certifications
21. If you were to compare the preferred regression model based upon a single independent variable with the preferred regression model based upon two independent variables, which model would be preferred overall?
- ☐ Preferred regression model based upon a single independent variable
 - ☐ Preferred regression model based upon two independent variables
22. When you consider all seven regression models, which is the overall preferred regression model?
- ☐ Annual salary vs. average performance rating for the past 3 years
 - ☐ Annual salary vs. years of service
 - ☐ Annual salary vs. number of relevant certifications

- Annual salary vs. average performance rating for the past 3 years and years of service
- Annual salary vs. average performance rating for the past 3 years and number of relevant certifications
- Annual salary vs. years of service and number of relevant certifications
- Annual salary vs. average performance rating for the past 3 years, years of service and number of relevant certifications

23. Do any of the regression models offer a higher confidence level for the model as a whole, or a lower standard error in comparison to the overall preferred model?
- Yes
 - No

The personnel is interested in comparing each employee's actual salary to their predicted salary in order to determine if there are any prevailing salary inequities. Suppose the personnel director considers an employee's current salary to be fair and reasonable if it is within plus or minus 1.5 standard errors of the value estimated by the regression model selected in response to question 22. For each individual employee, calculate his or her estimated salary using the regression model selected in response to question 22, as well calculate his or her upper and lower limits for a fair reasonable salary, in order to answer questions 24 and 25.

24. Of the 24 employees, how many employees' current salary is below what is considered fair and reasonable?
- 0
 - 1
 - 2
 - 3
 - 4
 - 5

25. Of the 24 employees, how many employees' current salary is above what is considered fair and reasonable?
- 0
 - 1
 - 2
 - 3
 - 4
 - 5