Week 2 Application Assignment

Quiz, 13 questions

Congratulations! You passed!

Next Item

1 / 1 point

1.
In this assignment we are continuing to work with customer reward programs (review assignments from Week 1 if you haven't completed them). The data is in the file

crp_cleandata.xlsx

In this exercise, you will complete a predictive modeling task where the target variable is continuous based on the data in the shared file. First remove all rows where either the Reward or NumStores column takes the value 0. Also remove all rows where the rewards do not expire (ExpirationMonth=999). [Hint: You can sort the relevant columns to quickly find the rows to delete.] How many rows are left after deleting these irrelevant rows, not counting the header row? What is the sum of the ExpirationMonth column?

- 48, 1335
- 48, 336
- 46, 1335
- 46, 336 (selected)

Correct
Bravo!

---

1 / 1 point

2.

Consider linear regression models with ExpirationMonth column as the target variable. Find the model with one predictor variable and the highest R-squared. Consider the following set of predictor variables: Salerank, X2013USSales, X2013WorldSales, NumStores,RewardSize, and ProfitMargin. Which variable did you choose?

- X2013USSales
- X2013WorldSales
- Salerank
- NumStores

Correct

1 / 1 point

3.

What is the estimated intercept coefficient of the model?

- 5.7082
- 4.8285

Correct
Bravo!

- 0.8898
- 0.2537

1 / 1 point

4.

What is the estimated slope coefficient of the model?

- 34.9427
- 4.8285
- 0.8898

Correct
Bravo!

Week 2 Application Assignment

Quiz, 13 questions

0.2537

---

1 / 1 point

5.

Data transformation is a great way to improve model fit. Now consider the log transformation for the model identified in the previous question. [Hint: Use log function to create the transformed columns.] You can choose to transform neither of them, one of them, or both of them. You should have four different models.

- Model 1: neither variable is transformed; this gives you the same model as in the previous question.
- Model 2: only the target variable is transformed
- Model 3: only the explanatory variable is transformed
- Model 4: both variables are transformed.

Report the R-squared values of all four models.

What is the R-squared for Model 1?

4.8285

0.8898

0.2537

Correct
Bravo!

6.6175

---

1 / 1 point

6.

R-squared for Model 2 is (report answer using 4 decimal places **i.e. x.xxxx**):

0.0704

Correct Response
Bravo!

Week 2 Application Assignment

Quiz, 13 questions

1 / 1 point

7.
R-squared for Model 3 is (report answer using 4 decimal places **i.e. x.xxxx**):

0.1446

Correct Response
Bravo!

1 / 1 point

8.
R-squared for Model 4 is (report answer using 4 decimal places **i.e. x.xxxx**):

0.0652

Correct Response
Bravo!

1 / 1 point

9.
Which model gives the best fit based on the R-squared value?

- [x] Model 1

Correct

- [ ] Model 2
- [ ] Model 3

Model 4

Week 2 Application Assignment

Quiz, 13 questions

1 / 1 point

10.

Our analysis so far shows that variable transformation does not improve the model fit. Another way to improve model fit is to add more explanatory variables on the right side. Again consider the following set of predictor variables: Salerank, X2013USSales, X2013WorldSales, NumStores, RewardSize, and ProfitMargin. Add one more variable to the best model you identified in the previous question. Which variable will you add? Hint: The correct additional variable gives the highest R-squared value.

- RewardSize
- X2013USSales

Correct
Bravo!

- ProfitMargin
- Salerank

1 / 1 point

11.

What is the R-squared for the model with the additional variable added (report answer using 4 decimal places **i.e. x.xxxx**)?

0.2844

Correct Response
R-squared value (keep four significant digits): 0.2845

Week 2 Application Assignment

Quiz, 13 questions

1 / 1 point

12.

One way to figure out whether a linear regression model explains a particular data point well is to look at the residual. For which retailer do you have the highest absolute value of residual based on your result in the previous question?

- Macy's
- Whole Foods
- TJX

Correct

Bravo.

To find the residual for each retailer, check Residuals - Unstandardized in the second step of multiple linear regression. A new sheet will be created by XLMiner, reporting the residual values for each row.

- Starbucks

---

1 / 1 point

13.

For which retailer do you have the lowest residual based on your result in the previous question?

- Macy's
- Whole Foods
- TJX
- Starbucks

Correct

Bravo

To find the residual for each retailer, check Residuals - Unstandardized in the second step of multiple linear regression. A new sheet will be created by XLMiner, reporting the residual values for each row.

Week 2 Application Assignment

Quiz, 13 questions