Kellogg School of Management
Northwestern University

EMP 96 Final Examination

Part I: True or false.

1. A 95%-confidence interval for the estimate of a population mean will always be at least as wide as a 90%-confidence interval based on the same sample data.
2. In a regression study, an observation with a positive residual is one for which the current model under-predicts the dependent variable.
3. In testing the null hypothesis that it takes *no more than* 1 hour to drive from Wilmette (near the main Northwestern campus) to Wieboldt Hall (in downtown Chicago, to teach an evening class), a professor finds that his sample data has a significance level of 4%. His sample mean must have been greater than 1 hour.

Part II: Unique Unicycles

In questions 1-9, work only with the originally-given variables.

1. Estimate the mean site rating across all visitors during the experiment, and give the margin of error in your estimate.
2. Predict the site rating that would be given by a visitor from a locality with a median family income of $60,000 (we'll call this an "attributed-income" of $60,000), who is shown the banner at a point-size of 24 and is shown the recreationally-oriented text. Give the margin of error in your prediction.
3. Estimate the mean rating given by visitors shown the banner at a point-size of 20, and shown the health-oriented text. Give the margin of error in your estimate.
4. During the experiment, who has been assigning higher ratings: Visitors with high attributed-incomes, or those with low attributed-incomes? Cite a number justifying your answer.
5. Estimate the proportion of site visitors who arrive via a Yahoo search, and give the margin of error in your estimate.
6. After regressing rating onto all of the original variables, we find that the two with the highest leverage are on rows 74 and 155 of the "Data" tab (they assigned ratings of 77 and 68, respectively). Why is their leverage so high? (Be as specific as you can.)

7. After looking at a model including all of the originally-given variables, you conclude that there's not much evidence that the location from which visitors arrive at UU's website (the "origin-site") plays a role in helping to predict the site rating given by the visitor. Cite a significance level of relevance to this conclusion.

   Before moving on, however, you wonder if attributed-income varies systematically with the origin-site.

8. Estimate the average attributed-income of site visitors who come in via a Google search, and give the margin of error in your estimate.

9. How much greater is the average attributed-income from-Yahoo visitors than is the average attributed-income from-Google visitors?

From this point on, we won't use the origin-site data again. Please omit it from all further analyses.

10. Regress site rating onto the remaining three original explanatory variables. What fraction of the variance in rating is potentially explained by variations in these three variables?

11. Having done this regression, where would you look to see if point-size actually might be related to ratings in a nonlinear manner?

12. What new variable would you introduce into your model to capture this suspected nonlinearity?

13. Create that new variable. Regressing site rating onto the three original variables and the newly-created one, what appears to be the optimal point-size for the slogan banner?

For the next series of questions (and for exam purposes), we'll leave out the variable created in (13).

14. Returning to the regression of rating onto the three original variables, for which is the evidence weakest that it actually belongs in your model? Cite a relevant significance level.

15. You decide to explore whether the choice of text should vary with the attributed-income of site visitors. What new variable would you add to your model to see if this is so?

16. Create that new variable, and add it to your model. How strong is the evidence that it indeed belongs in your model? Cite a relevant significance level.

17. What is the optimal text-targeting policy?

Finally, we pull the two previous sections together.

18. Bring both of your "created" variables (from (13) and (16)) into your model (so you now have 5 explanatory variables). Predict the rating that a visitor with an attributed-income of $50,000 would give to your site, if presented with a 20-point banner and the health-oriented text. Give the margin of error in your prediction.

Name : _____ **Answers**

EMP 96 Final Examination

Part I: Please circle your answers.

1. True False
2. True False
3. True False

Part II:

1.

_____ ±_____rating points

2.

_____% ±_____%

3.

_____ ±_____rating points

4.

_____ ±_____rating points

5.

Circle your answer: high low

Justifying number: _____

6.

Row 74: Because _____and _____

Row 155: Because _____and _____

7.

_____% (significance level)

8.

$_____± $ _____

9.

$ _____

10.

_____%

11.

Where: _____

12.

New variable: _____

13.

_____type-points (include 2 decimal places)

14.

Which variable?_____ _____%

15.

New variable: _____

16. Please circle your answer, and fill in a number:

very strong moderately strong not strong at all _____%

17. Circle and fill in:

Show the recreationally-oriented / health-oriented text to those with attributed-incomes above $ _____

18.

_____ ±_____rating points