



NORTHWESTERN  
UNIVERSITY

SCHOOL OF  
CONTINUING  
STUDIES

**Final Exam: Part I**

**PREDICT 401: Introduction to Statistical Analysis**

*Points possible:* 100

*Description:* The final exam will cover topics from sessions 1-9.

*Resources:* The exam is completely open book. You may use course textbooks, materials provided on Blackboard, a calculator, Excel and/or SPSS to answer the questions. For questions that require calculations, all calculations should be shown, not just the final answer. This will allow for partial credit for those answers that might be set up correctly but have calculation errors. For questions that require an answer using SPSS or Excel, graphs or output tables should be copied and pasted into your exam, with suitable discussion.

*Restrictions:* All answers are to be your work only. You are not to receive assistance from any other person.

*To complete the final exam:*

1. Answer all questions on the exam thoroughly. Create a Microsoft Word or PDF document, including the question number, the question, your typed answer, and SPSS/Excel output if relevant. You may use Word's equation editor to complete your answers.
2. Once you have completed your exam, return to the Assignment item where you downloaded the exam file, click View/Complete Assignment, and submit your document.

1. Your consulting group has been hired as political pollsters for the upcoming election. Your job is to estimate the proportion of voters who will vote for the Democratic ticket. You want to be 90% confident that your prediction is within  $\pm 0.04$  of the actual population proportion. (25 points total)
  - a. In order to achieve this, what sample size is needed? (5 points)
  - b. Suppose your fickle employers now want to have 95% confidence. What sample size will be needed? (5 points)
  - c. They have changed their mind once more and now, along with 95% confidence, they want to know the sample size needed with a margin of error of  $\pm 0.03$ . What sample size is needed? (5 points)
  - d. Based on your answers to these first three questions, what general conclusions can be drawn about how both the confidence level and margin of error affect the sample size needed, as well as how they interact with each other? (5 points)
  - e. Assuming that it will cost your organization roughly \$1 for each person you survey and you have a total of \$250,000 to spend, what sampling plan in terms of confidence level, margin of error, and sample size would you recommend to management, and why? Please be as specific and quantitative as possible and be sure to justify your statements. (5 points)

2. [Zagat's](#) publishes restaurant and hotel ratings for various establishments all over the world. The Excel file, Zagat Ratings, contains the ratings for food, décor, service, and price per person for a sampling of 50 restaurants located in New York City and 50 located in Long Island. Your group has been asked by Mr. Zagat himself to develop a regression model to predict the average price per person based on the sum of the ratings for food, décor and service. For this question, you must use SPSS or Excel for the relevant parts. Please use the attached data set. (45 points total)
- Set up a scatter diagram with the summated ratings on the horizontal axis and price per person on the vertical axis. (5 points)
  - Assuming a linear relationship, use the least-squares method to find the regression equation and the coefficients  $b_0$  (intercept) and  $b_1$  (slope) and interpret their meaning. (5 points)
  - Use the regression equation from part b to predict the average price of a restaurant with a summated rating of 50. (5 points)
  - The predicted value  $\hat{y}$  in part c) is an estimated average price per person. The quantity SSE (Sum of Squared Errors) divided by  $n-2$  gives you an estimate of the variance  $\sigma^2$  of the error term. Using this information, compute a standard error and a 95% confidence interval (NOT prediction interval) for the actual average price per person for a restaurant with a summated rating of 50. (5 points)
  - Remember that the computed value of the slope is a sample-based estimate of the unknown actual slope. Compute a 95% confidence interval estimate of the actual slope,  $\beta_1$ . (5 points)
  - Compute the coefficient of determination and interpret its meaning. (5 points)
  - Perform a simple residual analysis on your results and determine the adequacy of the fit of the model. You will need to plot the errors (difference between actual and predicted  $y$ -values) and determine to what extent they are independent of the  $x$ -values. (5 points)
  - At the 0.05 level of significance, is there evidence of a linear relationship between the price per person and the summated ratings? (5 points)
  - How useful do you think the summated rating is as a predictor of price per person? Explain. (5 points)

3. Your company is considering organizational changes based on adopting the use of self-managed work teams. To assess the attitudes of employees of the company toward this change, a sample of 400 employees is selected and asked whether they favor the introduction of self-managed work teams in the organization. Three responses were permitted: like, indifferent, or despise. The results are in the table below, classified by job. (30 points total)

<b>Attitudes toward self-managed work teams</b>				
<i>Job Type</i>	<i>Like</i>	<i>Indifferent</i>	<i>Despise</i>	<i>Total</i>
Hourly Worker	108	46	71	225
Supervisor	18	12	30	60
Middle Management	35	14	26	75
Upper Management	24	7	9	40
<i>Total</i>	185	79	136	400

- a. At the 0.05 level of significance, is there evidence of a relationship between attitude toward self-managed work teams and type of job? Use the chi-squared method to answer this question, and show your calculations. (10 points)

The survey went on to ask respondents about their attitudes toward instituting a policy whereby an employee could take one additional vacation day per month without pay. The results are in the table below.

<b>Attitudes toward vacation sans pay</b>				
<i>Job Type</i>	<i>Like</i>	<i>Indifferent</i>	<i>Despise</i>	<i>Total</i>
Hourly Worker	135	23	67	225
Supervisor	39	7	14	60
Middle Management	47	6	22	75
Upper Management	26	6	8	40
<i>Total</i>				400

- b. At the 0.05 level of significance, is there any evidence of a relationship between attitude toward vacation time without pay and type of job? Use the chi-squared method to answer this question, and show your calculations. (10 points)
- c. Explain how the answers to parts a and b may or may not be related. Also explain one other measure of association that you could use to analyze these data and answer the questions. It is not necessary to re-solve the questions in entirety from scratch with your alternate method. Just explain the procedure. (10 points)