**Data Analytics: Group Project**

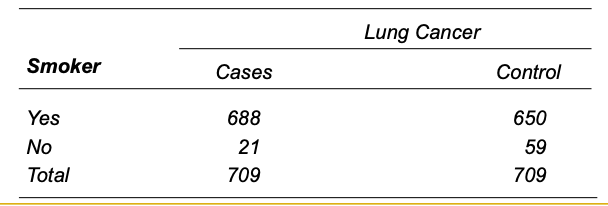
**Mini-Case 2**

[Total 30%, 10% for oral presentation using PowerPoint, 20% for written report including group combined effort]

Instruction:

1. Please give necessary details on how to get your numerical answers, keep 4 decimals in your answer.
2. Each group member is supposed to choose ONE question as follows. Each student is supposed to use PowerPoint to prepare his/her oral presentation file and use MS Word to prepare the answer file for the question. Upload the written report to iSpace on or before the last class. Each group consists of five members. Each group has 10 minutes (2 minutes for each member) for oral presentation.

***Cross classification of Smoking by Lung Cancer: Data is reported by Doll and Hill’s paper (1950).***

****

Question 1. Define

From the above survey data, find the estimated values for the above probabilities and summarize your conclusion on whether smokers have higher chance to get lung cancer.

Question 2. Define two categorical variables

Find the conditional probabilities:

and summarize your conclusion on whether smokers have higher chance to get lung cancer.

Question 3. Use the method of the Z-CI (confidence interval) for proportion to construct the 95%-CI for the four probabilities in Question 1.

Question 4. Use the method of the Z-CI (confidence interval) for proportion to construct the 95%-CI for the four probabilities in Question 2.

Question 5. Use alpha=5% and the two-sample Z-test for comparing two proportions to test the hypothesis

Group combined effort: Summarize your group’s conclusion on whether smokers have higher chance to get lung cancer based on the probability and statistical evidence in Questions 1-5.