Part 1: Death on the Titanic

The data set `titanic.dat` (titanic.xls) consists of survival status (dead or alive) on the Titanic by age, gender and cabin class, including crew. In this exercise, we actually have the entire population, that is, we have data on every passenger on the Titanic. This provides us with an opportunity to compare sampling estimates with the true value.

- Using the random number table or a statistical analysis package, draw a random sample of size 10, 20, and 50 from the full data set.
- Calculate the proportion that survived and construct a confidence interval around it for each sample size.
- Calculate the proportion that survived for the entire population. Report the difference between the estimates for each sample size and the true value.
- Calculate the proportion that survived by cabin class. Does survival status appear to be independent of cabin class?
- Create a new binary gender-age category with outcomes ‘women and children’ and ‘adult male’. Calculate the proportion that survived by ‘women and children’ and ‘adult males’. Does the ‘women and children first’ rule seem to have worked?
- Create a table or graph that appropriately displays the survival by cabin class for the population.
- Create a table or graph that appropriately displays survival by the new gender-age category.
- Report your findings in ‘journal article’ format.

Part 2: Airline Demand Effects of 9/11

The dataset `airq4.dat` (airq4.xls) contains data on average fare and average weekly passengers for the fourth quarters of years 2000 and 2001 for a population of 4177 markets. The file `airq4.txt` contains variable descriptions and file layout. Complete the following steps and write up a brief, but informative summary of results.

- Generate the percent difference (2001q4-2000q4) in total revenue for each market. Note that to obtain total revenue for each quarter, you multiply average fare by average weekly passengers by 13 weeks for each quarter.
- Obtain a histogram of percent difference in total revenues and comment on its shape.
- Obtain the mean and variance of this population.
- What size of sample would be needed to estimate this population mean percent within plus or minus 4% with 95% confidence?
- What is the power of testing \( H_0: \mu = 0 \) versus \( H_A: \mu < 0 \) for the true value of \( \mu \) for sample sizes of: \( n=25,64,100,400 \)? Use \( \alpha=0.05 \) significance level.
• Generate a random sample of n=50 markets and complete the following parts (use the middle 4 digits of UFID as a seed if prompted):
  • Obtain a 95% Confidence Interval for \( \mu \)
  • Test \( H_0: \mu = 0 \) versus \( H_A: \mu < 0 \) at \( \alpha=0.05 \) significance level
  • Write up your results in a “journal article” format.

Part 3: Comparison of 2 Yields

A firm wants to compare the yield of two types of animal feed. They wish to determine whether the true mean weight gain from the higher price type is higher than that for the lower price type. The response measured is weight gain (in kilograms) in a 3-month period on the diet. Complete the following steps and write up a brief, but informative summary of results:

• Based on a small pilot study, they estimate the standard deviations of weight gains to be 12 kg. How many animals should be assigned to each diet if they wish the power to detect a difference in means of 4 kg to be .90 (with \( \alpha=0.05 \))?
• How many animals should be assigned to each diet if they wish the margin of error to be less than or equal to 0.5 kg with 95% confidence?
• The dataset animfeed.dat (animfeed.xls) contains simulated data from an experiment with \( n_H=n_L=12 \) animals that were randomly assigned to the 2 diets.
• Obtain a 95% confidence interval for the difference in true mean weight gains (high price – low price). Note that the labels for diet are 1=Low, 2=High.
• Conduct a test of whether the population variances are equal.
• Based on your results of your previous part, use the appropriate t-test to determine whether the true mean weight gain is higher for the high price formulation (\( \alpha=0.05 \)).
• Use the Wilcoxon Rank-Sum Test to test whether the high price formulation is more effective in weight gain (\( \alpha=0.05 \)).
• Report your findings in ‘journal article’ format.