

Note: Solutions must be either typed or written neatly, and questions must be answered in order. Where SAS is used to answer a question, relevant SAS output must be copied as an image file and included with your answer to the question and not at the end of the assignment. Be sure to submit your assignment as a PDF and follow the instructions specified on the submission system. Name your file according to the convention <ASSIGNMENT NO>_<USERNAME>.pdf (*e.g.*, `Assignment_1_SmithSusa.pdf`).

1. Which of the following variables are categorical?
 - a. Water pressure (bars).
 - b. Course grade (A, B, C, D, E, F).
 - c. Level of approval of the Prime Minister's performance (1 = "Strongly disapprove", 2 = "Disapprove", 3 = "Neither approve nor disapprove", 4 = "Approve", 5 = "Stongly approve").
 - d. Hospital admissions (patients per day).
 - e. Yearly rainfall (centimeters).
 - f. Phone number.
2. Results from the 2013 New Zealand Census suggest that 20% of adults in New Zealand had a university degree or equivalent at the time of the census. Consider a random sample of 40 New Zealanders who participated in the 2013 census, and suppose that the number of these people who reported having a university degree or equivalent at the time of the 2013 census can be represented by a random variable following a binomial distribution.
 - a. Cleary explain what we are assuming about these 40 people in representing the number of them who reported having a university degree or equivalent at the time of the 2013 census by a binomial distribution? Provide an example of when this assumption would likely be violated. (Your answer must clearly refer to the situation described in the problem.)
 - b. What is the mean number of these 40 people that would be expected to have reported having a university degree or equivalent at the time of the 2013 census? What are the corresponding variance and standard deviation?
 - c. Using SAS, calculate the probability that exactly half of these 40 people reported having a university degree or equivalent at the time of the 2013 census.
 - d. What is the probability that fewer than 10 of these 40 people reported having a university degree or equivalent at the time of the 2013 census? Calculate this probability
 - exactly using SAS and
 - by hand using a normal approximation and the normal probability table.

3. Medical diagnostic tests are subject to one of two types of errors:
- **false positive:** A person who does not have the disease or condition returns a test result that suggests that they do have the disease or condition.
 - **false negative:** A person who has the disease or condition returns a test result that suggests that they do not have the disease or condition.

These two errors typically have quite different probabilities of occurring with the probability of a false positive most commonly being higher than the probability of a false negative because it is nearly always more catastrophic to miss those who have the disease or condition.

A recent report by the European CanCer Organisation (2017) into a non-invasive diagnostic test for stomach and esophageal cancers reported results on test results for 335 people across three different hospitals. The diagnosis test was administered to roughly equal numbers of people with and without stomach or esophageal cancer to assess the efficacy of the test. Results for the test are as shown in the table below.

Have stomach or esophageal cancer?	Tested positive for stomach or esophageal cancer?		<i>n</i>
	No	Yes	
No	140	32	172
Yes	32	131	163

Source: The European CanCer Organisation (29 January 2017). “Breath test could help detect stomach and esophageal cancers.” *ScienceDaily*.

- Suppose we wish to separately estimate the proportion of false positives and the proportion of false negatives and produce 95% confidence intervals for these proportions. Find the most conservative minimal sample sizes required for those who have stomach or esophageal cancer and those who do not have stomach or esophageal cancer to produce confidence intervals with an approximate margin of error of 0.06. (Note that you need only carry out one sample size calculation. The most conservative minimal sample size required will be the same sample size required to estimate each of the proportion of false positives and the proportion of false negatives to within the specified margin of error.)
- Using these data, produce both a standard and an Agresti-Coull 95% confidence interval for the proportion of false positives. Be sure to show all working.
- Test whether the proportion of false positives is significantly different from the proportion of false negatives. Carry out the test at the $\alpha = 0.05$ significance level, showing all working. Be sure to report the test statistic, p -value, and your conclusion based on the p -value. What does this result suggest about this particular diagnostic test?