Special Topics Comp Stat & Pro MAT5999 and Computational Stats & Prob. AIM 5002
Written Assignment 2 (25 points)
2/3/22

Solutions to be submitted on Canvas by the beginning of class on Wednesday, 2/9/22.

1. **(5 points)** The Associated Press (October 9, 2002) reported that in a survey of 4722 American youngsters aged 6 to 19, 15% were seriously overweight (a body mass index of at least 30; this index is a measure of weight relative to height). Calculate and interpret a confidence interval using a 99% confidence level for the proportion of all American youngsters who are seriously overweight.

2. **(5 points)** A sample of 14 joint specimens of a particular type gave a sample mean proportional limit stress of 8.48 MPa and a sample standard deviation of .79 MPa ("Characterization of Bearing Strength Factors in Pegged Timber Connections", J. of Structural Engr., 1997: 326-332). Calculate a 95% lower confidence bound for the true average proportional limit stress of all such joints. What, if any, assumptions did you make about the distribution of proportional limit stress?

3. **(5 points)** Solve Data Analysis exercise # 32a on this link. (the one where you need to construct 95% CI for the unknown speed of light given 100 measurements). For your information, the sample average is 852.4 and the sample standard deviation is 79.01055 (you can check these numbers if you click on the dataset, download it and use your favorite software to compute sample mean and sample standard deviation).

4. **(5 points)** Background: In the paper "Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy" by C. R. Charig; D. R. Webb; S. R. Payne; J. E. Wickham (Br Med J (Clin Res Ed). 292 (6524): 879-882.) two treatments for kidney stone were compared. Treatment A is a surgical procedure and Treatment B is percutaneous nephrolithotomy (which involves only a small puncture).

   It is found that treatment A was successfull in 81 of the 87 cases when applied for small stones. It was also successful in 192 of the 263 cases when applied for large stones. Treatment B was successful in 234 of the 270 cases when applied for small stones and in 55 of the 80 cases when applied for large stones.

   Compute the success rate of treatment A and B in three cases: when applied for small stones; when applied for large stones and when applied to any stone (that is, small and large stone data combined). Which procedure seems more successful? Why would you draw the wrong conclusion if you only looked at the combined data (that is, large and small stones data combined)?

   Remark: the phenomenon explored in this problem is called Simpson's paradox.

5. **(5 points)**

   (a) Let X be uniformly distributed on $[0, 1]$ and F be any continuous c.d.f. Show that the c.d.f. of $Y = F^{-1}(X)$ is F.

(b) The result of part (a) is useful in random number generation. For example, try

```
n <- 100
x <- qnorm(runif(n))
hist(x,prob=TRUE)
curve(dnorm(x), add=TRUE)
```

in RStudio (then choose different n and see how the plot changes). Explain how the output justifies numerically the results of part (a). No need to turn in your results from R, just explain what you saw and how it relates to part (a).

The following problems form the extra homework. They will not contribute to your final grade.

6. Given an random number generator *rand5()* that generates a uniform random integer in the range [1, 5], design an algorithm for *rand7()* that generates a uniform random integer in the range [1, 7]. You can only call the API *rand5()*, and you shouldn't call any other API.

   What is the expected number of times you call *rand5()*? How small can you make this expected number?