Problem Set 10

STAT-S 520

Due on April 3rd, 2023

Instructions:

- Submit your answers in Canvas as a single PDF file with answers in proper order.
- Include your R code, graphs, and relevant output.
  - Check that only the relevant output is included in your submission. Pages and pages of output that are not relevant can be penalized.
- You are allowed to collaborate with your classmates as long as you write your own solutions.

Questions:

```r
CPA = c(2.2041, 0.2744, 1.8050, 0.2822, 1.8062, 0.9600, 0.3175, 0.2953, 0.3704, 0.3828,
        7.8867, 5.6250, 4.4694, 4.8133, 0.6840, 0.6086, 1.5651, 0.5600, 2.2969)
x = round(log2(CPA),4)
x
```

```
##  [1]  1.1402 -1.8656  0.8520 -1.8252  0.8530 -0.0589 -1.6552 -1.7597 -1.4328
## [10] -1.3853  2.9794  2.4919  2.1601  2.2670 -0.5479 -0.7164  0.6463 -0.8365
## [19]  1.1997
```

1. The following exercise elaborates on the case study explicated in Section 10.4. You are welcome to read this case study (ISIR pp 257 - 260) but it's not necessary. The only relevant result is the vector of log_2(CPA) values given above as the vector `x`.

   a. Obtain the boxplot, normal probability plot, and kernel density plot of `x`. Does the sample seem to be drawn from a normal distribution?
   b. Is it possible that the difference with the normal is just do to sampling variation? To investigate whether or not this is the case, please do the following:
      i. Use`rnorm` to generate six samples from a normal distribution, each with `n = 19` observations.
      ii. Construct a normal probability plot for each simulated sample. Compare these plots to the normal probability plot of `x`
      iii. Compute the ratio of the sample interquartile range to the sample standard deviation for `x` and for each simulated sample.
      iv. Reviewing the available evidence, are you comfortable assuming that `x` was drawn from a normal distribution? Do six simulated samples provide enough information to answer the preceding question? Explain to receive full credit.
   c. Assuming that $X_1, \ldots, X_{19} \sim Normal(\mu, \sigma^2)$, perform a test of significance to determine whether the random sample `x` provides evidence to conclude that μ is different than zero. State the hypotheses, find the test statistic, p-value, and conclusion.
   d. Assume we learn that actually the true population mean for our problem is $\mu = 0.20$. Have we made a correct decision or have we committed a Type I or a Type II error? If an error, identify which one and explain why.

2. Assume that the hypotheses of a test are given by $H_0 : \mu \leq 14$ vs $H_1 : \mu > 14$ and we know $\sigma = 6$.

   a. Assume that we obtain a sample of size $n = 36$ with $\bar{x} = 15.2$. Perform a test with $\alpha = 0.1$, what is your conclusion?
   b. What is the sample mean that corresponds exactly to the boundary of the significance level? (i.e., the area under the PDF of $\bar{X}_n$ to the right of this sample mean has to be 0.1).
   c. Now we learn that actually $\mu = 15$. Have you committed a Type I error, Type II error, or made a correct decision.
   d. Use the value obtain in part b, alongside the true distribution (with $\mu = 15$) to obtain β, the probability of committing a type II error
   e. What is the power of the test?

3. Do the following

   a. Using the sample obtained in PS09 question 3c, obtain a 92% confidence interval for the average arrival delay for NY flights. Use the theory-based approach
   b. Repeat part a, using the simulation-based approach
   c. Using the sample obtained in PS09 question 3c, obtain a 96% confidence interval for the proportion of NY flights without arrival delays. Use the theory-based approach.

d. Repeat part c, using the simulation-based approach

4. ISI 9.5. Question 10 but use a 0.99 confidence level and L = 0.002.
5. ISI 11.4. Problem Set B, questions 3, 6, 7, and 8.

Reading assignments

- ISI Chapter 11, Section 11.1