1. Suppose $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{N}(\mu, \sigma^2)$ population, where $\mu \in \mathbb{R}$ is unknown and $\sigma^2 = 9$. The goal is to test

$$H_0 : \mu = 10$$
versus
$$H_a : \mu < 10.$$

Suppose we use $T = T(Y_1, Y_2, ..., Y_n) = \overline{Y}$ as a test statistic and the rejection region

$$\text{RR} = \{\overline{y} < k\},$$

where k is a constant.
(a) Show T is a sufficient statistic in this problem.
(b) Find the values of n and k which ensure $\alpha = 0.01$ and $\beta = 0.10$ when $\mu = 8$.
(c) For any value of $\mu \in \mathbb{R}$, show

$$P(\text{RR}) = P\left(\overline{Y} < k\right) = F_Z\left(\frac{k - \mu}{3/\sqrt{n}}\right),$$

where $F_Z(\cdot)$ is the $\mathcal{N}(0, 1)$ cdf. For the values of n and k you identified in part (b), graph $P(\text{RR})$ as a function of μ. (Later, we will call this the *power function*).

2. Suppose $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{N}(0, \sigma^2)$ population, where $\sigma^2 > 0$ is unknown (note the population mean here is $\mu = 0$). The goal is to test

$$H_0 : \sigma^2 = \sigma_0^2$$
versus
$$H_a : \sigma^2 \neq \sigma_0^2,$$

where σ_0^2 is known. Suppose we use $T = T(Y_1, Y_2, ..., Y_n) = \sum_{i=1}^n Y_i^2$ as a test statistic and the rejection region

$$\text{RR} = \{t < k_1 \ \text{ or } \ t > k_2\},$$

where $t = \sum_{i=1}^n y_i^2$ and $k_2 > k_1$.
(a) Show T is a sufficient statistic in this problem.
(b) Determine values of k_1 and k_2 to ensure a level α test.
(c) For any value of $\sigma^2 > 0$, show

$$P(\text{RR}) = F_{\chi^2_n}\left(\frac{k_1}{\sigma^2}\right) + 1 - F_{\chi^2_n}\left(\frac{k_2}{\sigma^2}\right),$$

where $F_{\chi^2_n}(\cdot)$ is the $\chi^2(n)$ cdf. Graph $P(\text{RR})$ as a function of σ^2 when $n = 10$, $\alpha = 0.05$, and $\sigma_0^2 = 1$.

3. Suppose $Y_1, Y_2, ..., Y_n$ is an iid sample from a population modeled by the pdf

$$f_Y(y|\theta) = \begin{cases} e^{-(y-\theta)}, & y > \theta \\ 0, & \text{otherwise.} \end{cases}$$

The population parameter θ satisfies $-\infty < \theta < \infty$ and is unknown. The goal is to test

$$H_0 : \theta = 0$$
$$\text{versus}$$
$$H_a : \theta > 0.$$

(a) When H_0 is true, which "named" distribution does $f_Y(y|\theta)$ represent? On the same graph, sketch what $f_Y(y|\theta)$ looks like when (i) H_0 is true and (ii) H_a is true.
(b) Suppose we use $T = T(Y_1, Y_2, ..., Y_n) = Y_{(1)}$ as a test statistic and the rejection region

$$\text{RR} = \left\{ y_{(1)} > k \right\},$$

where k is a constant. Find the value of k which ensures a level α test.

4. Suppose Y_1, Y_2 is an iid sample of size $n = 2$ from a $\mathcal{U}(\theta, \theta+1)$ population distribution where $\theta \in \mathbb{R}$ is unknown. The goal is to test

$$H_0 : \theta = 0$$
$$\text{versus}$$
$$H_a : \theta > 0.$$

Consider the two competing tests:

- Test 1: Reject H_0 when $Y_1 > 0.95$.
- Test 2: Reject H_0 when $Y_1 + Y_2 > c$.

(a) Determine the value of c so that the probability of Type I Error is the same for both tests.
(b) What if $n = 3$? How would you choose the corresponding critical value, say c^*, such that the rejection region $Y_1 + Y_2 + Y_3 > c^*$ has the same probability of Type I Error as the two tests above? *Hint:* I would think about using a Monte Carlo simulation strategy to approximate c^* here. The nice thing about this strategy is that it could be quickly generalized to accommodate any sample size n.

5. On August 17, 2021, Dr. Pastides reinstated the mask mandate for all faculty, staff, and students at the University of South Carolina. Let p denote the population proportion of UofSC students who support this decision. A random sample of $n = 100$ students is obtained and each student is asked if s/he supports the mandate. The goal is to test

$$H_0 : p = 0.5$$
$$\text{versus}$$
$$H_a : p > 0.5.$$

To make this problem concrete, let $Y_1, Y_2, ..., Y_{100}$ denote an iid sample of size $n = 100$ from a Bernoulli population distribution with probability p, which is unknown. Let $T = T(Y_1, Y_2, ..., Y_{100}) = \sum_{i=1}^{100} Y_i$ denote the sample sum.
(a) What is the sampling distribution of T when H_0 is true?
(b) We will use a rejection region of the form

$$\text{RR} = \left\{ t = \sum_{i=1}^{100} y_i \geq k \right\}.$$

Find the constant k which ensures $\alpha \approx 0.05$.
(c) Like the Poisson example in the notes, randomization would be needed to "hit" $\alpha = 0.05$ exactly. Describe how to perform a randomized test which will guarantee this.