1. In January 2010, the United States Supreme Court considered *Berghuis v. Smith*, the case of Diapolis Smith, a black man convicted in 1993 of second-degree murder by an all-white jury in Kent County, Michigan. Smith requested a new trial, arguing that blacks residents were under-represented in the jury pool. Census data confirmed black residents made up 7.28% of Kent County's age-eligible population at the time of the trial. By "age-eligible," one means those individuals who met the age requirements to sit in the jury pool. The jury pool consisted of 929 individuals, of whom 54 were black. Is this statistical evidence of under-representation of blacks in the jury pool? What assumptions are you making here?

2. Suppose $Y_1, Y_2, ..., Y_n$ is an iid sample from a Poisson distribution with mean $\theta > 0$, where θ is unknown. We would like to test

$$H_0 : \theta = \theta_0$$
$$\text{versus}$$
$$H_a : \theta > \theta_0.$$

In Example 8.3 (notes), we developed an "exact" hypothesis testing procedure using the sufficient statistic $T = \sum_{i=1}^{n} Y_i$ and its (discrete) sampling distribution. In this problem, we will develop a "large-sample" procedure and derive a sample size formula. This formula is applicable for the Poisson distribution only. We will use the test statistic

$$Z^* = \frac{\overline{Y} - \theta_0}{\sqrt{\frac{\theta_0}{n}}}.$$

(a) When H_0 is true, show that $Z^* \xrightarrow{d} \mathcal{N}(0,1)$, as $n \to \infty$. Conclude that

$$\text{RR} = \{z^* > z_\alpha\}$$

is an approximate level α rejection region. Note that z_α is the upper α quantile of the $\mathcal{N}(0,1)$ distribution. *Hint:* Showing the convergence result is a straightforward CLT argument, but explain all the details.

(b) Suppose we use the rejection region above to perform an approximate level α test of H_0 versus H_a. Derive a formula for the sample size n needed to maintain a Type II Error probability equal to β when $\theta = \theta_a > \theta_0$. *Hint:* The sample size n will satisfy

$$\beta \approx P_{\theta_a}(Z^* < z_\alpha).$$

First, convince yourself this is true by drawing a good picture. Then show

$$P_{\theta_a}(Z^* < z_\alpha) = P_{\theta_a}\left(Z < \frac{z_\alpha \sqrt{\frac{\theta_0}{n}} + \theta_0 - \theta_a}{\sqrt{\frac{\theta_a}{n}}} \right),$$

where $Z \sim \mathcal{N}(0,1)$. You should now be able to set up an equation which can be solved for n. The solution will depend on α, β, θ_0, and θ_a.

3. A high quality bullet proof glass will stop a bullet 90 percent of the time. A low quality glass will stop a bullet only 75 percent of the time.

You have received a large shipment of bullet proof glass, but you do not know if the shipment is high or low quality. To test which type you have, you decide to experiment by shooting bullets at n sheets of glass selected at random from the shipment. Your decision as to whether the glass is high or low quality will be based on the number (or the proportion) of the n sheets that stop the bullet.

To make your decision, you will enforce the following requirements:

- if the glass is of high quality, the probability of incorrectly concluding it is low quality is approximately 0.05.
- if the glass is of low quality, the probability of incorrectly concluding it is high quality is approximately 0.20.

What is the smallest sample size n that meets these requirements? If you can not find a numerical answer for n, you can outline an approach that would lead you to the correct answer; e.g., from solving an equation (or equations).

4. Suppose $Y_1, Y_2, ..., Y_n$ is an iid sample from a Pareto-type distribution with population level pdf

$$f_Y(y|\theta) = \begin{cases} \dfrac{\theta}{y^{\theta+1}}, & y > 1 \\ 0, & \text{otherwise.} \end{cases}$$

The population parameter θ satisfies $\theta > 0$ and is unknown.
(a) Construct a large-sample level α test for

$$H_0 : \theta = 5$$
$$\text{versus}$$
$$H_a : \theta > 5.$$

Hint: Apply the CLT to the sample mean $\overline{Y}$. State your (large-sample) test statistic and your rejection region.
(b) Does your large-sample procedure in part (a) involve a sufficient statistic in this problem? What might this suggest?

5. Airplanes approaching the runway for landing are required to stay within a "localizer region" (a certain distance left and right of the runway). When an airplane deviates from the localizer region, the FAA calls this an "exceedence." At the Schiphol airport in Amsterdam, two airlines (SAS and Lufthanza) were under investigation. In a one-week period, SAS had 8 out of 86 observed flights classified as exceedences. Lufthanza had 10 out of 142 observed flights classified as exceedences.

(a) Let p_1 and p_2 denote the population proportions of exceedances for the two airlines SAS and Lufthanza, respectively. Using the information above, perform a large-sample level $\alpha = 0.10$ test for

$$\begin{gathered} H_0 : p_1 - p_2 = 0 \\ \text{versus} \\ H_a : p_1 - p_2 \neq 0. \end{gathered}$$

(b) FAA officials want to plan a larger study that assumes a level $\alpha \approx 0.05$ analysis and equal numbers of airplanes sampled for both airlines (so that $n_1 = n_2 = n$). If the population proportions p_1 and p_2 differ by at least 0.05, officials would like to maintain a Type II Error probability of $\beta = 0.10$. Find the smallest sample size n (per airline) to accomplish these goals.