1. The United States Supreme Court recently considered *Berghuis v. Smith*, the case of Diapolis Smith, an African-American man convicted in 1993 of second-degree murder by an all-white jury in Kent County, Michigan. Smith requested a new trial, arguing that minorities were under-represented in the jury pool. Census data confirmed African-Americans made up 7.28% of Kent County's age-eligible population at the time of the trial (by "age-eligible," I mean those individuals who met the age requirements to sit in the jury pool). The jury pool consisted 929 individuals, of whom 54 were African-American. None of the 54 potential African-American jurors were ultimately selected. Perform a hypothesis test to answer the following question:

> Was there under-representation of African-Americans in the jury pool in the case of *Berghuis v. Smith*?

Choose an α level you feel is appropriate.

2. Coach Spurrier is concerned that his team is being penalized excessively, so he has asked you investigate this issue statistically. Let Y_i denote the number of penalties USC receives in quarter i during its next game, for $i = 1, 2, 3, 4$; that is, there are $n = 4$ quarters in the game. Coach Spurrier is very proficient at probability and mathematical statistics, and he is convinced that Y_1, Y_2, Y_3, Y_4 are independent and identically distributed Poisson random variables with mean $\theta > 0$. Under these model assumptions, Coach Spurrier asks you to test

$$H_0 : \theta = 1$$
versus
$$H_a : \theta > 1$$

using the rejection region

$$\text{RR} = \{t : t \geq 6\},$$

where $T = Y_1 + Y_2 + Y_3 + Y_4$ is the sample sum.

(a) Why do you think Coach Spurrier asked you to use the statistic T to perform this test?

(b) For this RR, compute the probability of Type I Error.

(c) Write out an expression, in terms of a Poisson mass function, that equals the probability of Type II Error when $\theta = 2$. You don't have to evaluate your expression.

3. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from an exponential distribution with mean $\theta > 0$. Consider testing

$$H_0 : \theta = 1$$
versus
$$H_a : \theta > 1$$

using the statistic $T = Y_1 + Y_2 + \cdots + Y_n$ and a rejection region $\text{RR} = \{t : t > k\}$, where k is chosen so that the test is of level α.
(a) Provide an expression for k.
(b) Show that the power function $K(\theta)$ is given by

$$K(\theta) = 1 - F_U\left(\frac{2k}{\theta}\right),$$

where $F_U(\cdot)$ is the cumulative distribution function (cdf) of $U \sim \chi^2(2n)$.

4. Suppose that Y follows a Pareto distribution with probability density function (pdf)

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

Consider testing

$$\begin{array}{c} H_0 : \theta = 2 \\ \text{versus} \\ H_a : \theta > 2 \end{array}$$

based on the single observation Y. Find the uniformly most powerful (UMP) level $\alpha = 0.10$ rejection region.

5. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from an exponential distribution with mean $\theta > 0$. Consider testing

$$\begin{array}{c} H_0 : \theta = 1 \\ \text{versus} \\ H_a : \theta > 1 \end{array}$$

using the sufficient statistic $T = Y_1 + Y_2 + \cdots + Y_n$ and a rejection region of the form $\text{RR} = \{t : t > k\}$, where k is a constant. We would like

- to use a level $\alpha = 0.01$ test
- our test to have a Type II Error probability be $\beta = 0.10$ when $\theta = 2$.

Describe how you would determine the minimum sample size n necessary to meet these requirements. Provide as much detail as possible.

6. Suppose that $Y \sim \mathcal{U}(-\theta, \theta)$, where $\theta > 0$, and that we would like to test $H_0 : \theta = 1$ versus $H_a : \theta > 1$, based on the value of Y using the rejection region $\text{RR} = \{y : |y| > k\}$.

(a) Find the value of k that provides a level $\alpha = 0.05$ test.
(b) Using your rejection region from part (a), graph the power function $K(\theta)$. What is $K(1.5)$?

7. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from

$$f_Y(y;\theta) = \begin{cases} \theta^2 y e^{-\theta y}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Show that the uniformly most powerful (UMP) level α test of

$$H_0 : \theta = 1$$
$$\text{versus}$$
$$H_a : \theta < 1$$

has a rejection region of the form $\text{RR} = \{\boldsymbol{y} : \sum_{i=1}^n y_i > c\}$. Find an expression for c.

8. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of Bernoulli observations with mean p, $0 < p < 1$, p fixed. In case you have forgotten, the Bernoulli(p) probability mass function is given by

$$f_Y(y;p) = \begin{cases} p^y(1-p)^{1-y}, & y = 0, 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Suppose that the goal is to test,

$$H_0 : p = p_0$$
$$\text{versus}$$
$$H_a : p > p_0.$$

Show that the uniformly most powerful (UMP) level α rejection region has the form

$$\text{RR} = \left\{\boldsymbol{y} : \sum_{i=1}^n y_i \geq k^*\right\},$$

where k^* is chosen so that the test is of level α. Write out an equation that describes how k^* would be chosen. Make sure to explain why is this rejection region is UMP. Do not use any type of normal approximation on this part.
(b) We have shown in class that the test in part (a) can be performed by using a normal approximation to the sampling distribution of

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^n Y_i.$$

In particular, we showed that an approximate level α rejection region has the form

$$\text{RR} = \{z : z > z_\alpha\},$$

where

$$Z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

and z_α is the upper α quantile of the standard normal distribution. Using this approximate level α rejection region, derive $K(p)$, the power function for the test. You should be able to express the power function in terms of the $\mathcal{N}(0,1)$ cumulative distribution function.

(c) The rejection region in part (a) is based on the exact distribution of the sufficient statistic

$$U = \sum_{i=1}^{n} Y_i.$$

The rejection in part (b) is based on the large-sample distribution of $\widehat{p} = U/n$. In a few sentences, explain (to an investigator) what the difference is between exact statistical procedures and procedures based on asymptotic theory.

9. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid $\mathcal{N}(0, \sigma^2)$ sample, where $\sigma^2 > 0$ is unknown. We are interested in testing

$$H_0 : \sigma^2 = 1$$
versus
$$H_a : \sigma^2 > 1.$$

(a) Show that $\text{RR} = \{t : t > \chi^2_{n,\alpha}\}$ is a level α rejection region, where the sufficient statistic

$$T = \sum_{i=1}^{n} Y_i^2.$$

(b) Suppose that we would like to maintain

- a Type I Error probability $\alpha = 0.05$.
- a Type II Error probability $\beta = 0.10$ when $\sigma^2 = 1.2$.

Describe an approach to find the smallest sample size n that meets these requirements. You don't have to come up with a numerical solution; just describe your approach as much as possible.

10. Suppose $Y_1, Y_2, ..., Y_n$ is an iid $\mathcal{N}(\mu, \sigma^2)$ sample, where $\sigma^2 = 9$, and that we would like to test $H_0 : \mu = 10$ versus $H_a : \mu > 10$ using a rejection region $\text{RR} = \{\boldsymbol{y} : \overline{y} > k\}$.

(a) Find the values of n and k that provide a Type I Error probability of $\alpha = 0.02$ and a Type II Error probability, when $\mu = 12$, of $\beta = 0.1$.

(b) With your values of n and k from part (a), graph the power function $K(\mu)$.

11. Is there an exceptionally high percentage of the executives of large corporations that are right-handed? It is assumed that 85 percent of the general population is right-handed. In a random sample of 60 chief executive officers, it was found that 57 were right-handed. Is this a statistically significant result at the $\alpha = 0.05$ level? Answer this question by performing an exact test (based on the binomial distribution) and a large-sample test based on the normal approximation. Do the two tests provide different conclusions?

12. The distribution of starting salaries (in \$1000s) of USC undergraduates majoring in statistics is well modeled by a Pareto distribution. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of n recent graduates, modeled as arising from a Pareto pdf of the form

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

where $\theta > 1$. The parameter ν represents the minimum starting salary; we will assume $\nu = 35$ (known). Find the form of the most powerful level α rejection region to test $H_0 : \theta = 2$ versus $H_a : \theta = 3$. Explain, in as much detail as possible, how to find the critical value for the test; you don't have to express the critical value in terms of a well known probability distribution. Just tell me how one would find it.

13. The Environmental Testing Agency (EPA) often collects data on LC50 measurements (the concentration that kills 50 percent of test animals) for certain chemicals likely to be found in freshwater rivers and lakes. For a certain species of fish, a scientist will take n LC50 measurements (in parts per million) for a certain chemical yielding data $Y_1, Y_2, ..., Y_n$. The scientist wants to determine whether or not the mean LC50 reading exceeds 7.5 parts per million.
(a) Set up an appropriate null and alternative hypothesis for this problem. Define any notation that you use.
(b) Give the form of the test statistic suitable for your hypothesis test in (a). What is the distribution of your test statistic when H_0 is true? State all necessary assumptions. Suppose $n = 12$ and the LC50 measurements are as follows:

```
16  5  21  19  10  5  8  2  7  2  4  9
```

(c) Calculate the numerical value of the test statistic with the above data. Simple calculations show that $\sum_{i=1}^{12} y_i = 108$ and $\sum_{i=1}^{12}(y_i - \overline{y})^2 = 453.38$.
(d) Find the probability value for this test that uses the above data. Would you reject your H_0 at the five percent significance level? State your conclusion as a meaningful sentence to the scientist.

14. Suppose that $X_1, X_2, ..., X_{n_1}$ is an iid sample from a $\mathcal{N}(0, \sigma_X^2)$ distribution. Also, suppose that $Y_1, Y_2, ..., Y_{n_2}$ is an iid sample from a $\mathcal{N}(0, \sigma_Y^2)$ distribution. Finally, assume that the samples are independent, and define $\lambda = \sigma_Y^2/\sigma_X^2$. Derive the form of a level α hypothesis test for $H_0 : \lambda = 1$ versus $H_a : \lambda \neq 1$. Be explicit.

15. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from a Weibull distribution with pdf

$$f_Y(y;\beta) = \begin{cases} \frac{\alpha_0}{\beta} y^{\alpha_0 - 1} e^{-y^{\alpha_0}/\beta}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

where α_0 is known and $\beta > 0$ is unknown.
(a) Prove that $Y_1^{\alpha_0}$ has an exponential distribution with mean β.
(b) Derive the form of a level α hypothesis test for $H_0 : \beta = 1$ versus $H_a : \beta \neq 1$. Use the test statistic

$$T = \sum_{i=1}^{n} Y_i^{\alpha_0}.$$