Assignment #4: Chapter 10 Questions 38, 54, 70, 96, 98

Question 10.38:   The Rockwell hardness index for steel is determined by pressing a diamond point into the steel and measuring the depth of penetration. For 50 specimens of an alloy of steel, the Rockwell hardness index averaged 62 with standard deviation 8. The steel is sufficiently hard to meet usage requirements if the mean Rockwell hardness measure does not drop below 60. Using the rejection region at the 1% significance level, find β for the specific alternative $\mu_a = 60$.

At the $\alpha = 0.01$ significance level, the rejection region is given by $z = \texttt{qnorm(0.01)} = -2.326348$. To find β, solve for $\bar{x}$ first.

$$z \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$\bar{x} \geq \mu + z\frac{\sqrt{n}}{\sigma} = 62 - \left(2.326348 \cdot \frac{\sqrt{50}}{8}\right)$$

$$\geq 59.94377$$

Now, solve for β, or the probability that $\bar{x}$ is greater than the above result.

$$\beta = \mathbb{P}(\bar{x} \geq 59.94377)$$

$$= \mathbb{P}\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq \frac{59.94377 - 60}{8/\sqrt{50}}\right)$$

$$= \mathbb{P}(z \geq -0.04969)$$

$$= \texttt{1 - pnorm(1.718075)}$$

$$= 0.5198153$$

Question 10.54:   Do you believe that an exceptionally high percentage of the executives of large corporations are right-handed? Although 85% of the general public is right-handed, a survey of 300 chief executive officers of large corporations found that 96% were right-handed.

a. Is this difference in percentages statistically significant? Test using $\alpha = 0.01$.
   At the $\alpha = 0.01$ significance level, $z = 2.326348$. Calculate z using the given values and determine whether $z > 2.326348$.

$$z = \frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.96 - 0.85}{\sqrt{\frac{0.85 \cdot 0.15}{300}}} = 5.335783 \geq 2.326348$$

   This shows that the difference in percentages is significant.

b. Find the p-value for the test and explain what it means.

$$p\text{-value} = \mathbb{P}(z > 5.335783) = 1 - \texttt{pnorm(5.335783)} \approx 0$$

   This means that if the null hypothesis, $p = 0.85$, the probability of finding a difference as large or greater is basically 0. This means that the null hypothesis is impossible to occur and can be rejected.

Question 10.70:   A study was conducted by the Florida Game and Fish Commission to assess the amounts of chemical residues found in the brain tissue of brown pelicans. In a test for DDT, random samples of $n_1 = 10$ juveniles and $n_2 = 13$ nestlings produced the results shown in he accompanying table (measurements in parts per mission, ppm).

Juveniles	Nestlings
$n_1 = 10$	$n_2 = 13$
$\bar{y}_1 = 0.041$	$\bar{y}_2 = 0.026$
$s_1 = 0.017$	$s_2 = 0.006$

a. Test the hypothesis that mean amounts of DDT found in juveniles and nestlings do not differ versus the alternative, that the juveniles have a larger mean. Use $\alpha = 0.05$. (This test has important implications regarding the accumulation of DDT over time.) The hypotheses are:

$$H_0 : \mu_{\text{juveniles}} - \mu_{\text{nestlings}} = 0$$
$$H_A : \mu_{\text{juveniles}} - \mu_{\text{nestlings}} > 0$$

To calculate the test statistic, first calculate the pooled standard deviation s_p.

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(10 - 1) \cdot 0.017^2 + (13 - 1) \cdot 0.006^2}{10 + 13 - 2}} = 0.01201$$

Then the test statistic is

$$T = \frac{\bar{y}_1 - \bar{y}_2 - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.041 - 0.026 - 0}{0.01201\sqrt{\frac{1}{10} + \frac{1}{13}}} = 2.96737$$

Now

$$t = \texttt{qt(0.05, df = 10 + 13 - 2)} = 1.720743$$

Since $2.96737 > t_\alpha$, the null hypothesis is rejected and the mean amounts of DDT found in juveniles and nestlings do differ.

b. Is there evidence that the mean for juveniles exceeds that for nestlings by more than 0.01 ppm?
The hypotheses are:

$$H_0 : \mu_{\text{juveniles}} - \mu_{\text{nestlings}} = 0.01$$
$$H_A : \mu_{\text{juveniles}} - \mu_{\text{nestlings}} > 0.01$$

Then the test statistic is

$$T = \frac{\bar{y}_1 - \bar{y}_2 - 0.01}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.041 - 0.026 - 0.01}{0.01201\sqrt{\frac{1}{10} + \frac{1}{13}}} = 0.98912$$

With this, the p-value is

$$p\text{-value} = \mathbb{P}(T > 0.98912) = \texttt{1 - pt(0.98912, 10 + 13 - 2)} = 0.1669612$$

At the significance level of $\alpha = 0.05$, the null hypothesis cannot be rejected and it can be concluded that there is no significant evidence that $\mu_{\text{juveniles}} - \mu_{\text{nestlings}} = 0.01$ is wrong.

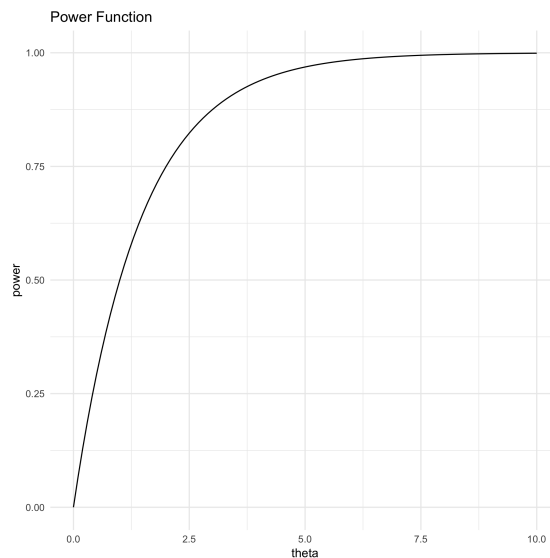**Question 10.96:** Suppose Y is a random sample of size 1 from a population with density function

$$f(y \mid \theta) = \begin{cases} \theta y^{\theta-1} & \text{if } 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

where $\theta > 0$.

a. Sketch the power function of the test with rejection region: $Y > 0.5$.
   To find the power function of the test, compute $\int_{0.5}^{1} \theta y^{\theta-1} \, dy$.

$$\int_{0.5}^{1} \theta y^{\theta-1} \, dy = \frac{\theta y^{\theta-1+1}}{\theta - 1 + 1} \bigg|_{y=0.5}^{y=1} = y^\theta \bigg|_{y=0.5}^{y=1} = 1 - 0.5^\theta$$

Then the power function looks like:

Power Function

Code to generate plot in `R`:

```
library(tidyverse)
p = ggplot(data = data.frame(x=0), mapping = aes(x=x))
fun.1 = function(x) 1 - 0.5^x
plt = p + stat_function(fun = fun.1) + xlim(0, 10) +
labs(x = "theta", y = "power", title = "Power Function") +
theme_minimal()
ggsave("powerfunc.png", plt, height = 7, width = 7, units = "in")
```

b. Based on the single observation Y, find a uniformly most powerful test of size α for testing $H_0 : \theta = 1$ versus $H_A : \theta > 1$.
   The likelihood ratio is

$$\frac{L(\theta_0)}{L(\theta_A)} = \frac{1}{\theta_A y^{\theta_A - 1}} < k$$

Now solve for y.

$$1 \leq k\theta_A y^{\theta_A - 1}$$

$$\frac{1}{k\theta_A} \leq y^{\theta_A - 1}$$

$$\left(\frac{1}{k\theta_A}\right)^{\frac{1}{\theta_A - 1}} \leq y$$

Given a size α, the equation for y becomes

$$y \geq \left(\frac{1}{\theta_A k}\right)^{(\theta_A - 1)^{-1}} = c$$

where c is a constant. This is the rejection region and since $\theta_A > 1$ is the only criterion, it is the uniformly most powerful test for testing H_0 against H_A.

Question 10.98: Let $Y_1, \ldots, Y_n$ be a random sample from the probability density function given by

$$f(y \mid \theta) = \begin{cases} \left(\frac{1}{\theta}\right) m y^{m-1} e^{-y^m/\theta} & \text{if } y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

with m denoting a known constant.

a. Find the uniformly most powerful test for testing $H_0 : \theta = \theta_0$ against $H_a : \theta > \theta_0$. First,

$$f(y_1, \ldots, y_n \mid \theta) = \begin{cases} \left(\frac{1}{\theta}\right)^n m^n (y_1 \ldots y_n)^{m-1} e^{-\left(\frac{\sum_i^n y_i^m}{\theta}\right)^n} & \text{if } \min(y_1, \ldots, y_n) > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The likelihood ratio is

$$\frac{L_0}{L_A} = \frac{\theta_0^{-n} m^n (y_1 \ldots y_n)^{m-1} e^{-\frac{\sum_i^n y_i^m}{\theta_0}}}{\theta_A^{-n} m^n (y_1 \ldots y_n)^{m-1} e^{-\frac{\sum_i^n y_i^m}{\theta_A}}} < k$$

Solve for in terms of $\sum_i^n y_i$.

$$k > \frac{\theta_0^{-n} m^n (y_1 \ldots y_n)^{m-1} e^{-\frac{\sum_i^n y_i^m}{\theta_0}}}{\theta_A^{-n} m^n (y_1 \ldots y_n)^{m-1} e^{\frac{\sum_i^n y_i^m}{\theta_A}}}$$

$$= \left(\frac{\theta_A}{\theta_0}\right)^n e^{-(\theta_0^{-1} - \theta_A^{-1})\sum_i^n y_i^m}$$

$$\ln k > n \ln\left(\frac{\theta_A}{\theta_0}\right) - (\theta_0^{-1} - \theta_A^{-1}) \sum_i^n y_i^m$$

$$\sum_i^n y_i^m > \frac{\ln k - n \ln\left(\frac{\theta_A}{\theta_0}\right)}{\theta_0^{-1} - \theta_A^{-1}} = c$$

Given a size α, the rejection region is

$$RR = \sum_i^n Y_i^m > c$$

where c is a constant. Furthermore, since the distribution of Y^m is exponential,

$$\frac{L(\theta_A)}{L(\theta_0)} = \frac{2RR}{\theta_0} = \frac{2\sum_i^n Y_i^m}{\theta_0} > \frac{2c}{\theta_0}$$

the distribution under H_0 is distributed as $\chi^2(2n)$. This does not depend on $\theta_A > \theta_0$ and so this test is the uniformly most powerful test for testing the hypotheses.

b. If the test in part (a) is to have $\theta_0 = 100$, $\alpha = 0.05$, and $\beta = 0.05$, when $\theta_a = 400$, find the appropriate sample size and critical region.
   If H_0 is true, then $\frac{2T}{100} = \frac{T}{50} \sim \chi^2(2n)$. If H_A is true, then $\frac{2T}{400} = \frac{T}{200} \sim \chi^2(2n)$. Therefore the following is needed:

$$\beta = \mathbb{P}\left(\frac{T}{50} \leq \chi^2_{0.05} \mid \theta = 400\right) = \mathbb{P}\left(\frac{T}{200} \leq \frac{\chi^2_{0.05}}{4} \mid \theta = 400\right) = \mathbb{P}\left(\chi^2 \leq \frac{1}{4}\chi^2_{0.05}\right) = 0.05$$

At the size of $\alpha = 0.05$, it must be the case that

$$\frac{1}{4}\chi^2_{0.05} = \chi^2_{0.95}$$

Look for a value of `df` where

$$\frac{\text{qchisq(0.95, df)}}{\text{qchisq(0.05, df)}} \approx 4$$

The best degrees of freedom that equate this is 12 and so, $2n = 12 \rightarrow n = 6$. The appropriate sample size is 6 and the critical region is given by

$$\beta = \mathbb{P}\left(\chi^2 \leq \frac{1}{4}\chi^2_{0.05}\right) = 0.05$$