STAT 323 Assignment 4

Instructor: Claudia Marie Mahler
Name: Yifeng Pan
UCID: 30063828

Summer 2019

1 Let X_1 and X_2 constitute a random sample from a population which is normally distributed with $\sigma^2 = 1$. If the null hypothesis $\mu = \mu_0$ is to be rejected in favor of the alternative hypothesis $\mu = \mu_a$, when $\overline{x} = \mu_0 + 1$, what is the size of the critical region (rejection region)? Assume $\mu_a > \mu_0$.

$\sigma = 1, \overline{x} = \mu_0 + 1, P(R\ H_0|H_0) = \alpha$
$\frac{\overline{x}-\mu_0}{\sigma}\sqrt{n} = \frac{\mu_0+1-\mu_0}{\sqrt{1}}\sqrt{2} = \sqrt{2}$
$1 - pnorm(\sqrt{2}) \approx 7.865\%.$

2 Let X_1 represent a random sample of size 1 from a population having the following probability density function:

$$f(x) = \begin{cases} \theta x^{(\theta-1)} & 0 < x < 1 \\ 0 & \textbf{elsewhere} \end{cases}$$

If the critical region $x_1 \geq 0.75$ is used to test the null hypothesis $\theta = 1$ against the alternative hypothesis $\theta = 2$, what is the power of the test at $\theta = 2$?

$P(RH_0|H_a) = p$
$P(x_1 > 0.75|\theta = 2) = \int_{0.75}^{1} 2xdx = 43.75\%$

3 A Calgary physician wants to determine whether a weight-reducing drug has a different effect on adults over 40 years of age than on adults that are no more than 40 years of age. For individuals who are no more than 40 years of age, it is known that the mean weight loss on this drug is $\mu = 11.3$ pounds. Twelve people over the age of 40 are given the drug; the mean weight loss is 8.9 pounds with a sample standard deviation of 4.1 pounds. Does the data suggest that the drug has a different effect on adults over 40 years of age compared to adults that are no more than 40 years of age? Conduct a hypothesis test using $\alpha = 0.05$.

$H_0 : \mu_{\leq 40} = \mu_{>40}$
$H_a : \mu_{\leq 40} \neq \mu_{>40}$
$t = \frac{8.9-11.3}{4.1}\sqrt{12} \approx -2.028$
p-value $= pt(t, 11) * 2 \approx 6.75\% > 5\%$
Therefore the H_0 is accepted.

4 It has been suggested that the price of condominiums in Calgary have increased in the past six months. In order to test this claim, a sampling design was employed where the selling price of 10 randomly selected condos were chosen from both January and July of this year. The selling prices are in terms of $\$1000$s and the data is in the Condo.R data file. Assuming that the selling price of condos are normally distributed, does the data support the claim that the mean selling price of a condo in July is greater than the selling price of a condo in January of this year? Use $\alpha = 0.05$.

$H_0 : \mu_{Ju} \leq \mu_{Ja}$
$H_a : \mu_{Ju} > \mu_{Ja}$
R code:

```
July = c(153.3, 155.9, 176.2, 189.9, 200.0, 214.9,
         229.9, 231.5, 257.9, 299.9)
January = c(151.1, 154.2, 169.9, 169.9, 185.9, 199.5,
            229.9, 232.9, 279.9, 289.9)

#Variance are equal
var4 = var.test(July, January, alternative = "greater")$conf.int

pv4a = t.test(July, January, var.equal = T,
              alternative = "greater", conf.level = 0.95)$p.value

pv4b = 1 - pt(mean(July - January)
              / sqrt( (var(July) + var(January)) / 2 )
              / sqrt(2 / 10), 18)

cat("p-value for p4, first method: ", pv4a, '\n')
cat("p-value for p4, second method: ", pv4b, '\n')
```

Output:

```
p-value for p4, first method:  0.4159381
p-value for p4, second method:  0.4159381
```

p-value $\approx 41.59\% > 5\%$
Therefore the H_0 is accepted.

5 A precision instrument is stated to have a measurement variation of no more than 0.49 units. A sample of four instrument readings on the same object yielded the measurements $351.4, 351, 351.9,$ and 350.3. Does this data suggest that the measurement variation of the instrument is at most the stated 0.49 units? Use $\alpha = 0.07$.

$H_0 : \sigma^2 > 0.49$
$H_a : \sigma^2 \leq 0.49$
R code:

```r
library(EnvStats)
p5 = c(351.4, 351, 351.9, 350.3)
pv5 = varTest(p5, alternative = "less", conf.level = 0.93)$p.value
cat("p-value for p5: ", pv5, '\n')
```

Output:

```
p-value for p5:  0.2874183
```

p-value $\approx 28.74\% > 7\%$
Therefore H_0 is accepted.

6 A major court case on the health effects of drinking contaminated water took place in the town of Picture Butte, Alberta. A town well in Picture Butte was contaminated with fecal bacteria due to run-off from a local cattle feedlot. During the period that residents drank water from this well, there were 16 birth defects among 414 births. In years when the contaminated well in question was not used and water was supplied from clean wells, there were two birth defects among 228 births. The plaintiffs suing thefeedlot responsible for the water contamination claimed that these data show that the rate of birth defects was higher when the contaminated well was in use. Conduct a hypothesis test with $\alpha = 0.01$ to determine if the rate of birth defects was significantly higher when the contaminated well was in use.

$H_0 : p_c \leq p$
$H_a : p_c > p$
$z = \frac{\hat{p}-p}{\sqrt{p(1-p)}}\sqrt{n} = \frac{16/414-2/228}{\sqrt{2/228(1-2/228)}}\sqrt{414} \approx 6.519$
p-value $= 1 - pnorm(z) \approx 3.539 \times 10^{-11} << 0.01$
Therefore H_0 is rejected.

7 Let $X_1, X_2, \ldots, X_n$ represent a random sample from a normal distribution where the value of the mean μ is unknown and the variance is $\sigma^2 = 1$. Derive the uniformly most powerful test criterion with $\alpha = 0.05$ used to test the hypotheses

$$H_0 : \mu = 0$$

$$H_a : \mu = 1$$

$$\begin{aligned}
k > \frac{L(\theta_0)}{L(\theta_a)} &= \prod_{i=1}^{n} \frac{\frac{1}{1\sqrt{2\pi}}\exp(-(\frac{(x_i-0)^2}{2}))}{\frac{1}{1\sqrt{2\pi}}\exp(-(\frac{(x_i-1)^2}{2}))} = \prod_{i=1}^{n} \frac{\exp(-(\frac{(x_i-0)^2}{2}))}{\exp(-(\frac{(x_i-1)^2}{2}))} = \prod_{i=1}^{n} \exp(-(\frac{(x_i-0)^2}{2}) + (\frac{(x_i-1)^2}{2})) \\
&= \prod_{i=1}^{n} \exp(\frac{1}{2}((x_i-1)^2 - x_i^2)) = \prod_{i=1}^{n} \exp(\frac{1}{2}(1-2x_i))
\end{aligned}$$

$$\begin{aligned}
\ln(k) > \ln(\frac{L(\theta_0)}{L(\theta_a)}) &= \ln(\prod_{i=1}^{n} \exp(\frac{1}{2}(1-2x_i))) = \sum_{i=1}^{n} \frac{1}{2}(1-2x_i) = \sum_{i=1}^{n} \frac{1}{2} - x_i = \frac{n}{2} - \sum_{i=1}^{n} x_i = \frac{n}{2} - n\overline{x} \\
\ln(k) + n\overline{x} &> n/2 \\
\overline{x} &> \frac{n/2 - \ln(k)}{n} = \frac{1}{2} - \frac{\ln(k)}{n} \\
\frac{\overline{x} - \mu}{\sigma}\sqrt{n} &> \frac{(\frac{1}{2} - \frac{\ln(k)}{n}) - 0)}{\sqrt{1}}\sqrt{n} = \frac{\sqrt{n}}{2} - \frac{\ln(k)}{\sqrt{n}}
\end{aligned}$$

(Note: The following approach is needlessly convoluted, but it avoids having to introduce a new varible.)

$$\begin{aligned}
1 - pnorm(\frac{\sqrt{n}}{2} - \frac{\ln(k)}{\sqrt{n}}) &= 0.05 \\
pnorm(\frac{\sqrt{n}}{2} - \frac{\ln(k)}{\sqrt{n}}) &= 0.95 \\
\frac{\sqrt{n}}{2} - \frac{\ln(k)}{\sqrt{n}} &= qnorm(0.95)
\end{aligned}$$

Solve for k and we get $k = \exp(\frac{n}{2} - qnorm(0.95)\sqrt{n})$.
Therefore the test criterion for μ is:

$$\begin{aligned}
\overline{x} > \frac{1}{2} - \frac{\ln(k)}{n} &= \frac{1}{2} - \frac{\ln(\exp(\frac{n}{2} - qnorm(0.95)\sqrt{n}))}{n} \\
&= \frac{1}{2} - \frac{\frac{n}{2} - qnorm(0.95)\sqrt{n}}{n} \\
&= \frac{1}{2} - (\frac{1}{2} - qnorm(0.95)/\sqrt{n})) \\
&= \frac{qnorm(0.95)}{\sqrt{n}} \approx \frac{1.645}{\sqrt{n}}
\end{aligned}$$