NORTHWESTERN UNIVERSITY | SCHOOL OF PROFESSIONAL STUDIES

R Lesson 9 - Solutions
MSPA 401 – Introduction to Statistical Analysis

1) Assume a random sample of size 100 is drawn from a normal distribution for which the mean and variance are unknown. Assume the sample mean is 50 and the standard deviation of the sample is 2. Test the hypothesis that the true mean is 56, and also test the hypothesis that the true mean is 40. Report p-values and comment on the results.

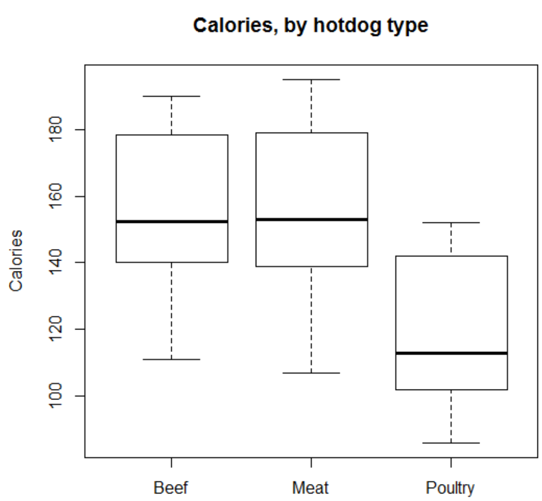
   The null hypothesis that the true mean = 56 is rejected, $p < 0.001$ (p = 1.70085e -51)

   The null hypothesis that the true mean = 40 is rejected, $p < 0.001$ (p = 4.595366e-72)

Data Set: hot_dogs.csv

1) Use hot_dogs.csv data and hypothesis tests to determine which type of hot dog has average calories less than 140 with 95% confidence. Present boxplots of calories by type of hot dog.

```
> with(beef, t.test(Calories)$conf.int[1])
[1] 146.2532
> with(meat, t.test(Calories)$conf.int[1])
[1] 145.7308
> with(poultry, t.test(Calories)$conf.int[1])
[1] 107.1698
```



2) Using hot_dogs.csv data and hypothesis tests at the 95% confidence level, determine which type of hot dog has an average Sodium level different from 425 milligrams.

   None of the three p-values is less than 0.05, so we do not reject the null hypothesis: mu = 425 mg.

3) Using the **hot_dogs.csv** data and hypothesis tests, determine if the variance in Sodium values for beef hot dogs is different from 6000 with 95% confidence.

```
> var.conf.int = function(x, conf.level = 0.95) {
+     df <- length(x) - 1
+     chilower <- qchisq((1 - conf.level)/2, df, lower.tail = TRUE)
+     chiupper <- qchisq((1 - conf.level)/2, df, lower.tail = FALSE)
+     v <- var(x)
+     c(df * v/chiupper, df * v/chilower)
+     }
> with(beef, var.conf.int(Sodium))
[1]  6068.506 22384.122
>
> # If this logical is TRUE, then we reject the null hypothesis that mu = 6000:
> with(beef, (6000 < var.conf.int(Sodium)[1]) || (6000 > var.conf.int(Sodium)[2]))
[1] TRUE
```

4) A coin is flipped 100 times. If it is unbiased the probability of a heads should equal the probability of a tails. At the 95% confidence level, test the null hypothesis the coin is unbiased versus the alternative that it is biased if 43 heads are obtained. Test the same hypothesis if 63 heads are obtained. Use one-sided hypothesis tests.

prop.test(x = 43, n = 100, alternative = "less") # see p-value 0.0968 > 0.05 (do not reject null hypothesis)

1-sample proportions test with continuity correction
data: 43 out of 100, null probability 0.5
X-squared = 1.69, df = 1, p-value = 0.0968
alternative hypothesis: true p is less than 0.5
95 percent confidence interval: 0.000000 0.517194

prop.test(x = 63, n = 100, alternative = "greater") # see p-value 0.00621 < 0.05 (reject null hypothesis)

1-sample proportions test with continuity correction
data: 63 out of 100, null probability 0.5
X-squared = 6.25, df = 1, p-value = 0.00621
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval: 0.5430629 1.0000000

Neither hypothesis can be rejected.

5) s**alaries.csv** contains data derived from a November 8, 1993 article in Forbes titled "America's Best Small Companies". The file gives the CEO age and salary for 60 small business firms. Use these data to test the hypothesis at 95% confidence that at least 50% of the CEOs are 45 years old or older. Also test the hypothesis at 95% confidence that at least 50% of the CEOs earn less than $500,000 per year. Use one-sided hypothesis tests.

First, test the hypothesis that at least 50% of the CEOs are 45 years old or older.

age <- salaries$AGE >= 45
count <- sum(age)
total<- length(age)
prop.test(x = count, n = total, alternative = "greater")

1-sample proportions test with continuity correction
data: count out of total, null probability 0.5
X-squared = 22.8167, df = 1, p-value = 8.911e-07
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval: 0.7121941 1.0000000 reject the null hypothesis
sample estimates: p = 0.8166667   Note: 81.7% are 45 years old or older

Now use prop.test to test if 50% of the CEO's earn less than $500,00 per year.

```
salary <- salaries$SAL < 500
count <- sum(salary)
total <- length(salary)
prop.test(x= count, n = total, alternative = "greater")
```

1-sample proportions test with continuity correction
data: count out of total, null probability 0.5
X-squared = 10.4167, df = 1, p-value = 0.0006244
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval: 0.6045034 1.0000000 reject the null hypothesis
sample estimates: p = 0.7166667 Note: 71.7% earn less than $500,000 per year.