Statistics and Data Analysis
Suggested Solution for Homework 5

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

1. (a) Let μ be the average price of cleaning a 12' by 18' wall-to-wall carpet. The hypothesis is

$$H_0 : \mu = 50$$
$$H_a : \mu > 50.$$

   (b) Because the population variance is known and the population is normal, we may use the z test. In this one-tailed test, the p-value is 0.0001. The R scripts that finds the p-value are

```
install.packages("BSDA")
library("BSDA") # for z.test()
price <- c(53, 51, 57, 49, 50, 52, 50, 48, 53, 52,
           51, 48, 58, 55, 51, 52, 53, 56, 53, 56,
           56, 57, 49, 48, 51)
z.test(price, alternative = "g", mu = 50, sigma.x = 3.25)
```

   In MS Excel, one may use "Z-Test: Mean" in "Data Analysis Plus" to find the p-value.

   (c) For $\alpha = 0.05$ or 0.01, as the p-value is smaller than α, we reject H_0. With a 95% or 99% significance level, there is a strong evidence showing that the average price in the region in which the company operates is higher than \$50.

2. (a) Let μ be the average diameter (in centimeters) of the punched hole. The hypothesis is

$$H_0 : \mu = 1.9$$
$$H_a : \mu \neq 1.9.$$

   (b) Because the population variance is unknown and the population is normal, we may use the t test. In this two-tailed test, the p-value is 0.00128 (without being multiplied by 2). The R scripts that finds the p-value are

```
hole <- c(1.92, 1.89, 1.87, 1.82, 1.85, 1.87, 1.87, 1.86, 1.85, 1.84)
t.test(hole, alternative = "t", mu = 1.9)
```

   In MS Excel, one may use "T-Test: Mean" in "Data Analysis Plus" to find the p-value.

   (c) For $\alpha = 0.05$ or 0.01, as the p-value is smaller than $\frac{\alpha}{2}$, we reject H_0. With a 95% or 99% significance level, there is a strong evidence showing that the average diameter (in centimeters) of the punched hole is not 1.9 cm.

3. (a) Let μ be the average yearly dental expenditure (in \$) per family. The hypothesis is

$$H_0 : \mu = 1200$$
$$H_a : \mu \neq 1200.$$

   (b) Because the population variance is unknown and the population is normal, we may use the t test. In this two-tailed test, the p-value is 0.0091 (without being multiplied by 2). The R scripts that finds the p-value are

```
dental <- c(1018,  850, 1127, 1325, 1308, 1414, 836, 1023, 1288, 855,
             931,  733,  698,  875,  915,  945, 955,  988, 1696, 998,
             968, 1724, 1008,  896, 1241)
t.test(dental, alternative = "t", mu = 1200)
```

   In MS Excel, one may use "T-Test: Mean" in "Data Analysis Plus" to find the p-value.

(c) For $\alpha = 0.05$, as the p-value is smaller than $\frac{\alpha}{2}$, we reject H_0. With a 95% significance level, there is a strong evidence showing that the average yearly dental expenditure per family has increased. However, for $\alpha = 0.01$, as the p-value is greater than $\frac{\alpha}{2}$, we do not reject H_0. We can make no conclusion at a 99% confidence level.

4. (a) Let p be the proportion of consumers who consumer at least one bottle of milk per month. The hypothesis is

$$H_0 \colon p = 0.76$$
$$H_0 \colon p < 0.76.$$

(b) By counting the number of 1s in the given data set, we found that 223 out of 315 consumers consume at least one bottle of milk per month. Therefore, the sample proportion $\hat{p} = \frac{223}{315} = 0.7079$. Because the sample size $n \geq 30$, $n\hat{p} \geq 5$, and $n(1 - \hat{p}) \geq 5$, we may use the z test to test the population proportion. In this left-tailed test, the p-value is 0.0152. The R scripts that finds the p-value is[1]

```
prop.test(223, 315, alternative = "l", p = 0.76, correct = FALSE)
```

In MS Excel, one may use "Z-Test: Proportion" in "Data Analysis Plus" to find the p-value.

(c) As the p-value is larger than $\alpha = 0.01$, we do not reject H_0. With a 99% confidence level, there is no strong evidence showing that the proportion of consumers who consume at least one bottle of milk has decreased. A new promotion should not be launched.

[1] The last argument `correct` determines whether a correction of continuity will be conducted. This is beyond the scope of this course. In general, if the sample size is large enough, whether to do a correction of continuity does not matter a lot.