Question 1 (22 points)

Scarlett and Heather, the owners of an upscale restaurant in Dayton, Ohio, what to study the dining characteristics of their customers. They decide to focus on two variables: the amount of money spent by customers and whether customers order dessert. The results from a sample of 60 customers are as follows:

Amount spent: X=$38.54, s=$7.26

Eighteen customers purchased dessert.

a. (5 points) Construct a 95% confidence interval estimate for the population mean amount spent per customer in the restaurant. Provide the step calculations. Use Excel or R function to compute the multiple. State the confidence interval in the two forms: **point estimate ± margin of error** and**[lower limit, upper limit].**

Solution:

In this problem, as the standard deviation for the population is ***unknown***, and our sample size $n = 60 > 30$, the normal sampling distribution can be justified, and the confidence interval (***CI***) for the population mean under *t-multiple* is:

$$
\begin{aligned}
CI &= \bar{X} \pm MOE \\
MOE &= (t - multiple) * SE \\
\because SE &= \frac{s}{\sqrt{n}} \\
t - multiple &= t_{n-1,1-\frac{\alpha}{2}} \\
\therefore CI &= \bar{X} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \\
&= 38.54 \pm t_{59,1-\frac{0.05}{2}} \frac{7.26}{\sqrt{60}} \\
&= 38.54 \pm \mathrm{T.INV.2T}(0.05, 59) \frac{7.26}{\sqrt{60}} \\
&= 38.54 \pm 1.87546 \\
&= [36.66454,\ 40.41546]
\end{aligned}
$$

b. (5 points) Construct a 90% confidence interval estimate for the population proportion of customers who purchase dessert. Provide the step calculations. Use Excel or R function to compute the multiple. State the confidence interval in the two forms: **point estimate ± margin of error** and**[lower limit, upper limit].**

Solution:

Let p be the proportion of all customers that purchased dessert. From this random sample of size $n = 60$,the sample proportion is $\hat{p} = \frac{18}{60} = 30\%$. As $\boldsymbol{np}$ & $\boldsymbol{n(1-p) \geq 5}$, we say that $Binomial(n,p)$ is approximately normally distributed, with ***CI*** of proportion as:

$$
\begin{aligned}
CI &= \hat{p} \pm MOE \\
MOE &= (Z - multiple) * SE(\hat{p}) \\
\because SE(\hat{p}) &\approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
Z - multiple &= z_{1-\frac{\alpha}{2}} \\
\therefore CI &= \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
&= 0.3 \pm z_{1-\frac{0.1}{2}} \sqrt{\frac{0.3 * 0.7}{60}} \\
&= 0.3 \pm \mathrm{NORM.S.INV}(0.95) \sqrt{\frac{0.3 * 0.7}{60}} \\
&= 0.3 \pm 0.097 \\
&= [0.203,\ 0.397]
\end{aligned}
$$

Jeanine, the owner of a competing restaurant, wants to conduct a similar survey in her restaurant. Jeanine does no have access to the information that Scarlett and Heather have obtained from the survey they conducted. Answer the following questions:

c. (5 points) What sample size is needed to have 95% confidence of estimating the population mean amount spent in her restaurant to within ± \$1.50, assuming that the standard deviation is estimated to be \$8? Provide the step calculations.

Solution:

Here we assume that the standard deviation is estimated as: $\sigma_{est} = \$8$, and then we can use *Z-multiple* to calculate suitable sample size, as:

$$
\begin{aligned}
n &= (\frac{(Z - multiple)\sigma_{est}}{B})^2 \\
\because Z - multiple &= z_{1-\frac{\alpha}{2}},\ B = \$1.50 \\
\therefore n &= (\frac{z_{1-\frac{\alpha}{2}}\sigma_{est}}{B})^2 \\
&= (\frac{\mathrm{NORM.S.INV}(0.975) * 8}{1.5})^2 \\
&= 110
\end{aligned}
$$

d. (5 points) How many customers need to be selected to have 90% confidence of estimating the population proportion of customer who purchase dessert to within ± 0.04? Provide the step calculations.

Solution:

Here we have no prior information about p, we select the worst-case and use $p_{est} = 0.5$, and get the sufficient sample size as:

$$
\begin{aligned}
n &= (\frac{Z - multiple}{B})^2 p_{est}(1 - p_{est}) \\
\because Z - multiple &= z_{1-\frac{\alpha}{2}},\ B = 0.04 \\
\therefore n &= (\frac{z_{1-\frac{\alpha}{2}}}{B})^2 p_{est}(1 - p_{est}) \\
&= (\frac{\mathrm{NORM.S.INV}(0.95)}{0.04})^2 * 0.5 * (1 - 0.5) \\
&= 423
\end{aligned}
$$

e. (2 points) Based on your answers to (c) and (d), how large a sample should Jeanine take?

Solution:

In order to satisfy both the requirement, Jeanine should use the larger one of the required sample size, **423**, as the sample size.

Question 2 (13 points)

A drugstore manager needs to purchase adequate supplies of various brands of toothpaste to meet the ongoing demands of its customers. In particular, the company is interested in estimating the proportion of its customers who favor the county's leading brand of toothpaste, Crest. The Data sheet of the file 'Toothpaste.xlsx' contains the toothpaste brand preferences of 200 randomly selected customers, obtained recently through customer survey.

a. (3 points) Compute the proportion of customers who prefer each brand (Aquafresh, Colgate, Crest, Mentadent, and others)

Hint: You can use pivot table or 'Excel IF' function.

Solution:

Using pivot table, we can find the given dataset has no error(blank), and the table shows the proportion for the proportion of customers who prefer each brand as:

	Aquafresh	Colgate	Crest	Mentadent	Others	(blank)	Grand Total
Count of Customer	9.00%	15.50%	30.00%	13.50%	32.00%	0.00%	100.00%

b. (5 points) Calculate a 95% confidence interval for the proportion of customers who prefer Crest toothpaste.

Solution:

As $\boldsymbol{np}$ & $\boldsymbol{n(1-p) \geq 5}$, we say that $Binomial(n,p)$ is approximately normally distributed, with ***CI*** of proportion as:

$$
\begin{aligned}
CI &= \hat{p} \pm MOE \\
MOE &= (Z - multiple) * SE(\hat{p}) \\
\because SE(\hat{p}) &\approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
Z - multiple &= z_{1-\frac{\alpha}{2}} \\
\therefore CI &= \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
&= 0.3 \pm z_{1-\frac{0.05}{2}} \sqrt{\frac{0.3 * 0.7}{200}} \\
&= 0.3 \pm \mathrm{NORM.S.INV}(0.975) \sqrt{\frac{0.3 * 0.7}{200}} \\
&= 0.3 \pm 0.06351 \\
&= [0.23649,\ 0.36351]
\end{aligned}
$$

c. (5 points) How large should the sample be if they would like to *ensure* that MOE is no larger than 0.05 or 5% at 95% confidence?

Solution:

Since our data sheet size 200 is relatively large, we can use the sample proportion of Crest to estimate:

$$
\begin{aligned}
n &= (\frac{Z - multiple}{B})^2 p_{est}(1 - p_{est}) \\
\because Z - multiple &= z_{1-\frac{\alpha}{2}},\ B = 0.05,\ p_{est} = 0.3 \\
\therefore n &= (\frac{z_{1-\frac{\alpha}{2}}}{B})^2 p_{est}(1 - p_{est}) \\
&= (\frac{\mathrm{NORM.S.INV}(0.975)}{0.05})^2 * 0.3 * (1 - 0.3) \\
&= 323
\end{aligned}
$$

Question 3 (15 points)

The data 'Bottles.xlsx' represents the amount of soft drink filled in a sample of 50 consecutive 2-liter bottles. At the 5% level of significance, is there evidence that the mean amount of soft drink filled is different from 2.0 liter?

a. (5 points) Specify the null hypothesis and the alternative hypothesis. State whether you are using a one- or two-tailed test and explain your reasoning.

Solution:

- H_0**(Null Hypothesis): we hypothesize that the mean amount of soft drink filled is** *equal to*** 2.0 liter, as $\mu = 2.0$

- $\mathbf{H_a}$(Alternative Hypothesis): we hypothesize that the mean amount of soft drink filled is ***different from*** 2.0 liter, as $\mu \neq 2.0$

It's a two-tailed test, because our goal is to test whether the mean is different from 2.0 liter, and both less than and greater than should be counted.

b. (5 points) Conduct calculations by hand to compute the critical value and p-value of the test. Determine your decision.

Solution:

1. : Identify the null hypothesis and the alternative hypothesis

   Let μ be the mean amount of soft drink filled.

   H_0: $\mu = 2.0$

   H_a: $\mu \neq 2.0$

   - The hypothesis test is a two-tailed test
   - Set the significance level α to 5%
2. Compute the *t-value*:

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{2.00072 - 2.0}{\frac{0.044563662}{\sqrt{50}}} = 0.114244848 \tag{1}$$

3. Make a decision by hand.

   Method 1: Critical value method by hand

   Critical value = T. INV(0.975,49) = 2.009575>|0.114244848(*t-value*)|

   Method 2: p-value method by hand

   T. DIST(-0.114244848,49,1) =0.454755

   p-vale = 2*0.454755 = 0.909511>0.05

c. (5 points) Communicate the results of the test in “plain English” including the meaning of the p-value.

Solution:

4. Interpret the results:

   - Conclusion: We fail to reject H_0 (null hypothesis) because our sample didn’t provide enough evidence against it, and the mean amount of soft drink filled is the usual 2.0 liter.

   The p-value of 0.909511 means that:

   1. Assuming that the true mean amount of soft drink filled is 2.0 liter.
   2. there is a 90.9511% chance
   3. that we get a mean of 2.00072 or more on a sample of 50.