Question 1 (15 points)

A consumer products firm has recently introduced a new brand. The firm would like to estimate the proportion of people in its target market segment who are aware of the new brand. As part of a larger market research study, it was found that in a sample of 125 randomly selected individuals from the target market segment, 84 individuals were aware of the firm's new brand.

The manager in charge of the new brand has stated that the brand awareness is greater than 0.65, meaning that more than 65% of the population is aware of the brand. At the 5% significance level, conduct a hypothesis test with the goal of proving his claim.

a. (4 points) Specify the null hypothesis and the alternative hypothesis. Define the parameter of your interest and write the mathematical expressions of your hypotheses in terms of your parameter.

Solution:

1. Identify the null hypothesis and the alternative hypothesis

- $\mathbf{H_0}$(Null Hypothesis): we hypothesize that the brand awareness is ***less than or equal to*** 0.65, as $p = 0.65$
- $\mathbf{H_a}$(Alternative Hypothesis): we hypothesize that the brand awareness is ***greater than*** 0.65, as $p \neq 0.65$

b. (1 point) What test should be run?

(One) sample, (one) tailed, (Z)- test.

c. (8 points) Determine your decision using two methods: **the critical value** and **the p-value approach**. Conduct calculation by hand to compute the critical value and p-value of the test. Please specify the Excel functions you use.

Solution:

2. Compute the *Z-value*:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$\because \hat{p} = \frac{84}{125} = 0.672$$

$$\therefore Z = \frac{0.672 - 0.65}{\sqrt{\frac{0.65(1-0.65)}{125}}} \approx 0.515688$$

3. Make a decision by hand.

   Method 1: Critical value method by hand

   $z_{1-\alpha}$ = NORM.S.INV(0.95) = 1.645 > 0.515688

   Method 2: p-value method by hand

   p-value = 1 - NORM.S.DIST(0.515688,1) = NORM.S.DIST(-0.515688,1) = 0.303062 > 0.05

d. (2 points) Provide 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 proportion of brand awareness is not statistically significantly greater than 65%.

   The p-value of 0.303062 means that:

   1. Assuming that the true proportion of brand awareness is 0.65.
   2. there is a 30.3062% chance
   3. that we get a proportion of 67.2% or more on a sample of 125.

Question 2 (10 points)

A company would like to estimate the effect of traditional and new styles of soft-drink cans.

To determine whether a new style is popular with the consumers, the company rounds a number of focus group session around the country. At each of these sessions, randomly selected consumers are allowed to examine the new and traditional styles, exchange ideas, and offer their opinions. Eventually, they fill out a form where, among other questions, they are asked to respond to the following items, each on a scale of 1 to 7, 7 being the best:

- Rate the attractiveness of the traditional style can (AO).
- Rate the attractiveness of the new style can (AN).

The data are listed in the file Soft-Drink Can.xlsx.

The company wants to compare the means to see whether consumers rate the attractiveness higher for a new-style can than the traditional-style can.

a. (4 points) Specify the null and the alternative hypotheses to compare the means.

Solution:

- $\mathbf{H_0}$(Null Hypothesis): we hypothesize that the the new style can has a ***less or equal grading than*** the traditional style can, as $\mu_n \leq \mu_o$, or $d = \mu_n - \mu_o \leq 0$
- $\mathbf{H_a}$(Alternative Hypothesis): we hypothesize that the the new style can has a ***greater grading than*** the traditional style can, as $\mu_n > \mu_o$, or $d = \mu_n - \mu_o > 0$

b. (1 point) What test should be run?

(Paired) sample, (one) tailed, (t)- test.

c. (5 points) Use R to conduct the hypothesis test. Copy and paste the command and the report generated by R. Provide the results of the test in “plain English”.

Solution:

```
> attach(Soft_Drink_Cans)
"
The following objects are masked from Soft_Drink_Cans (pos = 3):

    AN, AO, Consumer
"
> t.test(AN, AO , alternative = "greater", paired = TRUE )
"
  Paired t-test

data:  AN and AO
t = 5.3514, df = 179, p-value = 1.323e-07
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.3723901       Inf
sample estimates:
mean of the differences
              0.5388889
"
```

Interpret the results:

- From the result, we get $\bar{d} = 0.5388889, t = 5.3514$, p-value $= 1.323e-07 < 0.05 $
- Conclusion: We reject H_0 (null hypothesis) because our sample provided enough evidence against it, and the the new style can has a ***greater grading than*** the traditional style can.

The p-value of 1.323e-07 means that:

1. Assuming that the new style can has a ***same grading as*** the traditional style can
2. there is a 1.323e-07 chance
3. that we get a mean of difference as 0.5388889 or more on a sample of 180.

Question 3 (25 points)

Financial analysts specializing in credit markets are often interested in creating models to predict whether a firm will go bankrupt within some fixed period of time. If there is a good chance that a particular firm will go bankrupt, then the firm will have to pay a very high interest rate on any debt (bonds) that it may issue.

In practice, statistical models to predict bankruptcy are fairly difficult to construct. One of the variables that may be useful in distinguishing between firms that go bankrupt and firms that stay solvent is the return on assets (ROA). The accompanying file Bankrupcy.xlsx contains financial data on 44 firms. Of these 44 firms, 20 firms went bankrupt within 1 year after the data were collected; the other 24 firms remained solvent after 1 year. For this assignment, ignore all financial measures other than ROA.

To investigate the ROA variable, first **unstack** the ROA variable using "Bankrupt" as the **code variable;** see worksheet "Unstacked Data". You may also try to create the unstacked data using R for practice. Here is the code:

```
#Create a data frame using two variabels ROA and BANKRUPT

ROA.Bkcy<-data.frame(Bankruptcy$ROA,Bankruptcy$BANKRUPT)

# Unstack ROA,Bkcy

Unstack.ROA=unstack(ROA.Bkcy)
```

Then, you have two columns of ROA data. For the purpose of this exercise we will assume that firms' ROA is, in general, normally distributed.

Part I

a. (5 points) Conduct a hypothesis test that the average ROA of firms that *went bankrupt* is less than -5%.

- (4 points) Specify the null hypothesis and the alternative hypothesis. Define the parameter of your interest and write the mathematical expressions of your hypotheses in terms of your parameter.

  Solution:

  - $\mathbf{H_0}$(Null Hypothesis): we hypothesize that the average ROA of firms that *went bankrupt* is ***greater than or equal to*** -5%, as $\mu \geq -5\%$
  - $\mathbf{H_a}$(Alternative Hypothesis): we hypothesize that the average ROA of firms that *went bankrupt* is ***less than*** -5%, as $\mu < -5\%$
- (1 point) What test should be run?

(One) sample, (one) tailed, (t)- test.

b. (8 points) Conduct calculation by hand to compute the p-value for this test.

Solution:

```
> Unstack.ROA=unstack(ROA.Bkcy)
> View(Unstack.ROA)
> t.test(Unstack.ROA[["Yes"]], mu = -0.05, alternative = "less")
"
  One Sample t-test

data:  Unstack.ROA[["Yes"]]
t = -0.64514, df = 19, p-value = 0.2633
alternative hypothesis: true mean is less than -0.05
95 percent confidence interval:
        -Inf -0.01639542
sample estimates:
mean of x
    -0.07
"
```

Note: Sorry I just choose to use R here, because it's part of the entire problem in R, and R is more convenient. Also, in question 1 there's already one similar question calculating by hand, with similar method. Don't deduct my score because of that :)

c. (2 points) At the 10% level of significance, is the average ROA of firms that went bankrupt less than -5%? Provide the results of the test in “plain English” including the meaning of the p-value.

Solution:

Interpret the results:

- From the result, we get $\mu = -0.07, t = -0.64514$, p-value $= 0.2633 > 10\%$
- Conclusion: We fail to reject H_0 (null hypothesis) because our sample didn’t provide enough evidence against it, and there’s no strong support to say that the average ROA of firms that *went bankrupt* is ***less than*** -5%

The p-value of 0.2633 means that:

1. Assuming that the **true** average ROA of firms that *went bankrupt* is 5%
2. there is a 26.33% chance
3. that we get a mean of ROA as -7% or less on a sample of 20.

Part II

d. (10 points) Conduct a hypothesis to test whether the average ROA of firms that *remained solvent* is greater than the average ROA of firms that *went bankrupt*.

- (4 points) Specify the null hypothesis and the alternative hypothesis. Define the parameters of your interest and write the mathematical expressions of your hypotheses in terms of your parameter.

  Solution:

  - $\mathbf{H_0}$(Null Hypothesis): we hypothesize that the average ROA of firms that *remained solvent* is ***less than or equal to*** the average ROA of firms that *went bankrupt*, as $\mu_s \leq \mu_b$
  - $\mathbf{H_a}$(Alternative Hypothesis): we hypothesize that the average ROA of firms that *remained solvent* is ***greater than*** the average ROA of firms that *went bankrupt*, as $\mu_s > \mu_b$
- (1 point) What test should be run?

(Two) sample, (one) tailed, (t)- test.

e. (5 points) Use R to conduct the hypothesis test. Copy and paste the command and the report generated by R. Provide the results of the test in “plain English”.

Solution:

Here we don’t know whether the two sample has the equal variance. To make the most robust conclusion, we do it in both situations:

Equal-variance assumption:

```
> t.test(Unstack.ROA[["No"]],Unstack.ROA[["Yes"]],alternative = "greater",var.equal = TRUE)
"
  Two Sample t-test

data:  Unstack.ROA[["No"]] and Unstack.ROA[["Yes"]]
t = 4.1602, df = 42, p-value = 7.686e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.07520753        Inf
sample estimates:
mean of x mean of y
  0.05625  -0.07000
"
```

Unequal-variance assumption:

```
> t.test(Unstack.ROA[["No"]],Unstack.ROA[["Yes"]],alternative =
"greater",var.equal = FALSE)
"
  Welch Two Sample t-test

data:  Unstack.ROA[["No"]] and Unstack.ROA[["Yes"]]
t = 3.8707, df = 23.065, p-value = 0.0003861
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.07035491        Inf
sample estimates:
mean of x mean of y
  0.05625  -0.07000
"
```

Interpret the results:

- From the result, we get $p_1 = 7.686e-05 < 1\%, p_2 = 0.0003861 < 1\%$, meaning that our conclusion would be very robust.
- Conclusion: We reject H_0 (null hypothesis) because our sample provide enough evidence against it, and the average ROA of firms that *remained solvent* is ***greater than*** the average ROA of firms that *went bankrupt*.