Question 1. [10 marks]

A labour union representing workers from a large manufacturing company claims that the true unemployment rate based on their own administrative data is higher than the official unemployment rate of 6.0% used by the government to calculate unemployment benefits for their members. To strengthen their argument, they commission a survey of 1,000 residents where the company is located and find a 7.5% unemployment rate. *Assuming that the sampling method used by the labour union survey is consistent with that used to compute the official unemployment rate*, answer the following questions.

a) Does the unemployment rate estimated by the labour union provide enough evidence to support its claim? Use the critical value approach and a 0.05 level of significance.

[4] $H_0: \cancel{\mu}\ p = 0.06$ $H_a: \cancel{\mu}\ p > 0.06$ one-sided

4

$$Z_{stat} = \frac{0.075 - 0.06}{\sqrt{0.06 \times 0.94 / 1000}} = \frac{0.015}{0.0075} \approx 1.997$$

$Z_{stat} > Z_{crit}$

Z_{crit} @ Level of Significance of 0.05 = 1.645

Because $Z_{stat} > Z_{crit}$ we have sufficient evidence to reject H_0 and conclude that unemployment rate is higher than 6%

b) Calculate an appropriate 95% 1-sided confidence interval to estimate the unemployment rate based on the labour union survey data (5 decimal places). Is this interval consistent with your conclusion in part a) above? Explain.

[3] $LB = \bar{x} - Z_\alpha * \sqrt{\hat{p}\hat{q}/n}$

3

$= 0.075 - 1.645 * \sqrt{0.075 \times 0.925 / 1000}$

$= 0.075 - 0.0137 = 0.06130$

$CI = (0.06130, 1)$ This is consistent with part A, because the CI does not include the hypothesised value (0.06) so we reject H_0 JUST like in part a.

c) What sample size would be required to estimate the unemployment rate with a margin of error of plus-or-minus 0.015 (or ± 1.5%) using a 95% confidence interval? Assuming that the value of the sample proportion remains unchanged, what would be a 95% 2-sided confidence interval under these conditions? Explain if and why such an interval would change your conclusions in parts (a) and (b) above.

$2\frac{1}{2}$

[3] $n = \dfrac{1.96^2 \times 0.075 \times 0.925}{0.015^2}$

≈ 1185 people

$CI = 0.075 \pm 1.96 \times \sqrt{0.075 \times 0.925 / 1000}$

$= 0.075 \pm 0.016325$

$= (0.0586, 0.0913)$

Because it's 2-sided the Interval would ~~change~~ your conclusion because in the new CI, the hypothesised value is ~~now~~ included inside, where as before we had the hypothesised value outside of the CI

Question 2. "Can you handle it?" [10 marks]

The vast majority of people have hand dominance--they are either left-handed or right-handed. *In this question we will ignore the very small percentage of the population who are ambidextrous (i.e., those who use both hands equally well or equally badly!).*

A researcher believes that men are more likely to be left-handed than women. To check this she checked the dominant hand of random samples of 200 men and 200 women, and found 24 men and 18 women to be left-handed. Let p_M and p_W be the proportions of left-handed men and left-handed women, respectively, in the population.

a) Construct an appropriate 95% 1-sided confidence interval to estimate the difference $p_M - p_W$.

[2]

$H_0: p_m - p_w = 0 \quad H_a: p_m - p_w > 0$

$\hat{p}_m = 0.12 \quad \hat{p}_w = 0.09$

$\hat{q}_m = 0.88 \quad \hat{q}_w = 0.91$

$LB = (.12 - .09) - 1.645 \times \sqrt{.12 \times .88/200 + .09 \times .91/200}$

$= 0.03 - 0.0504 = -0.0204 \quad CI = (-0.0204, 1)$

2

b) Does your answer to part a) support the researcher's claim that men are more likely to be left-handed? Please explain your reasoning briefly.

[1] No it does not because the CI includes the 0 (from the RHS of the hypothesis), IE the hypothesised value is included in the CI, so Ha is rejected.

1

c) Confirm your conclusion in part b) by formulating and testing the hypothesis (at the 5% significance level) that men are more likely to be left-handed than women in the population. Make sure to compute a p-value.

[5]

$Z_{stat} = \frac{(0.12 - 0.09)}{\sqrt{\bar{p} \times \bar{q} \times (1/n_1 + 1/n_2)}} = \frac{0.03}{\sqrt{0.105 \times 0.895 \times (1/200 + 1/200)}} = \frac{0.03}{0.0306553421} \approx 0.98$

$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{24 + 18}{400} = 0.105$

$\bar{q} = 1 - \bar{p} = 0.895$

concl?

P value = 0.8363

1 - ANS = 0.1635 < 0.05

4½

Because 0.05 (level of significance) < Pvalue 0.1635 there is insufficient evidence to reject Ha.

* d) In a different study, a 95% confidence interval for the proportion of **right-handed** people in the general population was constructed. The interval was 0.89 ± 0.04. Using the same sample, would a 95% confidence interval for the proportion of **left-handed** people in the general population be 0.11 ± 0.04? Please circle your answer choice.

[1]

True   False   Can't tell

0

e) Explain if the calculations in parts (a) and (c) require any assumptions on the distribution of data on left-handedness in the general population.

[1] - The sample proportions must have a normal distribution to use the z-score tables

0

Question 3. [10 marks]

The beta of a stock measures its volatility relative to the stock index. A beta less than 1 indicates a less volatile stock while a beta greater than 1 indicates a stock more volatile than the index. Energy stocks historically have been more volatile than the market in general. You take a sample of energy stocks to see if this is currently true. **Appendix A** gives a plot of the beta values for a sample of energy stocks and summarizes the observations.

a) Test whether there is sufficient evidence to claim that energy stocks currently are more volatile than the market in general. Use the p-value approach and a 0.05 level of significance.

[4]

$H_0: \mu = 1$ $H_a: \mu > 1$ One - sided Df = 32

4

$t_{stat} = \frac{1.1643 - 1}{0.4123/\sqrt{33}} = \frac{0.1643}{0.0718} \simeq 2.29$ with Df = 32

Pvalue = Between (0.025 and 0.10)

Pvalue (Between 0.025 and 0.10) < Significance level of 0.05

Because the Pvalue is < significance level there is sufficient evidence to reject Ho and assume energy stocks are more volatile

b) Calculate a 95% 1-sided confidence interval to estimate the average beta for current energy stocks. Explain how this confirms your decision and conclusion above.

[3] $LB = 1.1643 - 1.69 * (0.4123/\sqrt{33}) = 1.1643 - 0.1212950$
$= 1.043$

3

$CI (1.043, \infty)$

Because the hypothesized value (1) is not in the CI we have sufficient evidence to reject Ho like in part a).

c) Suppose you knew the standard deviation (σ) of beta values for energy stocks was 0.40, and decide to reject the null hypothesis if the z-statistic exceeds 1.645. What values for the sample mean would lead you to reject the null hypothesis, assuming a sample size of 33?

[1] $1.645 < \frac{\bar{x} - 1}{0.4/\sqrt{33}} \rightarrow 1.645 < \frac{\bar{x} - 1}{0.069631062} \rightarrow 0.11454 < \bar{x} - 1$
$\rightarrow \bar{x} > 1.11454$

1

Any values above 1.11454 for the sample mean would lead you to reject Ho

d) Explain whether the assumptions of the test and confidence interval above are valid. State the **minimum** requirements for the distribution of beta values in the population of energy stocks and explain whether these assumptions are warranted or justified.

1

[2] The distribution must be ~~normally distributed~~.

Explanation:

The boxplot does not show major Skewdness, however there is an outlier. This outlior can lead us to believe that the test and CI are not valid, because the outlier assumption voids normal distibution making the tests void.

Question 4. (10 Marks)

FedEx and United Parcel Service (UPS) are the world's two leading cargo carriers by volume and revenue (*The Wall Street Journal*, January 27, 2004). According to the Airports Council International, the Memphis International Airport (FedEx) and the Louisville International Airport (UPS) are 2 of the 10 largest cargo airports in the world. The following data show the volume of cargo (thousands of tons) handled by these airports for two randomly selected samples of days.

Memphis

9.1	15.10	8.8	10.0	7.5	10.5
8.3	9.1	6.0	5.8	12.1	9.3

Louisville

4.7	5.0	4.2	3.3	5.5
2.2	4.1	2.6	3.4	7.0

Appendix B includes side-by-side boxplots of the data and a partial Minitab output.

a) Test at the 5% level of significance whether the mean daily cargo volumes for the two airports are different using the critical value approach. Assume the two population variances are unequal. $H_0: \mu_m - \mu_L = 0$ $H_a: \mu_m - \mu_L \neq 0$

[4]

$$T_{stat} = \frac{(9.3 - 4.2)}{\sqrt{2.54^2/12 + 1.43^2/10}} = \frac{5.1}{0.86144658} = 5.920$$

4

$5.920 > 2.11$

T_{crit} @ 0.05 $DF = 17 = 2.11$

Because Tstat > Tcrit, there is sufficient evidence to reject Ho and conclude that the daily cargo volumes for each airport is different

b) What is the p-value for the test?

$\frac{1}{2}$

[1] 5.920 with DF = 17 → P value < 0.0001 × 2

c) Are the population means different? Justify your answer by calculating the 95% confidence interval for the difference between the means of the populations of daily cargo volumes for the two airports. Explain the meaning of the confidence interval.

[3] $(9.3 - 4.2) \pm 2.11 * \sqrt{2.54^2/12 + 1.43^2/10}$

$= 5.1 \pm 1.817692849 \rightarrow (3.28, 6.92)$

2

The CI says that 95% of the time that the difference between the means in daily cargo volume is between [3.28 and 6.92]. The hypothesized value of 0 is not included so we have sufficient evidence to reject Ho and conclude the population means are different.

d) Are the assumptions of the test above warranted or justified? Explain briefly.

2

[2] - The data for memphis has one ~~extreme~~ outlier @ 15.1.
This interferes with the outlier condition and does not allow for the sample to be normally distributed. The Lousiville data has no outliers + no skewness, so it can be concluded as normally distributed.
- However the samples are independent and that condition checks out because the data was randomly selected