**ADM 2304** **Midterm Exam Solutions** **Winter 2016**

**Question 1. [11 marks ]**

a. **(4 marks)** Test the hypothesis that the mean concentration of blood serum level in this population is less than 22 ng/mL. Use the **p-value approach**.

S1: H0: $\mu = 22$ Ha: $\mu < 22$ (1 mark)

S2: $t_{Calc} = \frac{\bar{X} - \mu_0}{SE(\bar{X})} = \frac{19.655 - 22}{4.507/\sqrt{22}} = \frac{-2.345}{0.9609} = -2.4404$ (1 mark)

S5: p_Val = P[t(21) < -2.4404] x 1 = P[t(21) > 2.4404] ≈ 0.0125 (1 mark)

{more precisely, p_Val = 0.0118} *From t-table, 0.01 < p-value < 0.025 deserves the full mark.*

Since p_Val < 0.025 < [α =.05] ==> Reject H0 (0.5 mark)

Based on the available statistical evidence, one can claim that the population mean of the serum concentration of vitamin D is less than 22 ng/mL. (0.5 mark)

b. **(2 marks)** For the test above, calculate the appropriate confidence interval for the mean concentration and explain why it is consistent with your conclusion.

$UB = \bar{X} + t_{\alpha=0.05}(df = 21)SE(\bar{X}) = 19.655 + 1.721 * 0.9609 = 21.3087$ (1.5 marks)

*Deduct 0.5 mark for wrong critical value (note table gives 1.72)*

Since $\mu_0 = 22$ is outside this upper bound, it cannot be 22 ng/mL and as such, it is consistent with the conclusion reached in part 'a' above. (0.5 mark)

c. **(3 marks)** Now test whether the median concentration of blood serum level in this population is different from 22 ng/mL.

$S1: H_0 : Md = 22$ $H_a : Md \neq 22$ (1 mark)

S5: [p_Val = 0.022] < [α = 0.05] ==> Reject $H_0$. (1.5 mark)

Based on the statistical evidence, one can claim that the population median is not 22 ng/mL.

(0.5 mark)

*Suggest 1 mark for hypotheses, 1 for p-value, 1 for rejection/conclusion*

d. **(2 marks)** Based on the evidence you have available, which test is more appropriate: the parametric or the non-parametric test? Explain briefly with reference to the relevant assumption for the t-test.

**Assumption:** For a small sample, the population must be normally distributed. (1 mark)

**Explanation:** In the boxplot, there are two outliers and as such the population cannot be considered normal. Therefore, a non-parametric Wilcoxon test on the population median is more appropriate. (1 mark)

**Question 2. [ 7 marks]**

a. Does the computer assisted program increase the proportion of students passing the examination? Use α = .05. Use the critical value approach.

[4] *(1 mark) H0: p1-p2 = 0 (or p1-p2 ≤ 0); Ha: p1-p2 > 0*

*1 mark for the decision rule :*
*We will reject H0 if Zcal, i.e. the test statistic is greater than $Z_{.05}$= 1.645.*

*Note: deduct .5 mark if student uses 1.96 as the critical value of Z*

*P1hat = 94/125 = .752 and p2hat = 113/175 = .646*
*Should calculate pooled proportion as p-bar = (94+113)/300 = 207/300 = 0.69*
*Thus the test statistic is: Zcal =( p1hat-p2hat)/ sqrt ((p-bar*q-bar)(1/n1 + 1/n2)*
*=(.752-.646)/ sqrt ((.69*.31)(1/125+1/175)) =0.106/0.0542 = 1.96 > $Z_{.05}$= 1.645.*

*0.5 mark for the standard error calculation using the pooled proportion and*
*0.5 mark for the test statistic*

*Conclusion: Reject Ho. There is enough evidence to support to claim that the computer assisted program is more effective than the traditional classroom sessions as it increases the proportion of students passing the examination.*
*1 mark. Deduct .5 mark if the student did not state the conclusion in managerial terms.*

b. For the test above, calculate an appropriate 95% confidence interval for the potential increase in the effectiveness of teaching English, for those in the computer assisted program over the traditional classroom session program.

[2] *The appropriate confidence interval is a right-sided interval. Thus we are interested in obtained the lower bound LB which is obtained as:*
*LB = .752-.646 – 1.645* sqrt ((.752*.248)/125) + (.646*.354)/175) = .106- 1.645*.053= .0188.*
- *Deduct .5 mark if the student estimated a two-sided interval;*
- *Give all the points if the student used 1.96 in (a) as well as in estimating the CI so long as it is a LB CI.*

c. Based on your answer in part b, can you conclude that the computer assisted program is more effective than the traditional classroom method? Explain briefly.

[1] *1 mark (Deduct .5 mark if the student did not express the conclusion in managerial terms)*
*Since the CI does not include zero, we can conclude with 95% confidence that the computer assisted program is more effective than the traditional classroom program.*

**Question 3. [8 marks ]**

a. At the 5% level of significance, test whether the percentage of students engaging in binge drinking at the university is greater than that found in the national survey.

[4] *(1 mark)* $H_0$ : *p = .44 or (p ≤ .44);* $H_a$ : *p > .44*

*The test statistic is Zcal =* $\frac{Phat - P0}{sqrt\ ((p0*q0)/n)}$

*1 mark: Decision: for Alpha = .05, one-sided test, Reject the Ho if Zcal > 1.645*

*Note: deduct .5 mark if student uses 1.96 as the critical value of Z*

*1 mark: Sqrt((*$p_0$*$q_0$*)/500) = sqrt ((.44*.56)/500) = .02219*

*So Zcal = Zcal =* $\frac{Phat - P0}{sqrt\ ((p*q)/n)}$ *=(.48-.44)/.02219 = 1.803 > 1.645, Reject H0.*

*Conclusion:*

*Reject the null hypothesis. There is enough evidence to conclude that the percentage of students at the university that participate in binge drinking is superior to the national percentage of 44%.*
*1 mark. Deduct .5 mark if the student did not state the conclusion in managerial terms.*

b. For the hypothesis test above, calculate an appropriate 95% confidence interval to estimate the proportion of binge drinkers at the university.

[2] *The appropriate confidence interval is right-sided and thus we are only interested in estimating the lower bound LB = .48-1.645*sqrt((.48*.52)/500)= .48-1.645*.0223=.48-.0365 =.444.* Deduct .5 mark if student estimated a two-sided interval.

c. To achieve a ± 3% margin of error, what sample size would be required to be 95% confident in the estimate?

[2] *n = Phat*Qhat* $(z / M)^2$ *= 0.48*0.52 ** $(1.96/0.03)^2$ ***= .48*.52*4266.67 = 1066***
*or for a n maximum, with p=.5, we get n = .5*.5*4266.67=1067*

*Thus, for a margin of error of ± 3%, we need to interview 1066 (1067 max) students if we want to be 95% confident in or estimates.*

*If a solution uses z = 1.645 for a one-sided 95% CI, p-hat =.48, q-hat =.52*

*n = Phat*Qhat* $(z / M)^2$ *= 0.48*0.52 ** $(1.645/0.03)^2$ *= .2496*3006.69= 751 students, there is a 0.5 deduction.*

*Give 1 mark for the appropriate formula and 1 mark for the calculation. Deduct 0.5 mark is calculation is wrong but formula is right.*

**Question 4. [9 marks ]**

a. If one wanted to compare the volatility of stocks in the energy and financial sectors using these data, should the analysis be done as an independent samples or a paired sample test? Explain briefly.

[1] *These are two samples of stocks from different sectors and no hint of any matching; therefore, the samples are independent.*

*If this answer is wrong, then student can get full marks on remaining parts if they are done properly from this wrong premise.*

b. Is a parametric test appropriate? Explain briefly, beginning with the distributional assumption(s) required for the parametric test and commenting on whether this(these) assumption(s) are warranted, with specific references to the appropriate boxplot(s).

[2] Distributional assumption(s): *With two large samples, we require the populations to be **not extremely skewed** for the sample means to be normally distributed. The populations do **not have** to be normally distributed.*

Comment: *Both boxplots of the data are relatively symmetric with a single (mild) outlier; therefore, these assumptions are warranted. The boxplot of differences is not relevant.*
*If student assumes **normality** of the two beta populations in part a, and proceeds to comment on the appropriateness of this assumption, then lose first mark but get the second. If part a specified matched samples, then this part must refer to boxplot of differences.*

c. Notwithstanding your answer to part b, but being consistent with your answer in part a, perform a parametric test to determine whether the volatility of energy stocks is higher than that of financial stocks, on average, by more than 0.25. Use a 5% significance level.

[4] *-Ho: mu(E) - mu(F) = 0.25 ; Ha: mu(E) - mu(F) > 0.25*

*-t = [(1.164 – 0.738) – 0.25] / sqrt(0.412^2/33 + 0.32^2/33) = 0.176 / 0.09 = 1.96; accept pooled variance; deduct 0.5 mark if 0.25 not subtracted unless zero diff specified in Ho.*

*-Reject Ho if t > 1.645 (using the normal approx. to t with 60 d.f.)*

*-Decide to reject Ho and conclude volatility of energy stocks exceeds that of financial stocks by more than 0.25. (0.5 mark each)*

d. Given your hypotheses above, calculate an appropriate confidence interval for the increase in volatility from financial to energy stocks.

[2] *At least 0.426 – 1.645* 0.09 = at least 0.426 - 0.148 or [ 0.278, ∞ )*

*Deduct 0.5 mark for wrong critical value or for 2-sided interval;*
*0 for UB (or LB) of 0.426 + 0.148; no need to note that the interval excludes 0.25.*