ADM 2304 -- ASSIGNMENT 1

Solutions and Marking Guide

1. [10 marks]

(a) The population mean is 25613.6 and the popn SD is 9523.7.

Note that the sample SD is 9529.3.

2 marks, 1 for each population value

(b)

Example of 20 sample means:

25384	25974	26730	24390	25866	22502	24285	25945	23063	24048
25400	25135	26224	25672	27332	26638	26515	25362	27605	25584

2 marks for showing the 20 sample means

(c)

For the above set of means, we have:

Descriptive Statistics: Means

```
Variable   Mean  StDev
Means     25483   1311
```

2 marks for showing the mean and standard deviation of the 20 sample means. Obviously every student will have a different set of summaries.

(d)

The result (mean) in part (c) is the mean of only 20 “observations” from the sampling distribution and we cannot expect that a sample mean based on 20 observations to be exactly equal to the population mean value of 25613.6 that it is estimating.

2 marks for an explanation that recognizes the result (sample mean) in part (c) is variable but is an estimate of the value 25613.6.

(e)

The sample standard deviation in part (c) is ________ which estimates the standard deviation of the sampling distribution which is sigma / sqrt(40). (in this example, the value in (c) is 1311.

Therefore the estimated population standard deviation in (a) is _____ * sqrt(40) = _____.
The solution should explain how different the estimate is from 9523.7.

In this example, the estimate is 1311*sqrt(40) = 8291, which is 1233 below the true value of 9524.

2 marks, 1 for the estimate based on the standard deviation in part c times sqrt(40) and 1 for noting how far it is from 9529.

2. [10 marks]

(a)

```
Test of p = 0.5 vs p > 0.5

                                     95%
                                   Lower
Sample     X    N  Sample p        Bound  Z-Value  P-Value
1        441  848  0.520047     0.491828     1.17    0.121
```

Ho: p = 0.5; Ha: p > 0.5
1-sided 95% CI is Lower Bound of 0.52 - 1.645*sqrt(0.52*0.48/848) = LB of 0.52 - 1.645*0.017 = LB of 0.52 - 0.028 = LB of 0.492
We do not reject the null H since the UB CI covers 0.50
We cannot conclude that a majority of Americans do not believe Trump's asseration

4 marks: 1 for hypotheses, 1 for CI, 1 for decision not to reject based on CI covering 0.50, 1 for the conclusion

(b) If we assume that 0.52 is the best estimate of p, then

we need the margin of error 1.645 sqrt(0.52*0.48/n) to be less than 0.02.

n must be greater than 0.52*0.48*1.645^2 / (0.02)^2 = 1688.56 or 1689

Since we do not reject the null H that p = 0.5, we could use n = 0.5*0.5*1.645^2/ 0.02^2 = 1691.3 or 1692.

2 marks for using the appropriate formula for finding n, with a deduction of 1 if 1.96 is used instead of 1.645.

(c)

Ho: p = 0.5; Ha: p > 0.5
We cannot use the normal approximation; therefore we go back to fundamentals and note that the p-value is $P(X \geq 13) = 1 - P(X \leq 12) = 1 - 0.9963 = 0.0037$ for a Binomial X with n=15 and p=0.5
We reject the null H since the p-value < 0.05 and conclude that the proportion … exceeds 50%.

```
Binomial with n = 15 and p = 0.5

 x   P( X <= x )
12      0.996307
```

- *4 marks, 1 for the hypotheses, 1 for recognizing the normal approximation does not apply, 1 for calculating the p-value using the binomial, 1 for the decision and conclusion.*

Note that Minitab will do the correct calculation of the p-value if we do not specify the normal distribution.
If the solutions uses the calculation of the z-statistic, please deduct the mark for not recognizing that the normal approximation does not apply. However, in this case, give the mark for the p-value of 0.002.

Test and CI for One Proportion

```
Test of p = 0.5 vs p > 0.5


                                  95%
                                Lower    Exact
Sample   X   N  Sample p        Bound  P-Value
1       13  15  0.866667     0.636558    0.004
```

Test and CI for One Proportion

```
Test of p = 0.5 vs p > 0.5


                                  95%
                                Lower
Sample   X   N  Sample p        Bound  Z-Value  P-Value
1       13  15  0.866667     0.722297     2.84    0.002


* NOTE * The normal approximation may be inaccurate for small samples.
```

3. [10 marks]

(a)

One-Sample T: salary

```
Test of mu = 50000 vs > 50000


                                                 95%
                                               Lower
Variable   N     Mean   StDev  SE Mean         Bound     T      P
salary    33  53243.0  9445.5   1644.2       50457.9  1.97  0.029
```

Ho: $\mu = 50000$ or $\mu \leq 50000$, Ha: $\mu > 50000$
T = (53243 – 50000) / [9445.5/sqrt(33)] = 3243 / 1644.2 = 1.9724
We reject the null H if T > 169 (based on t-distribution with 32 df)
We decide to reject the null H since 1.97 > 1.69 and we conclude the average starting salary exceeds $50K.

-4 marks: 1 for hypotheses, 1 for manual calculation of t-statistic, 1 for the decision rule or rejection region, 1 for decision/conclusion.

(b)

The p-value is Prob (t > 1.97) = 1 – Prob(t ≤ 1.97) = 1 – 0.971 = 0.029

Cumulative Distribution Function

```
Student's t distribution with 32 DF

   x  P( X <= x )
1.97     0.971228
```

1 mark for showing the p-value is Prob(t > 1.97)

(c)

For the > alternative hypothesis, we should calculate the "greater than" CI, which is

Lower Bound of 53243 – 1.69*1644.2 = LB of 53243 – 2779 = LB of 50464.

1 mark for the CI upper bound of 50464

(d) Explain why the p-value and confidence interval do or do not support your conclusion in part (a).

They do support the conclusion because the p-value of 0.029 is less than 0.05 and because the CI does not cover the hypothesized value of 50000.

-2 marks, 1 for each reason.

(e) Show a boxplot of the data and explain whether the assumptions of the test and confidence interval are reasonable.

The test includes the assumption that the population of salaries is not extremely skewed since we have a "large" sample.
This is a reasonable assumption since the boxplot shows that the sample is reasonably symmetric and certainly not too skewed, if at all skewed.

2 marks, 1 for stating the assumption that the population is not extremely skewed and 1 for explaining that the boxplot is consistent with that assumption.

Note: do not accept the assumption that the population data is normally distributed since this is not required.