BUS 310, Exam #1A, Fall 2017

Name KEY

Section

Instructions: This exam is in two parts: Part I is to be completed partly at home using the materials posted on Blackboard for Part I and you will answer questions about that work in class below; Part II is to be completed entirely in class. You may not use cell phones, and you may only access internet resources you are specifically directed to use. You may access your data file for Part I of the exam in Blackboard. You may access the data files posted to Blackboard for the Exam part II. Be sure you are using the data file that matches the exam version you are given.

Part I:

1. Use the information you calculated at home. Find a confidence interval for the mean from the students with undergraduate engineering majors. (8 points)

GMAT: (689, 744)
Salary (51068, 68652)
Expenses (1106, 2314)
Debt (-4055, 45335)

Age (29, 36)
Children (-.7, 1.5)

Your answers will vary

2. Use the information you calculated at home. Find a confidence interval for the difference of means for students age 25 vs. students aged 40 for their GMAT scores. Interpret the results of your interval. (15 points)

(-56, 20) we are 95% Confident that the difference in GMAT scores between ages 25 and 40 lies between -56, and 20 - which is to say there is no significant difference between them given our sample.

3. Record your χ^2 test of independence here. Clearly state the hypothesis, all key test statistics and the P-value. Interpret the results of the test in context. (12 points)

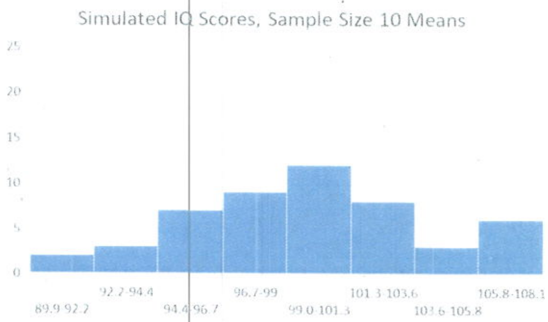
Ho: variables independent, Ha: variables dependent

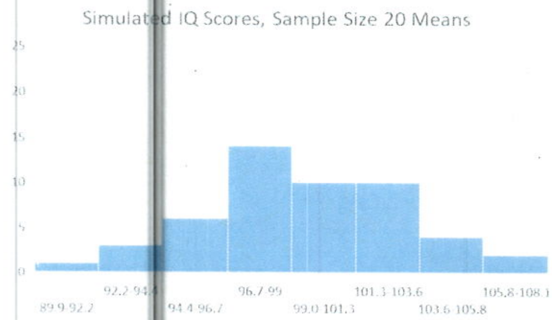
The χ^2 value is quite low, and the p-value nearly .9, Therefore we fail to reject the null and conclude marital status is independent of national origin.

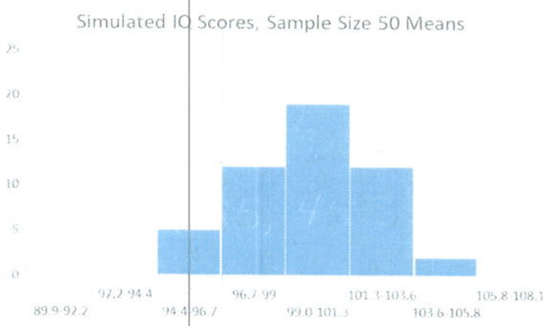
Calculations in Excel: (1) 32 points, (2) 24 points, (3) 24 points.

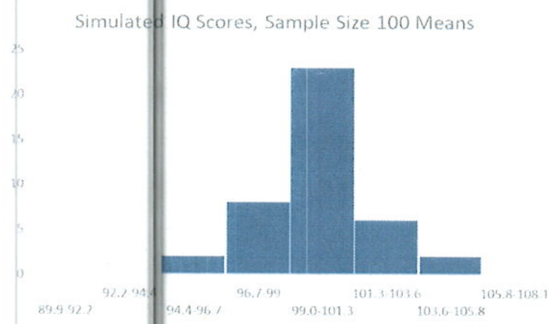
Part II:

4. Fifty (50) simulated samples of IQ scores are taken with each of 4 different sample sizes. Histograms of the means of the simulated data for each sample size are shown below, along with a table of summary statistics. Use this information to answer the questions that follow.









	Mean	Standard Deviation
Sample Size 10	98.7	4.708
Sample Size 20	100.2	3.198
Sample Size 50	100.4	2.481
Sample Size 100	100.2	1.529
Population	**100**	**15**

a. Describe what is happening to the histograms shown above as the sample sizes increase. (6 points)

The spread is shrinking and becoming more symmetric

b. The table shows the mean of the means from each sample size simulation, and the standard deviations of the means from 50 samples of each size. Calculate the standard error for a sample size of N=100 using the population values shown in the table. How does the simulated standard deviation compare to the value obtained from the simulation? (10 points)

very similar

Pop St. dev = 15

5. A 90% confidence interval for a population proportion is determined to be 0.64 to 0.75. If the confidence level is increased to 95%, and everything else remains the same, in what way will the confidence interval change? (6 points)

it will get wider

6. If the sample size increases and everything else remains the same, in what way will the confidence interval change? (6 points)

it will get narrower

7. Suppose that the alternative hypothesis is $H_a: \mu > 45$, is the hypothesis test one-tailed or two-tailed? (6 points)

one-tailed

8. Describe what a sampling frame is. (6 points)

it is the list of population members to be sampled

9. As the sample size increases, the t-distribution approaches what? (6 points)

the normal distribution

10. Suppose that a two-tailed test of a population proportion has a test-statistic of $z = -2.84$. Find the P-value. Use that information to determine whether the null hypothesis would be rejected at the 5% significance level. (10 points)

P-value $= .0045 < 5\%$

reject H_0

11. If the standard deviation of the lifetime of a vacuum cleaner is estimated to be 250 hours, how large of a sample, at minimum, must be taken to be 96% confident that the margin of error will not exceed 45 hours? (15 points)

n = 131

12. Give an example of a measurement error. Describe a situation in which it might occur and why it poses a problem for statistics. (6 points)

an instrument might be misreading
the data or a device used
improperly - answers wrong

13. Use the data in the data file for Exam #1 that matches your test. It contains data from a marketing company about the brand they market, and their competitor's brand. Find the proportion of the sample that uses "our brand". Find a 99% confidence interval and interpret the result in context. (20 points)

(45.5%, 61.7%)
We are 99% confident that between 45.5% & 61.7% of customers prefer our brand
- which is to say we can't tell which is more preferred from this data.

14. Use the data in the data file for Exam #1 that matches your test. It contains data from a sample of men and women matched for similar experience, age, education and other factors.
    a. Is this data paired or independent? (6 points)

paired

b. Calculate an appropriate hypothesis test for this data to determine if the difference between men's and women's salaries is 0 (they are the same). State the hypotheses, the test statistic and the P-value. What do you conclude about your hypotheses at the 5% level of significance for this data? (24 points)

Ho: $\delta = 0$ test-statistic $t = 7.07$

Ha: $\delta \neq 0$ P-value $= 2.57 \times 10^{-9} < 5\%$

reject Ho

c. Interpret a Type I error in this context. (6 points)

the difference between men's & women's salaries really is near 0, but we conclude they are different

d. Interpret a Type II error in this context. (6 points)

the difference between men's & women's Salaries is substantial, but we conclude it is not.

Upload your completed Excel files to the Exam #1 submission box in Blackboard, and submit your completed paper exam to your instructor. You may not modify anything once the exam is submitted.

Standard errors: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ $s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

$s_{x_1-x_2} = s_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

Sample sizes: $n > \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$ $n > \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$ $m = n = \frac{4z_{\alpha/2}^2(\sigma_1^2+\sigma_2^2)}{w^2}$

Confidence intervals:

One sample: $\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Two samples (independent): $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,n-1}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ $(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Test statistics:

One sample: $z \text{ or } t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ $z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$

Two samples: dependent: $z \text{ or } t = \frac{\bar{d}_0-\delta}{\frac{s_d}{\sqrt{n}}}$

Independent: $z \text{ or } t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ $z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}$

Degrees of freedom (two samples, unpooled) $\nu = \frac{\left(\frac{s_1^2}{m}+\frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1}+\frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}}$

χ^2Tests: $\chi^2 = \sum_{all\ cells}\frac{(obs-\exp)^2}{exp}$