Midterm 2

STAT-UB.0103 – Statistics for Business Control and Regression Models

The exam is closed book and notes, with the following exception: you are allowed to bring one letter-sized page of notes into the exam (front and back). You are also permitted use of a calculator. Each part of each problem is worth 5 points. There are 75 points total. There is no penalty for guessing incorrectly on a multiple choice problem. Partial credit may be awarded for all problems, as long as you show work.

For the problems involving calculations, you must show all work to get full credit. For short-answer problems, there should not be any symbols in your final answer (p, n, λ, etc.), but you do not need to fully simplify your answer. It is ok to have quantities like $\binom{5}{2}$, $e^{-3.1}$, etc. in your final answers on these problems.

NYU Stern Honor Code:

I will not lie, cheat or steal to gain an academic advantage, or tolerate those who do.

Signature: ______________________________ Date: ________________

Name: ____

Short Answer

1. (10 points) You want to estimate the average price of a dinner for all notable New York City restaurants. You take the 294 restaurants included in the Zagat guide to be an unbiased sample from the population of all such restaurants. For those restaurants listed in the guide, the average dinner price is \$36.56, and the standard deviation is \$14.77.

   (a) Construct a 95% confidence interval for the average dinner price of all notable New York City restaurants.

   (b) What is the interpretation of the confidence interval?

2. (20 points) Your consulting company wants to show that your clients earn higher profits after receiving your services. To show this, you select 50 past clients and record their annual profits the year before and the year after receiving your services. You compute the following statistics from the 50 sampled clients:

   - The mean and standard deviation of the "Before" profits (in millions of dollars) are 3.2 and 0.7.
   - The mean and standard deviation of the "After" profits (in millions of dollars) are 3.5 and 1.1.
   - The mean and standard deviation of the differences between "After" and "Before" profits for each compnay (in millions of dollars) are 0.3 and 0.86.

   You plan to perform a hypothesis test to demonstrate that your consulting services have an effect on profits. (Note: we will use a two-sided alternative in this problem, even though you probably want to demonstrate that your services have a *positive* effect.)

   (a) State the null and alternative hypotheses, using a two-sided alternative. Define any symbols you use (e.g., if you use a symbol, say θ, then define θ in the context of the problem.)

   (b) Compute an appropriate test statistic.

(c) Approximate the observed significance level (the p-value).

(d) In the context of the problem, explain in one or two sentences the meaning of the p-value you computed in part (c). Do not use the words "hypothesis," "null," "alternative," "significance," or "reject." The phrase "consulting services" should appear in your answer.

Multiple Choice

3. (5 points) You want to estimate the proportion of male Goldman Sachs employees who wear ties to work. To collect some data, you sit in the company's lobby in the morning and count the number of male employees wearing ties. You see 40 male employees with ties, and 10 male employees without. Based on this data, give an approximate 95% confidence interval for the proportion of all male Goldman Sachs employees who wear ties to work.

   A. $(.68, .91)$

   B. $(0.74, 0.86)$

   C. $(0.51, 1.08)$

   D. $(0.11, 0.39)$

   E. None of the above.

4. (5 points) Which of the following best describes why a 95% confidence interval for a population mean is more informative than the sample mean.

   A. The confidence interval gives the range of plausible values for μ.

   B. The expected value of the sample mean is equal to the population mean.

   C. The confidence interval involves a more sophisticated calculation than the sample mean does.

   D. The confidence interval contains the population mean.

   E. None of the above, because a confidence interval is *not* more informative than the sample mean.

5. (5 points) Which of the following is considered a Type II error?

   A. Failing to reject the null hypothesis when H_0 is true.

   B. Failing to reject the null hypothesis when H_0 is false.

   C. Rejecting the null hypothesis when H_0 is true.

   D. Rejecting the null hypothesis when H_0 is false.

   E. None of the above.

6. (5 points) Suppose that you are testing the null hypothesis $H_0 : \mu = 0$ against the alternative hypothesis $H_a : \mu \neq 0$, where μ is the population mean. You know the population standard deviation, σ. Let z denote the test statistic, and let p denote the p-value. Which of the following statements is true?

   A. If $z > -2$, then we must have that $p \geq .05$.

   B. If $z < 2$, then we must have that $p < .05$.

   C. If $z < -2$, then we must have that $p < .05$.

   D. If $z > 2$, then it is possible to have $p \geq .05$.

   E. None of the above.

7. (10 points) Suppose that we want to perform a hypothesis test on a population mean, μ. Specifically, we want to test the null hypothesis $H_0 : \mu = 2$ against the alternative $H_a : \mu \neq 2$. We collect a sample of size 60, observing a sample mean of 4.2 with a sample standard devation of 7.8.

   (a) The test statistic, t is equal to:

      A. 0.28
      B. 2.18
      C. 16.92
      D. 4.17
      E. None of the above.

   (b) Suppose that you want to perform the test at level 10%. Which of the following is correct rejection region (approximately)?

      A. Reject H_0 if $|t| > 2.000$
      B. Reject H_0 if $t > 1.296$
      C. Reject H_0 if $|t| > 1.671$
      D. Reject H_0 if $t < -1.296$
      E. Reject H_0 if $|t| > 1.296$

8. (5 points) There is a population with mean μ and standard deviation σ. You draw independent sample values $X_1, X_2, \ldots, X_n$ from this population, and compute the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Which of the following is true?

   A. $\mathrm{sd}(\bar{X}) = \sigma$ and $\mathrm{sd}(X_1) = \sigma$

   B. $\mathrm{sd}(\bar{X}) = \sigma$ and $\mathrm{sd}(X_1) = \sigma/\sqrt{n}$

   C. $\mathrm{sd}(\bar{X}) = \sigma/\sqrt{n}$ and $\mathrm{sd}(X_1) = \sigma$

   D. $\mathrm{sd}(\bar{X}) = \sigma/\sqrt{n}$ and $\mathrm{sd}(X_1) = \sigma/\sqrt{n}$

   E. Not enough information to determine.

9. (5 points) You draw a sample of size $n = 49$ from a population with mean $\mu = 2$ and standard deviation $\sigma = 14$. Let $\bar{X}$ denote the sample mean. Approximately what is $P(|\bar{X}| < 1)$?

   A. 38%

   B. 68%

   C. 24%

   D. 5.6%

   E. Not enough information to determine.

10. (5 points) Below is some Minitab output from a t-test of the null hypothesis $H_0 : \mu = 20$ against the alternative $H_a : \mu \neq 20$. Some of the values are missing.

```
Test of mu = 20 vs not = 20


 N    Mean  StDev  SE Mean       95% CI          T        P
??   22.10   8.10     ????  (?????, ?????)  ????  0.0332
```

Based on the output, you want to perform a hypothesis test at level 5%. What is the result of the test?

A. Reject H_0.

B. Reject H_a.

C. Do not reject H_0.

D. Do not reject H_a.

E. Not enough information to determine