Abantika Das Tonny

EXAM 1, DFSC5340.02

Fall 2020, SHSU

Name: Abantika Das Tonny Score: /240

Instruction: You may directly write your answers to the exam sheet. A total of 240 points are distributed over I and II questions. Additional 10 bonus points are given in III.

~~I. Multiple Choices (5 points each, a total of 100 points)~~

1. In a college class, the average IQ is 115. Assume that the distribution is normal and that the standard deviation is 15. What percentage of the class has an IQ between 105 and 130?
(Use a Z table, not provided. Please use the link in the chapter description or do a search.)

A. 28% B. 59% C. 54% D. 45%

B 59%

2. If z = 2 and -2, what area falls OUTSIDE these z values on a standard normal curve?

A. 0.081 B. 0.024 C. 0.00012 D. 0.046

D 0·046

3. James is a teacher in a small school is Boston. He is amused to find that his data is normally distributed, with 30 of his students taking a test which has 70 points. The average of their test scores is 36 and the standard deviation is 7. If the distribution is normal, towards which value is this symmetric data centered?

A. 30 B. 70 C. 36 D. 7

B 70

4. Which of the following is FALSE regarding the binomial probability distribution?

A. It is a mathematical construct that is used to model the probability of observing *r* successes in *n* trials.
B. It is the most often used discrete probability distribution in statistics.
C. It is calculated based on a formula where *p* is the significance level for a single event and *q* is the quantity of data present.
D. None of the answers are correct.

C

5. The mean score of a medical test is 72.21 with a standard deviation of 2.5. Find a 95% confidence interval for a random sample of 25 students with the following scores.

A. 71.18 to 73.24
B. 73.45 to 74.65
C. 72.25 to 72.30
D. 70.12 to 75.12

A

6. **A random sample is taken, and the sample size is 25. The sample is normally distributed, the sample mean is 89, and the standard deviation is 5.5. Find a 90% confidence interval for the population mean.**

6.

A. 88.56 to 89.16
B. 88.21 to 97.54
C. 78.69 to 98.34
D. 87.12 to 90.88

D

7. **What is the value of alpha for a 90% confidence interval?**

A. 0.01 B. 0.10 C. 0.25 D. 0.05

B

8. **When would you use the *t*-distribution procedure to find the confidence interval for the population mean?**

A. Only when you have the standard deviation and mean of a normally distributed population.
B. When you do not know the standard deviation of a normally distributed population.
C. When you are working a population that does not have a normal distribution.
D. When the only thing that you know about a population is its size.

B

9. **Which of the following is a way of increasing the power of a hypothesis test?**

A. Increasing the sample size.
B. Decreasing the sample size.
C. Using a stratified sampling technique.
D. Choosing always the 10% of the population.
E.

A

10. **It has been reported that the mean score for a student who takes the certain test is 80 with a standard deviation of 9. For a random sample of 100 students, what is the standard error?**

A. 0.8 B. 0.6 C. 0.5 D. 0.9

D

11. **A soccer coach wants to know how many hours per week his players spend training at home. He has 20 players and he decides to ask the first 4 players to arrive at the Monday's soccer practice how many hours they spend training per week. He then calculated that they spend an average of 10 hours per week. Therefore, he assumed that all the players train 10 hours per week. Is this an example of a simple random sample?**

A. No, because he didn't sample every soccer player.
B. No, because each student did not have an equal chance of being selected.
C. Yes, because each student had an equal chance of being selected.
D. Yes, the minimum number of students sampled needs to be four for it to be a simple random sample.

B

12. **A random sample of 10 items is taken from a normal population. The sample had a mean of 82 and a standard deviation is 26. Which is the appropriate 99% confidence interval for the population mean?**

A. $82 \pm z_{0.005}(26)$
B. $82 \pm t_{0.005}(26)$
C. $82 \pm z_{0.01} \frac{26}{\sqrt{10}}$
D. $82 \pm t_{0.005} \frac{26}{\sqrt{10}}$

E. None of the above

D

13. **A manufacturer of women's blouses has noticed that 80% of their blouses have no flaws, 15% of their blouses have one flaw, and 5% have two flaws. If you buy a new blouse from this manufacturer, the expected number of flaws will be**

A 0.15
B 0.20
C 0.80
D 1.00
E none of the above

E

14. **An inspector needs to learn if customers are getting fewer ounces of a soft drink than the 28 ounces stated on the label. After she collects data from a sample of bottles, she is going to conduct a test of a hypothesis. She should use**

A. a two tailed test.
B. a one tailed test with an alternative to the right.
C. a one tailed test with an alternative to the left.
D. either a one or a two tailed test because they are equivalent.
E. none of the above

C

15. **The manufacturer of Anthony Big's exercise equipment is interested in the relationship between the number of months (X) since the equipment was purchased by a customer and the number of hours (Y) the customer used the equipment last week. The result was the regression equation $Y = 12 - 0.5X$. The number 0.5 in the equation means that the average customer**

A. used the equipment for 30 minutes last week.
B. who has owned the equipment an extra month used the equipment 30 minutes less last week than the average customer who has owned it one month less.
C. who just bought the equipment used it 30 minutes last week.
D. bought the equipment one-half month ago.
E. none of the above

B

16. **A researcher is studying students in college in California. She takes a sample of 400 students from 10 colleges. The average age of all college students in California is**

A. a statistic.
B. a parameter.
C. the median.
D. a population.
E. none of the above

B

17. **A sample of 150 new cell phones produced by Yeskia found that 12 had cosmetic flaws. A 90% confidence interval for the proportion of all new Yeskia phones with cosmetic flaws is 0.044 to 0.116. Which statement below provides the correct interpretation of this confidence interval?**

A. There is a 90% chance that the proportion of new phones that have cosmetic flaws is between 0.044 and 0.116.
B. There is at least a 4.4% chance that a new phone will have a cosmetic flaw.
C. A sample of 150 phones will have no more than 11.6% with cosmetic flaws.
D. If you selected a very large number of samples and constructed a confidence interval for each, 90% of these intervals would include the proportion of all new phones with cosmetic flaws.
E. none of the above

D

18. **The standard deviation of a normal population is 10. You take a sample of 25 items from this population and compute a 95% confidence interval. In order to compute the confidence interval, you will use**

A. the t table because the degrees of freedom will be 24.
B. the t table because you have estimated the standard deviation from the sample.
C. the z table because the population standard deviation is known.
D. the z table because the sample size is small.
E. none of the above

C

19. **You are conducting a one-sided test of the null hypothesis that the population mean is 532 versus the alternative that the population mean is less than 532. If the sample mean is 529 and the p-value is 0.01, which of the following statements is true?**

A. There is a 0.01 probability that the population mean is smaller than 529.
B. The probability of observing a sample mean smaller than 529 when the population mean is 532 is 0.01.
C. There is a 0.01 probability that the population mean is smaller than 532.
D. If the significance level is 0.05, you will accept the null hypothesis.
E. none of the above

B

20. **The long-run average of a random variable is**

A. the expected value
B. the coefficient of determination
C. the standard deviation
D. the mode
E. none of the above

A

II. Problem Solving (140 points).

1. (20 points) The General Social Survey asks whether you agree or disagree with the following statement: "It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family." The sample proportion agreeing was 0.66 in 1977 and 0.31 in 2014 (n = 1655).

(a) Show that the estimated standard error in 2014 was 0.011.

As we know the sample proportion agreeing was ·31 in 2014.
The point estimation for the population proportion equals
$\hat{\pi} = \cdot 31$

Standard error of the point estimation in 2014

$SE = \sqrt{\frac{\cdot 31(1-\cdot 31)}{1655}} = 0\cdot 011$ (Showed)

[We Know standard deviation $\sigma = \sqrt{\pi(1-\pi)}$

(b) Construct the 95% confidence interval for 2014 and interpret it.

For 95% confidence level z value will be 1·96

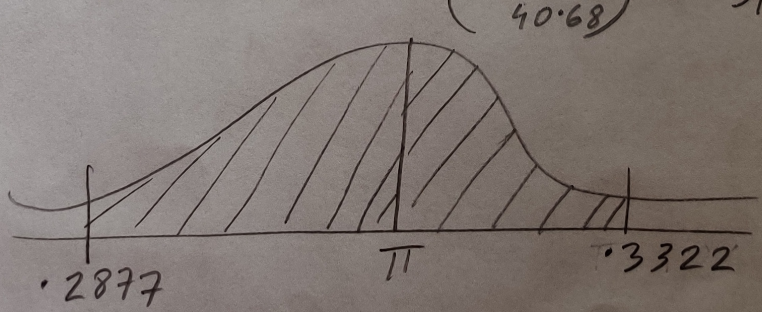
$\hat{\pi} = 0\cdot 31 = P(1)$ | $n = 1655$
$1-\hat{\pi} = \cdot 69 = P(0)$

Confidence level interval for 95% is $\hat{\pi} \pm 1\cdot 96\sqrt{\frac{\pi(1-\hat{\pi})}{n}}$

$= \cdot 31 \pm 1\cdot 96\sqrt{\frac{\cdot 31 \times \cdot 69}{1655}}$

Lower range $= \cdot 31 - 1\cdot 96\left(\frac{\cdot 4624}{40\cdot 68}\right) = \cdot 31 - \cdot 0222 = 0\cdot 2877$

higher range $= \cdot 31 + 1\cdot 96\left(\frac{\cdot 4624}{40\cdot 68}\right) = \cdot 31 + 0\cdot 0222 = 0\cdot 3322$



2. (20 points) For an exit poll of people who voted in a gubernatorial election, 40% voted for Jones and 60% for Smith. Assuming this is a random sample of all voters, construct a 99% confidence interval for the proportion of votes that Jones received, if the sample size was (a) 400, (b) 40. In each case, indicate whether you would be willing to predict the winner. Explain how and why the sample size affects the inference.

(a) $n = 400$, the point estimation of the population $\hat{\pi} = 0.40$

Standard error, $SE = \sqrt{\frac{.40(1-.40)}{400}} = 0.0245$

For 99% confidence interval, the probability $= \frac{1-.99}{2} = .005$

We know the table value at $\alpha = 0.01$ is 2.576

99% confidence interval $= .040 \pm (.0245)2.576$

$= (.34, .46)$ The interval is below 0.5, so we can predict Jones would lose the election.

(b) $n = 40$, the point estimation of the population $\hat{\pi} = 0.40$

Standard Error, $SE = \sqrt{\frac{.40 \times (1-.40)}{40}} = 0.0774$

For 99% confidence interval, Z value is 2.58

99% confidence interval $= .40 \pm (2.58)(0.0774) = (.20, .60)$

The interval is 20% to 60%. So, Jones has higher chance of winning the election.

3. (20 points) Find and interpret the P-value for testingH0: μ = 100 against H0: μ = 100 if a sample has (a) n = 400, y = 103, and s = 40. (b) n = 1600, y¯= 103, and s= 40. Comment on the effect of n on the results of a significance test

(a) $n = 400, \bar{y} = 103, s = 40$

$\mu = 100$

$se = \frac{40}{\sqrt{400}} = 2$

$t = \frac{103-100}{2} = 1.5$

For two tailed p-value

$P(T < -1.50 \text{ or } T > 1.50)$

$= P(Z < -1.50 \text{ or } Z > 1.50)$

$= 1 - P(Z < 1.50)$

$= 1 - 0.9332$

$= 0.0668$

For two tailed $= 2 \times 0.0668$

$= 0.1336$

P-value 0.1336

(b) $n = 1600, \bar{y} = 103, s = 40$

$se = \frac{s}{\sqrt{n}} = \frac{40}{\sqrt{1600}} = 1$

$t = \frac{\bar{y} - \mu_0}{n} = \frac{103-100}{1} = 3$

For two tailed p-value:

$P(t < -3 \text{ or } t > 3)$

$= P(Z < -3 \text{ or } Z > 3)$

$= 1 - P(Z < 3)$

$= 1 - .9987$

$= 0.0013 = 2 \times 0.0013 = 0.0026$

3a, 3b the more population, the less probability. Small P strong evidence against Ho.

4. (20 points) For a test of H0: $\pi = 0.50$, the sample proportion is 0.35 with n = 100.

(a) Show that the test statistic is z =-3.0.

Here, $\hat{\pi} = .35$, $n = 100$, $H_0: \pi = 0.50$

$$Z = \frac{\hat{\pi} - \pi_0}{SE_0} = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{.35 - .50}{\sqrt{\frac{.50(.50)}{100}}} = -3$$

(Showed)

(b) Find and interpret the P-value for Ha: $\pi < 0.50$.

We know for P value for Ha: $\pi < 0.50$ is:

$P(Z < -3) = 0.00135$

So we can get if the population proportion were .50 then the probability is .00135. The P value will be at .35 or below of .35.

(c) For a significance level of $\alpha = 0.05$, what decision do you make?

The P value is small. Ho is false here.

The population proportion is less than 0.50.

(d) If the decision in (c) was in error, what type of error was it? What could you do to reduce the chance of that type of error?

It was Type 1 error.

Type 1 error occurs when Ho rejected.

We have to reduce the significance level so that we can reduce the chance of that type of error.

5. (20 points) Jones and Smith separately conduct studies to test H0: μ = 500 against H0: μ = 500, each with n = 1000. Jones gets ȳ = 519.5, with se = 10.0. Smith gets ȳ = 519.7, with se = 10.0.

(a) Show that t = 1.95 and P-value = 0.051 for Jones. Show that t = 1.97 and P- value = 0.049 for Smith.

For Jones,

test statistics $= \frac{\bar{y} - \mu_0}{se} = \frac{519.5 - 500}{10}$

$= 1.95$ (Showed)

P-value $= P(T < -1.95 \text{ or } T > 1.95)$

$= P(Z < -1.95 \text{ or } Z > 1.95)$

$= 2 \times 0.0256$

$= 0.0512$

(Showed)

For Smith,

test statistics, $\frac{\bar{y} - \mu_0}{se} = \frac{519.7 - 500}{10}$

$= 1.97$ (Showed)

P-value $= P(T < -1.97 \text{ or } T > 1.97)$

$= P(Z < -1.97 \text{ or } T > 1.97)$

$= 2 \times 0.0244$

$= 0.049$

(Showed)

(b) Using α = 0.050, for each study indicate whether the result is "statistically significant."

Here For Jones P-value is larger than 0.05 so it does not contradict Ho. Jone's result is not statistically significant

Here, For Smith P-value is smaller than 0.05, so it does contradict Ho. Smith's result is statistically significant.

(c) Using this example, explain the misleading aspects of reporting the result of a test as "P ≤ 0.05" versus "P > 0.05,"oras"reject H0" versus "Do not reject H0," without reporting the actual P-value.

Jones's probability 0.0512, Smith's probability 0.049

If $P < \alpha$ = we can reject Ho

$P > \alpha$ = we can not reject Ho

Lets consider $\alpha = 0.05$

For Jones:

$P = 0.0512$, $\alpha = 0.05$. Here $P > \alpha$ so we can not reject Ho for Jones

For Smith:

$P = 0.049$, $\alpha = 0.05$. Here, $P < \alpha$, so we can reject Ho for Smith.

6. (20 points) In Great Britain, the Time Use Survey17 studied how a random sample of Brits spend their time on a typical day. For those who reported working full time, Table 7.18 reports the mean and standard deviation of the reported average number of minutes per day spent on cooking and washing up.

TABLE

		Cooking and Washing Up Minutes	
Sex	Sample Size	Mean	Standard Deviation
Men	1219	23	32
Women	733	37	16

(a) Estimate the difference between the means for women and men and find its standard error.

Difference between the means for women and men:

$\bar{y}_2 - \bar{y}_1 = 37-23 = 14$

Standard Error $= \sqrt{\frac{(16)^2}{733} + \frac{(32)^2}{1219}}$

$= \sqrt{1.1893}$

≈ 1.090 (Answer)

(b) Compare the population means using a two-sided significance test. Interpret the P-value.

Here, $SE = 1.090$, $\bar{y}_2-\bar{y}_1 = 14$, Z-score for 95% Confidence level 1.96

Test static $= \frac{(\bar{y}_2-\bar{y}_1)-0}{1.090} = \frac{14}{1.090} = 12.844$

We Know, $\mu_2-\mu_1 = (\bar{y}_2-\bar{y}_1) \pm t(SE)$

$= 14 \pm 1.96 \times 1.090$

$= (11.8636, 16.1364)$

Here, t value is large and P value is 0 to 0.00001. So we can say the population means differ. We can reject H_0 at $\alpha = 0.05$ level for two sided significance test. Result is significant as $P < 0.05$.

7. (20 points) In a study of the effect of the compound tomoxetine as a treatment for adult attention deficit hyperactivity disorder (ADHD), the 21 subjects had an ADHD rating scale mean of 30.0 ($s = 6.7$) at baseline and 21.5 ($s = 10.1$) after three weeks of treatment. The standard deviation was 9.84 for the 21 changes in rating. The authors reported a paired t statistic of 3.96 with df = 20.

Show how the authors constructed the t statistic, and report and interpret the P- value for a two-sided test.

Here, standard deviation 6·7, 10·1 | $n = 21$
mean = 30, 21·5
statistic value 3·96
DF 20

$$\text{Standard Error} = \frac{sd}{\sqrt{n}} = \frac{9 \cdot 84}{\sqrt{21}}.$$

$$= 2 \cdot 14$$

The test statistic,

$$t = \frac{\bar{x}_1 - \bar{y}_2}{se}$$

$$= \frac{30 - 21 \cdot 5}{2 \cdot 15}$$

$$= 3 \cdot 95$$

If we check Excel's TDIST (3·96, 20·2)

p- value = ~~0·001~~ <u>0·000075</u> which is less than 0·05, so we can say the report is significant at $P < 0 \cdot 05$.

There is strong evidence that tomoxetine is a effective treatment after 3 weeks. Here,

<u>We can say that Tomoxetine is effective.</u>

III. 1. (Extra Credit, 5 points). Which aspect(s) have you gained mostly from this class?

(a) Fundamental knowledge and concepts in statistical methods.
(b) Principles and Application skills by using Program to solve statistical problems
(c) None; or
(d) Others

Please describe and justify your answer.

(a) and (b)

This course helped me to grasp very complex statiscal subjects. This course's assignments are valuable emphasis on practical application. This course made things clearer and easier for me. I feel like I "Know" Statistics and its application.

2. (Extra Credit, 5 points). Do you have any suggestions on this class for the future teaching? Please describe your suggestions.

I like doing this course online. All lectures, assignments are straight forward. I like the flexibility of the course. The assignments are very beneficial to the entire learning process.

I dont have any suggestion on this class for now. We have a project to do. So, it would be less stressful if we dont have to take any final exam for this course.