Solutions

# Midterm 2 - STAT 301
## Spring 2015

Name:

UIN:

**Signature:**

**Version A:**

1. Do not open this test until told to do so.
2. This is a closed book examination, However you may use the cheat sheet provided to you, the $t$ and $z$ tables. You should have no other printed or written material with you on the exam. But scrap paper is allowed.
3. You have 60 minutes to work on this exam. There are 15 multiple choice questions.
4. On the scantron please state the version of exam that you have.
5. You may use a calculator in the exam.
6. If there is no correct answer or if multiple answers are correct, select the **best** answer.
7. Unless stated otherwise, do all tests at the 5% level.
8. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification (however we are limited in the amount of help we can offer).
9. Please only give one answer per question (the one that is closest to the solution).
10. Good Luck!!!

(1-3) Female heights are known to be normally distributed with mean $\mu = 64.5$ and standard deviation $\sigma = 2.5$ inches. I randomly sample **two** females.

(1) What is the mean and standard error of their average height (the sample mean)?

(A) $\mu = 64.5$, se $= \frac{2.5}{2}$ (B) $\mu$ is unknown and se $= \frac{2.5}{2}$ (C) $\mu = \frac{64.5}{2}$ and se $= \frac{2.5}{2}$
(D) $\mu = 64.5$, se $= \frac{2.5}{\sqrt{2}}$ (E) $\mu$ is unknown and se is unknown.

(2) What is the probability that their average (sample mean) height will be **less** than 67 inches?

$z = \frac{67-64.5}{\left(\frac{2.5}{\sqrt{2}}\right)} = 1.41$

(A) 1.4% (B) 98.6% (C) 8% (D) 92% (E) 84%

(3) What is the probability that their **total** height will be **less** than 130 inches?

(A) 53% (B) 55% (C) 57.9% (D) 61% (E) 63%

$z = \frac{65-64.5}{\left(2.5/\sqrt{2}\right)} = 0.28$

(4) A high-tech company wants to know how many portable devices people own. They intend to draw a random sample from the general population and use the data to construct a 95% confidence interval for the mean number of devices. It is believed that the standard deviation is between $1-3$. What is the minimum number of people required to be surveyed such that the margin of error is at **most** 0.25?

(A) 62 (B) 246 (C) 387 (D) 554 (E) 652.

(5) The 99% confidence interval for the mean height of a person (using the normal distribution) is [65, 66] inches (margin of error is 0.5). Which of the following methods reduces the margin of error to 0.25?

(A) Increase sample size by factor 4.

(B) Increase sample size by factor 2.

(C) Reduce the level of confidence from 99% to 95%

(D) (A) and (C).

(E) (B) and (C).

(6) A news outlet randomly samples 200 people from the Houston Area. They ask each person sampled whether they approved of Obama. 65 of the 200 said they approved, whereas 135 of the 200 disapproved. Based on the fact that 32.5% of those surveyed voiced approval, the news outlet makes the claim that less than 35% of the country approve of Obama. Using this information and the statistical test in Figure ?? comment of this statement.

$H_0: p \geq 0.35$

**Hypothesis test results:**
Outcomes in : Approve
Success : 1
p : Proportion of successes
$H_0 : p \geq 0.35$
$H_A : p < 0.35$

| Variable | Count | Total | Sample Prop. | Std. Err. | Z-Stat | P-value |
|---|---|---|---|---|---|---|
| Approve | 65 | 200 | 0.325 | 0.033726844 | -0.74124932 | 0.2293 |

22% pretty large!

Figure 1: Obama approval data

1/2 (A) The sample is biased so conclusions about the entire US population cannot be drawn from it.

(B) From the population from which the sample was drawn, there is evidence to suggest that Obama's approval rating is below 35%.

1/2 (C) From the population from which the sample was drawn, there is no evidence to suggest that Obama's approval rating is below 35%.

(D) (A) and (B)

1 (E) (A) and (C)

(7-9) A survey is conducted to estimate the average the number of people in a household below the age of 18 (defined as child). 15 households are randomly sampled and asked the number of children in that house. The summary statistics based on the data and a QQplot of the data is given below.

**Summary statistics:**

| Column | n | Mean | Variance | Std. dev. | Std. err. | Median |
|---|---|---|---|---|---|---|
| siblings | 15 | 2.6 | 4.2571429 | 2.0632845 | 0.53273776 | 2 |

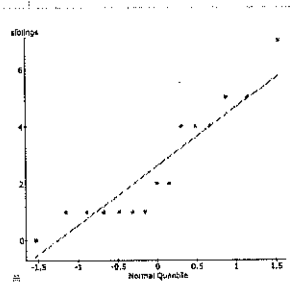


Figure 2: Sibling data

(7) Construct a 95% confidence interval for the mean number of children in a household (remember to use the t-distribution with 14df).

(A) $[2.6 \pm 2.145 \times 0.53]$ (B) $[2 \pm 1.761 \times 0.53]$ (C) $[4.2 \pm 2.145 \times 2.06]$

(D) Since the number of children is integer valued, the mean has to be integer values. Therefore the CI has no meaning.

(E) $[2.6 \pm 1.761 \times 2.06]$

(8) 60 years ago the national census showed that the number of children (under 18 year olds) in a household was 4. It is believed that the average number of children has decreased. Using the data collected, state the hypothesis of interest. Is there any evidence to back this claim (do the test at the 5% level)?

$t = \frac{2.6-4}{0.53}$

$= -2.628$

(A) $H_0 : \mu \geq 4$ vs $H_A : \mu < 4$. $t = -2.628$, the p-value is less than 2% thus there is evidence that the mean number of children has decreased.

(B) $H_0 : \mu \geq 4$ vs $H_A : \mu < 4$. $t = 2.628$, the p-value is greater than 98% thus there no evidence that the mean number of children has decreased.

(C) $H_0 : \mu \geq 2.6$ vs $H_A : \mu < 2.6$. $t = 2.628$, the p-value is greater than 98% thus there no evidence that the mean number of children has decreased.

(D) $H_0 : \mu \geq 4$ vs $H_A : \mu < 4$. $t = -0.68$, the p-value is greater than 5% thus there no evidence that the mean number of children has decreased.

(E) $H_0 : \mu \leq 4$ vs $H_A : \mu > 4$. $t = 2.628$, the p-value is less than 2% thus there actually evidence that the mean number of children has increased.

(9) Based on the data and the QQplot, comment on the reliablity of the confidence interval (do we really have 95% confidence in it?) and p-values obtained in (7) and (8).

(A) The data is not normally distributed, therefore with such a small sample size neither the CI or p-value are reliable.

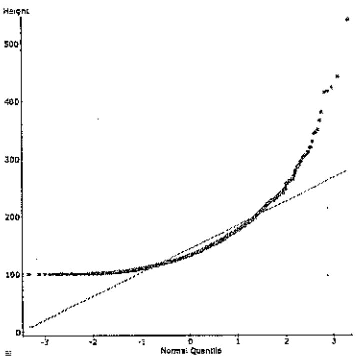
(B) The data is not normally distributed, however, by using the t-distribution we have corrected for the lack of non-normality, thus giving a reliable CI and p-value.

(C) The data is close to normal, thus both the CI and the p-values obtained are reliable.

(D) The central limit theorem means that the data will become normal for large sample sizes.

(E) (B) and (D)

(10-11) 2105 skyscrappers are randomly sampled. The QQplot of the heights of all these skyscrappers and a confidence interval is given in Figure ??.




**95% confidence interval results:**

μ : Mean of variable

| Variable | Sample Mean | Std. Err. | DF | L. Limit | U. Limit |
|---|---|---|---|---|---|
| Height | 143.69905 | 0.89919143 | 2104 | 141.93565 | 145.46245 |

Figure 3: Skyscrapper data

C.I = [141.93, 145.46]

(10) Which statement(s) are correct?

(A) The data is highly RIGHT skewed, however as the sample size is large we really do have 95% confidence the mean is in this interval.

(B) Despite the non-normality of the data, thanks to the central limit theorem, 95% of the data will lie in the interval [141.9, 145.46].

(C) The data is highly skewed, however if we collect a large enough sample size the data will become more normal.

(D) Two of the above. (E) None of the above.

(11) What are the conclusion of the following tests for the mean height of a skyscrapper (at the 5% level).

| | $H_0 : \mu \leq 140$ vs $H_A : \mu > 140$ | $H_0 : \mu = 140$ vs $H_A : \mu \neq 140$ | $H_0 : \mu \geq 140$ vs $H_A : \mu < 140$ |
|---|---|---|---|
| A | p-value > 97.5%, cannot reject null | p-value < 5% reject null | p-value < 2.5% reject null. |
| B | p-value < 2.5%, reject null | p-value < 5% reject null | p-value > 97.5% cannot reject null. |
| C | p-value > 5%, cannot reject null | p-value > 5% cannot reject null | p-value > 5% cannot reject null. |
| D | p-value > 97.5%, reject null | p-value < 5% cannot reject null | p-value < 2.5% cannot reject null. |
| E | p-value < 2.5%, cannot reject null | p-value < 5% cannot reject null | p-value > 97.5% reject null. |

$\bar{x} = 143.7 > 140$ 140 not in 95% CI

(12-13) It is notoriously difficult to assess the cognitive ability of babies (below the age of one). In a given assessment, child psychologists need to discriminate betweem random behaviour and real awareness in the baby.

10 months old babies are tested to see whether they can match socks. This is done by psychologist holding a sock in front of the baby. The baby is 'asked' (a struggle) to select the matching sock from a pile of different socks. The experiment is done five times and the chance of the baby obtaining the results by random chance is calculated. The tests are always done at the 5% significance level.

(12) A 10 month old baby is 'tested' to see whether she can match socks. In the test the baby matches 4 out of 5 socks correctly. The probability of her randomly matching 4 out of 5 socks is 0.672%. Based on this information state the hypothesis of interest to the psychologists and the result of the test (at the 5% level).

(A) $H_0$ :Baby was not guessing vs $H_A$: Baby was guessing. The p-value is less than 5% we reject the null and determine the baby is guessing.

(B) $H_0$ :Baby was guessing vs $H_A$: Baby was not guessing. The p-value is greater than 5% we cannot reject the null.

(C) $H_0$ :Baby was guessing vs $H_A$: Baby was not guessing. The p-value is less than 5% we reject the null and determine the baby is able to match socks.

(D) This is an example of biased sampling. (E) (B) and (D)

(13) Suppose 100 babies with no matching ability are tested. On average how many of these will falsely give the impression they have some sort of ability?

(A) 5 (B) 95 (C) 10 (D) 90 (E) 0.

(14) An expectant mother is diagnosed with gestational diabetes if her mean glucose level is over 140. Each time a blood sample is taken the glucose level will vary, however it is known that for all pregnant women the standard deviation is $\sigma = 2$. 4 blood samples from a expectant mother is taken, they are 138.62, 133.38, 140.55, 139.52. The sample mean is $\bar{x} = 138.0$. The obstetrician wants to see of there is any evidence of gestational diabetes in the expectant mother, based on her blood samples. What is the hypothesis of interest, and the result of the test at the 5% level.

(A) $H_0 : \mu \leq 140$ vs $H_A : \mu > 140$. The p-value is 2.3% there is evidence to suggest she has gestational diabetes.

(B) $H_0 : \mu \leq 140$ vs $H_A : \mu > 140$. The p-value is 97.7% there is no evidence to suggest she has gestational diabetes.

$\bar{x} = 138$ It is in the null

(C) $H_0 : \mu \geq 140$ vs $H_A : \mu < 140$. The p-value is 2.3% there is evidence to suggest she has gestational diabetes.

(D) $H_0 : \mu = 140$ vs $H_A : \mu \neq 140$. The p-value is 4.6%, there is evidence to suggest she has gestational diabetes.

(E) $H_0 : \mu \leq 138$ vs $H_A : \mu > 138$. The p-value is 50% there is no evidence to suggest she has gestational diabetes.

(15) The obstetrician wants to increase the detection of gestational diabetes. Which of these options will increase the detection rate?

(A) Decrease the signifiance level from 5% to 1%

$\sqrt{}/2$ (B) Increase the significance level from 5% to 10%

$\sqrt{}/2$ (C) Increase the number of blood samples taken.

(D) (A) and (C)

(E) (B) and (C).