**Final - STAT 301**
**Fall 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 cheatsheet provided. You should have no other printed or written material with you on the exam. But scrap paper is allowed.
3. You have 2 hours to work on this exam. There are 25 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. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification.
8. Please only give one answer per question (the one that is closest to the solution).
9. No wearing hats that can cover ones eyes.
10. Good Luck. Have a wonderful winter break, it was lovely to teach you.

(1) Which statement is correct?

(A) If the standard deviation of a data set is zero, all the values in the data set must be **the same**.

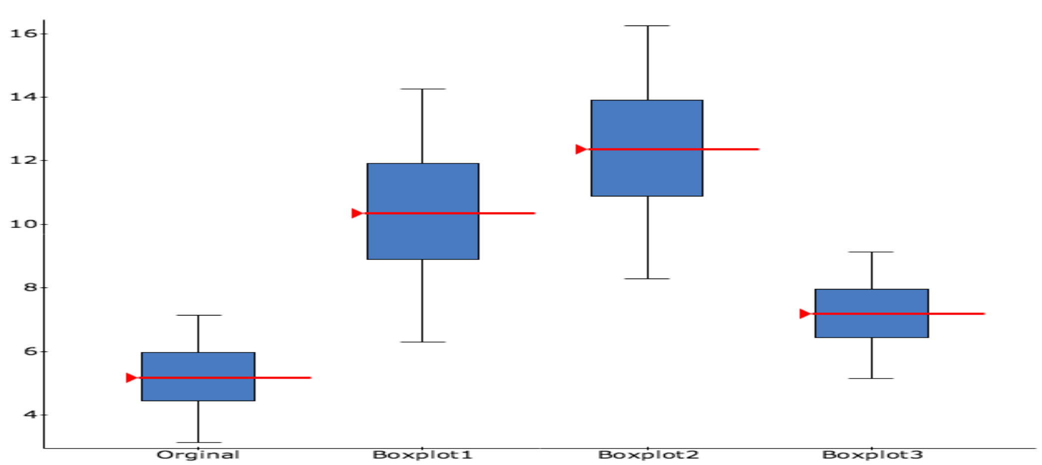
(B) If the standard deviation of a data set is zero, all the values in the data set must be **zero**.

(C) If the interquartile range in a data set is zero, at least 50% of the values in the data set must be the same.

(D) [B] and [C] (E) [A] and [C].

(2) Box plot of four data sets. The first (far left) is the original data, and the next three boxplots are transformations of the original data. The three transformations are
(i) 2×original data +1 (ii) original data +2 (iii) 2×original data.

Match the boxplot to the transformation.




| | Boxplot1 | Boxplot2 | Boxplot3 |
|---|---|---|---|
| (A) | original + 2 | 2×original | 2×original +2 |
| (B) | 2×original data +2 | 2×original | original +2 |
| (C) | 2×original +2 | original + 2 | 2×original |
| (D) | 2×original | 2×original +2 | original +2 |
| (E) | original + 2 | 2×original +2 | 2×original |

(3) Suppose the heights of horses are normally distributed. What proportion of horse heights will be within 0.5 standard deviations of the mean?

(A) 64% (B) 34% (C) 68% (D) 38% (E) 95%.

(4) Match the study to the data:

1. 1000 people were surveyed on their diet and health.
2. After a heart attack, 1000 people (who had a heart attack) were divided into two groups. 500 were placed on a Mediterranean diet and there other 500 were placed an a diet recommended by the American heart association. There health status was followed up over a number of years.
3. To understand the effect that temperature may be have on metabolism, 32 baby mice were divided into two groups. 16 mice spent three months in cold conditions and the other 16 were kept in warm conditions. Their weight and diversity of bacteria in their stomach was measured after 3 months.

| | 1 | 2 | 3 |
|---|---|---|---|
| (A) | Experimental | Observational | Observational |
| (B) | Observational | Experimental | Experimental |
| (C) | Observational | Anecdotal | Experimental |
| (D) | Observational | Experimental | Anecdotal |
| (E) | Anecdotal | Experimental | Observational |

(5) 15 people were placed on a diet. After one week the average weight loss was 1.15 kgs. We test the hypothesis that the diet results in weight loss (at least for the first week). That is $H_0 : \mu \leq 0$ vs. $H_A : \mu > 0$. A summary of the data (in kilos), and test result is given below.

The data is converted into pounds; this means all the weights are multiplied by 2 (doubled). What happens to the mean, standard deviation, T-value and p-value?

**Summary statistics:**

| Column | Mean | Variance | Std. dev. | Std. err. |
|---|---|---|---|---|
| Diet | 1.1527224 | 0.71669146 | 0.84657631 | 0.21858506 |

**Hypothesis test results:**
μ : Mean of variable
$H_0 : \mu = 0$
$H_A : \mu > 0$

| Variable | Sample Mean | Std. Err. | DF | T-Stat | P-value |
|---|---|---|---|---|---|
| Diet | 1.1527224 | 0.21858506 | 14 | 5.2735643 | <0.0001 |

| | sample mean | Std. dev. | T-stat | P-value |
|---|---|---|---|---|
| (A) | stays same | stays same | double | double |
| (B) | double | double | double | double |
| (C) | stays same | stays same | stays same | stays same |
| (D) | stay same | double | stay sample | double |
| (E) | double | double | stays same | stays same |

(6) The length of mice is known to be normally distributed with mean **2** inches and standard deviation **0.25** inches.

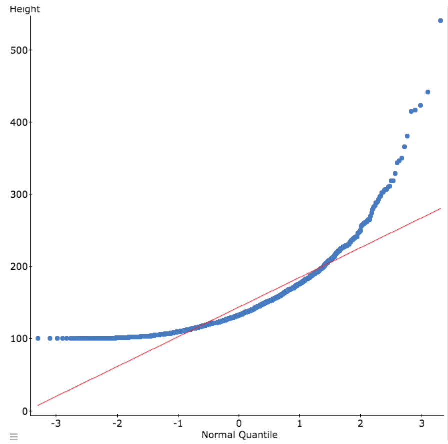
The height of human males is known to be normally distributed with mean **71** inches and standard deviation **3** inches.

Oscar the mouse is **2.5** inches long. Using the equivalent percentile, how tall would Oscar be if he were a human male?

(A) 65 inches (B) 71.5 inches (C) 73 inches (D) 73.5 inches (E) 77 inches.

(7) Below we give a QQplot of skyscrappers heights sampled from across the world.

Which statement(s) on the distribution of skyscrappers is correct?




(A) The distribution is right skewed.

(B) The distribution will become more normal as we increase the sample size.

(C) The distribution is heavy tailed.

(D) [A] and [B].

(E) [B] and [C].

(8-9) The length of monkeys from feet to shoulders is known to be normally distributed with mean $\mu = 30$ inches and standard deviation $\sigma = 3$ inches. Use this information to answer the following two questions.

(8) **Four** monkeys are randomly sampled and the average length from foot to shoulder is calculated. What is the population mean and standard error of the sample mean (average based on four)?

(A) $\mu = \frac{30}{\sqrt{4}}$, se $= \frac{3}{\sqrt{4}}$, (B) $\mu = 30$, se $= \frac{3}{\sqrt{4}}$, (C) $\mu = 30$, se $= \frac{3}{4}$, (D) $\mu = 30$, se$= 3$, (E) $\mu = \frac{30}{4}$, se $= \frac{3}{\sqrt{4}}$.

(9) A cookie jar is on a shelf which is **132** inches from the ground. **Four** monkeys decide to stand on each others shoulders to reach the cookie jar.

What is the chance the top monkey's shoulder reaches the shelf (thus reaching the cookie jar)?

(A) 2.28% (B) 3% (C) 12.58% (D) 49.6% (E) 97.7%.

(10) A pregnant person is said to have gestational diabetes if their mean level glucose level is **over** 140. To detect gestational diabetes, four blood samples are taken and the sample average is evaluated. We test the hypothesis $H_0 : \mu \leq 140$ vs $H_A : \mu > 140$. Below we give the mean level of 5 different ladies. For which mean level are we most likely to reject the null (most power)?

(A) $\mu = 125$ (B) $\mu = 130$ (C) $\mu = 140$ (D) $\mu = 150$ (E) $\mu = 155$.

(11-13) Students are investigating the **number of pets A&M students** have.

(11) In one project, **10** students are random sampled and asked how many pets they have. Below we summarize the results.

**Summary statistics:**

| Column | n | Mean | Variance | Std. dev. | Std. err. | Median | Range | Min | Max |
|---|---|---|---|---|---|---|---|---|---|
| pets | 10 | 1.8 | 1.9555556 | 1.3984118 | 0.44221664 | 1 | 4 | 0 | 4 |

Using the output, construct a 95% confidence interval for the mean number of pets (remember to use the t-distribution).

(A) $[1.8 \pm 1.383 \times 0.44]$ (B) $[1.8 \pm 1.383 \times 0.44]$ (C) $[1.8 \pm 2.262 \times 0.44]$

(D) $[1.8 \pm 1.383 \times 1.39]$ (E) $[1.8 \pm 2.262 \times 1.39]$.

(12) Another group asked **101** students for the number of pets they owned. A confidence interval based on this data set is given below.

**95% confidence interval results:**

μ : Mean of variable

| Variable | Sample Mean | Std. Err. | DF | L. Limit | U. Limit |
|---|---|---|---|---|---|
| Pets2 | 1.990099 | 0.1499142 | 100 | 1.6926735 | 2.2875245 |

Figure 1: Sample size of 101 students

When comparing the confidence interval from (Q11) with the confidence interval above which statement is correct?

(A) The pet data **is not** normal, therefore we cannot have 95% confidence in the interval constructed in (Q11).

(B) The pet data **is** normal, therefore we have 95% confidence in the interval constructed in (Q11).

(C) The pet data is not normal, but given the sample size of 101, we have close to 95% confidence in the CI given in (Q12).

(D) [A] and [C]. (E) [B] and [C].

(13) We want to see if there is any evidence that on average students have more than 2.5 pets? What is the hypothesis of interest? Use the output in Figure 1 to test the hypothesis. Give the $t-$value and p-value.

(A) $H_0 : \mu \leq 2.5$ vs $H_A : \mu > 2.5$, $t = -4.7$ and p-value is greater than 50%. There is **no evidence** to suggest the mean number of pets is greater than 2.5.

(B) $H_0 : \mu \leq 2.5$ vs $H_A : \mu > 2.5$, $t = -4.7$ and p-value is greater than 50%. There **is evidence** to suggest the mean number of pets is greater than 2.5.

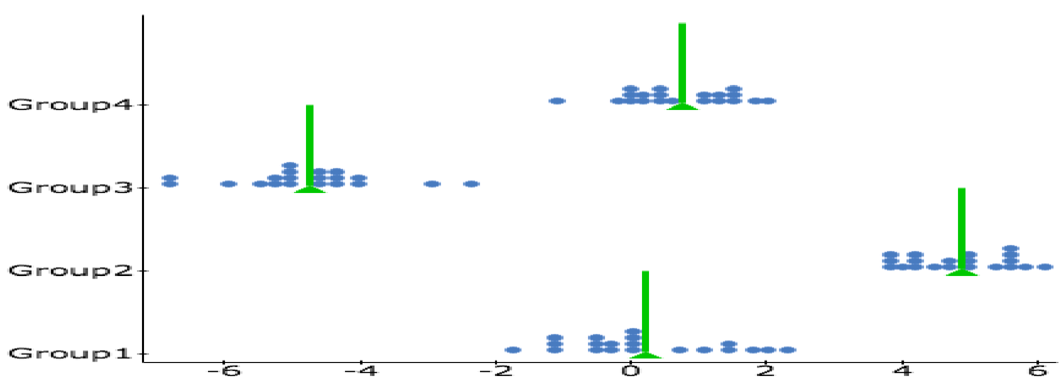
(C) $H_0 : \mu \geq 2.5$ vs $H_A : \mu < 2.5$, $t = 4.7$ and p-value is less than 0.5%. There **is evidence** to suggest the mean number of pets is greater than 2.5.

(D) $H_0 : \mu \leq 2.5$ vs $H_A : \mu > 2.5$, $t = -4.7$ and p-value is less than 5%. There **is some evidence** to suggest the mean number of pets is greater than 2.5.

(E) $H_0 : \mu \leq 1.99$ vs $H_A : \mu > 1.99$, $t = 0$ and p-value is 50%. There **is evidence** to suggest the mean number of pets is greater than 2.5.

(14) 4 different populations are sampled. It is plotted below. Let $\mu_1, \ldots, \mu_4$ denote the population means for group $1, \ldots, 4$. Take careful note of the numbers on the x-axis and the groups numbers (1 is at the bottom and 4 is at the top).

What are the correct p-values corresponding to each test (due to limited space only the the alternative is stated).




| | $H_A : \mu_2 - \mu_1 > 0$ | $H_A : \mu_3 - \mu_1 > 0$ | $H_A : \mu_4 - \mu_1 > 0$ |
|---|---|---|---|
| (A) | more than 50% | less than 5% | more than 50% |
| (B) | less than 1% | more than 50% | between 1 − 10% |
| (C) | more than 50% | more than 50% | less than 5% |
| (D) | between 1 − 10% | between 1 − 10% | more than 50% |
| (E) | less than 0.01% | less than 0.01% | less than 0.01% |

(15) The proportion of people who get a minor reaction when taking a pain-control pill is somewhere between **0-15%**.

The proportion of people who develop a reaction using PM pain-control pill is being assessed. To estimate the proportion a random sample is to be taken. What is the minimum number of people that should be sampled to ensure that Margin of Error is **3%** or less (using a **95%** confidence interval)?

(A) 545 (B) 747 (C) 1067 (D) 1668 (E) 2000

(16) Hersheys chocolate bar states that that its mean weight is 50g. However, there has been some speculation that one of its machines is putting **too much** chocolate in each bar and they are heavier than what is expected. Hersheys want to check this and sample 16 chocolate bars. What is the hypothesis of interest? Use the output below to do the test at the 5% level.

**Hypothesis test results:**
μ : Mean of variable
$H_0 : \mu = 50$
$H_A : \mu \neq 50$

| Variable | Sample Mean | Std. Err. | DF | T-Stat | P-value |
|---|---|---|---|---|---|
| chocolate | 47.839375 | 0.59971762 | 15 | -3.6027372 | 0.0026 |

(A) $H_0 : \mu = 50$ vs $H_A : \mu \neq 50$, the p-value is 0.26% there **is evidence** to suggest that the mean weight of chocolate bars is **greater** than 50g.

(B) $H_0 : \mu \leq 50$ vs $H_A : \mu > 50$, the p-value is 99.74% there is **no evidence** to suggest that the mean weight of chocolate bars is **greater** than 50g.

(C) $H_0 : \mu \leq 50$ vs $H_A : \mu > 50$, the p-value is 0.13% there **is evidence** to suggest that the mean weight of chocolate bars is **greater** than 50g.

(D) $H_0 : \mu \geq 50$ vs $H_A : \mu < 50$, the p-value is 0.13% there **is evidence** to suggest that the mean weight of chocolate bars is **greater** than 50g.

(E) $H_0 : \mu \leq 50$ vs $H_A : \mu > 50$, the p-value is 99.87% there is **no evidence** to suggest that the mean weight of chocolate bars is **greater** than 50g.

(17) In a recent WHO report the (average) life expectancy of males and females are given in 194 countries. Using the Statcrunch output, what can we say about the life expectancy of females verses males?

**Hypothesis test results:**

$\mu_D = \mu_1 - \mu_2$ : Mean of the difference between FemaleLife and Male

$H_0 : \mu_D = 0$

$H_A : \mu_D \neq 0$

| Difference | Sample Diff. | Std. Err. | DF | T-Stat |
|---|---|---|---|---|
| FemaleLife - MaleLife | 4.685567 | 0.19880388 | 193 | 23.56879 |

(A) If we test the hypothesis $H_0 : \mu_{Female} - \mu_{Male} \leq 0$ vs $H_A : \mu_{Female} - \mu_{Male} > 0$ we cannot reject the null.

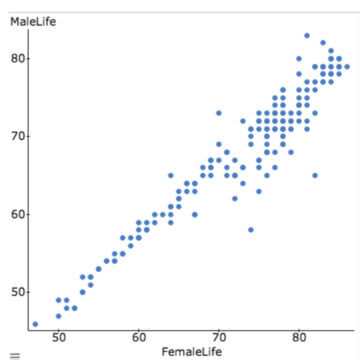
(B) If we test the hypothesis $H_0 : \mu_{Female} - \mu_{Male} = 0$ vs $H_A : \mu_{Female} - \mu_{Male} \neq 0$ we cannot reject the null.

(C) Without the t-tables for 193 df it is not possible to say.

(D) The t-value is extremely large, there is clear evidence that women, on average, outlive men.

(E) [C] and [D].

(18) We make a plot of the life expectancy of females against male expectancy in 194 countries.




(A) In countries where women live longer, men tend to live longer too.

(B) There is clear matching between the male and female life expectancy and this is why we have to do a matched paired t-test in (Q17).

(C) There is no dependence between female and male life expectancy.

(D) [A] and [B]    (E) [B] and [C].

(19) In a recent WHO report the road death rates (per 100,000) for 194 countries was given.

In the Statcrunch output below we compare the death rates in countries in the Americas (35 countries in total) with those in Europe (53 countries in total). Note that there were 7.54 **more deaths** in the Americas compared with Europe.

**Hypothesis test results:**
$\mu_1$ : Mean of AmericasDeathRate
$\mu_2$ : Mean of EuropeDeathRate
$\mu_1 - \mu_2$ : Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 \neq 0$
(without pooled variances)

| Difference | Sample Diff. | Std. Err. | DF | T-Stat | P-value |
|---|---|---|---|---|---|
| $\mu_1 - \mu_2$ | 7.5417892 | 1.6059147 | 45.794142 | 4.6962578 | <0.0001 |

(A) The difference of 7.54 is statistically significant suggesting, on average, there are more deaths on the roads of the Americas compared with Europe.

(B) The difference of 7.54 is **not** statistically significant, this difference can be explained by random variation.

(C) There must be a clear matching between the countries in the Americas and Europe and a matched paired t-test should have been done instead.

(D) [A] and [C] [B] and [C].

(20) Based on the data choose the correct testing procedure.

1. In 2006 the average mile per gallon of medium size car was 34 mpg. This year the EPA wants to investigate whether mileage has improved amongst medium sized cars. They take a random sample of 400 medium sized cars and measure their mileage.
2. The cycling habits of students from Texas A&M and UT are compared.
3. To see whether a new teaching technique works pupils in a school are assessed before the teaching technique is applied and after the teaching technique is applied. The difference in scores are analyzed.

| | [1] | [2] | [3] |
|---|---|---|---|
| (A) | Matched t-test | One sample t-test | Independent two sample t-test |
| (B) | Independent two sample t-test | Matched t-test | One sample t-test |
| (C) | One sample t-test | Independent two sample t-test | Matched t-test |
| (D) | One sample proportions | Matched t-test | Independent two sample t-test |
| (E) | Independent two sample t-test | one sample test on proportions | one sample t-test |

(21-22) A survey is conducted where 5890 females and males are asked whether they buy organic food (yes or no). The data is summarized in the Table below.

| Cell format |
|---|
| Count<br>(Row percent)<br>(Column percent) |

| | Female | Male | Total |
|---|---|---|---|
| No | 1656<br>(51.46%)<br>(49.97%) | 1562<br>(48.54%)<br>(60.64%) | 3218<br>(100%)<br>(54.63%) |
| Yes | 1658<br>(62.05%)<br>(50.03%) | 1014<br>(37.95%)<br>(39.36%) | 2672<br>(100%)<br>(45.37%) |
| Total | 3314<br>(56.26%)<br>(100%) | 2576<br>(43.74%)<br>(100%) | 5890<br>(100%)<br>(100%) |

**Chi-Square test:**

| Statistic | DF | Value | P-value |
|---|---|---|---|
| Chi-square | 1 | 66.536688 | <0.0001 |

(21) (A) Given the small p-value there is evidence to suggest that gender and preference to buy organic food are **statistically independent**.

(B) Given the small p-value there is evidence of **an association/dependence** between gender and a preference for organic food. Further, females are more likely to buy organic food than men (50% vs 39.36% based on this survey).

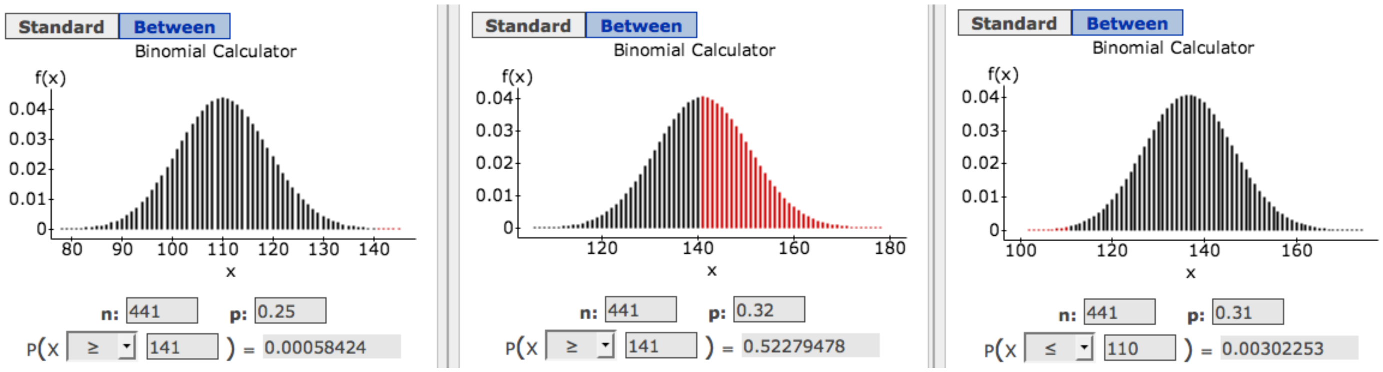
(C) 51.46% of females don't buy organic food and 48.54% of males don't buy organic food (based on this survey).

(D) [A] and [C] (E) [B] and [C].

(22) Suppose that there is **no association** between gender and buying organic food. Given 3314 females, how many females would you expect to buy organic food (look carefully at the table)?

(A) 1658 (B) 2054 (C) 1304 (D) 3314 (E) 1503

(23) In a recent poll 141 of 441 registered Republican voters voiced concern about Mr. Trump. Is there any evidence that over 25% of Republican voters are concerned about Mr. Trump? State the null and alternative and the result of the test.




(A) $H_0 : p \leq 0.25$ vs $H_A : p > 0.25$, the p-value is less than 0.1%. There is evidence to suggest that over 25% of Republican voters are concerned about Mr. Trump.

(B) $H_0 : p \leq 0.25$ vs $H_A : p > 0.25$, the p-value is over 50%. There is no evidence to suggest that over 25% of Republican voters are concerned about Mr. Trump.

(C) $H_0 : p \geq 0.25$ vs $H_A : p < 0.25$, the p-value is over 50%. There is no evidence to suggest that over 25% of Republican voters are concerned about Mr. Trump.

(D) $H_0 : p \geq 0.25$ vs $H_A : p < 0.25$, the p-value is 25%. There is no evidence to suggest that over 25% of Republican voters are concerned about Mr. Trump.

(E) $H_0 : p \geq 0.31$ vs $H_A : p < 0.31$, the p-value is less than 0.1%. There is evidence to suggest that less than 31% of Republican voters are concerned about Mr. Trump.

(24) Using the output below fill in the blank: "32% (Margin of Error **X%**) of registered Republican voters voiced concern about Mr. Trump (based on a 95% CI)". What is **X%**?

**95% confidence interval results:**
p : Proportion of successes
Method: Standard-Wald

| Proportion | Count | Total | Sample Prop. | Std. Err. |
|---|---|---|---|---|
| p | 141 | 441 | 0.31972789 | 0.022208146 |

Figure 2: Output on Mr. Trump poll

(A) 2.22% (B) 4.35% (C) 5% (D) 3.19% (E) 2.5%

(25) The proportion of females students attending a university is 49%, the proportion of males is 49% and the proportion of students with no designated gender is 2%.

The proportion of animal science majors at the university is 15%.

Assume there is no dependence between gender and major. What proportion of students have no designated gender **and are** animal science students?

(A) 17% (B) 2.15% (C) 0.3% (D) Between 1-2% (E) 0.13%.