Midterm 3 - STAT 301
Fall 2013

Name:

UIN:

Signature:

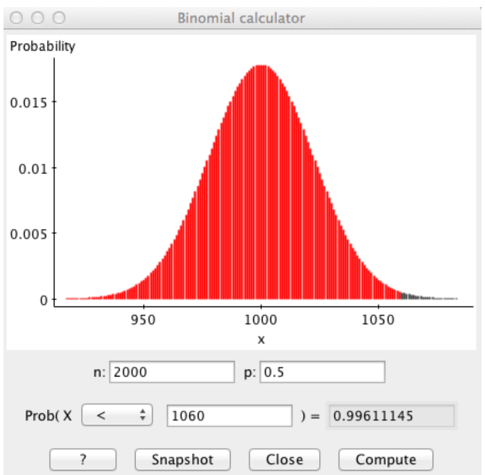
Version A:

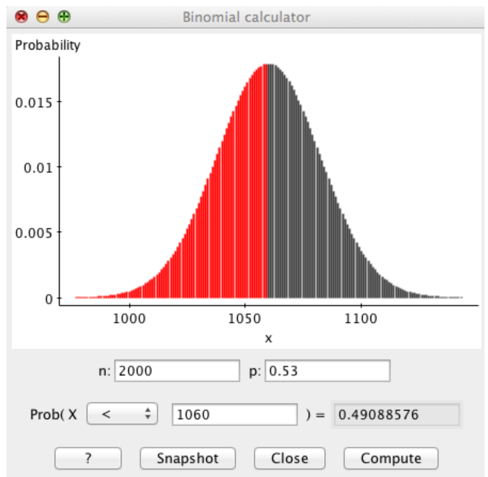
1. Do not open this test until told to do so.
2. This is a closed book examination, However you may use two single-sided sheet of formulas that you have brought with you and the tables (z, t and chi tables). You should have no other printed or written material with you on the exam. But scrap paper is 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. 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. You eyes should not be obstructed by caps.
10. Good Luck!!!

(1) TEXAS GOVERNOR. Greg Abbott and Wendy Davis are two possible candidates for Texan Governor. In a poll of 2000 Texans it was found that

Abott	Davis
53%	47%

. Is there any evidence that Greg Abbott will win the election (do the test at the 5% level)? Possibly useful information is given in the figures below.





(A) Testing $H_0 : p = 0.5$ against $H_A : p > 0.5$, the p-value is 0.4%, therefore we can reject the null. There is evidence to suggest that Abbott will win.

(B) Testing $H_0 : p = 0.53$ against $H_A : p > 0.53$, the p-value is 51%, therefore we can reject the null. There is evidence to suggest that Abbott will win.

(C) Testing $H_0 : p = 0.53$ against $H_A : p > 0.53$, the p-value is 51%, therefore we cannot reject the null. It is unclear who will win the election.

(D) Testing $H_0 : p = 0.5$ against $H_A : p < 0.5$, the p-value is 99.6%, therefore we cannot reject the null. It is unclear who will win the election.

(E) Testing $H_0 : p = 0.5$ against $H_A : p > 0.5$, the p-value is 0.4%, therefore we cannot reject the null. It is unclear who will win the election.

(2) TEXAS GOVERNOR. Gallop conducts a poll to see whether there is a difference between voting habits in Austin and Houston. They randomly sample 500 people in Austin and Houston. Each person is asked whether they will vote for Abbott or Davis. 60% of the Austin sample said they would vote for Davis, whereas 48% of the Houston sample said they would vote for Davis. What test must be done to see if these differences are statistically significant?

(A) Two sample test for proportions. (B) A chi-square test for association.

(C) An independent sample t-test. (D) (A) and (B). (A) and (C).

(3) TEXAS GOVERNOR How large a sample size should Gallop use for the margin of error of a 99% CI to be less than 1%?

(A) $n = (\frac{2.57\times 0.5}{1})^2 = 1.65$ (B) $n = (\frac{2.57\times 0.5}{1})^2 = 1.65$

(C) $n = (\frac{2.32\times 0.5}{0.01})^2 = 13456$ (D) $n = (\frac{2.57\times 0.5}{0.01})^2 = 16512$.

(E) p is unknown, without this knowledge it is impossible to say.

(4) WIN! You read the following news article:

The Davidson family from East Texas have good reason to celebrate. The Davidson's have won the lottery not once but *twice* in the past year! The chance of winning the lottery once is one in a hundred thousand (one in 10^5) the chance of winning it twice is extremely small (one in $10^5 \times 10^5 = 10^{10}$). What a luck family!

Comment on the accuracy of the probabilities in the article (assuming that the lottery was fair and no cheating took place).

(A) The probability is incorrect. The probability of winning the lottery *twice* is two in a hundred thousand ($10^{-5} + 10^{-5}$).

(B) The probabilities quoted are correct. Since both lottery wins are statistically independent events.

(C) Since this probability is extremely small, it is statistically significant and we can reject the null.

(D) The probability is calculated under the assumption that the events are statistically independent. However, as the same family is involved in both wins this assumption is unlikely to be true.

(E) C and D.

(5) WIN! You read the following news article:

The Smith family have good reason to celebrate. The Smith Siblings, Peter and Jane, are amateur long distance runners and *both* won the New York Marathon this year! The chance of an amateur winning a major marathon is one in a thousand. Therefore the chance of both siblings winning the marathon is one in a million (one in $10^{-3} \times 10^{-3} = 10^{-6}$). What a lucky family!

Comment on the accuracy of the probabilities in the article.

(A) The probability is incorrect, the probability of winning the lottery *twice* is two in a thousand ($10^{-3} + 10^{-3}$).

(B) The probabilities quoted are correct. Since both wins don't depend on each other.

(C) Since this probability is extremely small, it is statistically significant and we can reject the null.

(D) The probability is calculated under the assumption that the events are statistically independent. However, they are siblings therefore this assumption is unlikely to be true.

(E) A and C.

(6-7) FIT Are we getting less fit? It has been conjectured that children are slower than their parents.

To test this, a recent study compared the running time of children with their mother at the same age. 30 mother/10 years old child pairs were randomly selected and interviewed.

The running time of the mother at age 10 was asked and the running time of their 10 child was timed (the distance for both was 500m). It was found that the sample mean (average) of *mother's time was 4.5 minutes*, whereas the sample mean for the *child was 4.7 minutes*.

The results of a matched pair t-test is given in the Statcrunch output below (the differences between each mother and child pair was calculated and a one sample test done).

One sample T statistics with summary

Options

Hypothesis test results:

μ : population mean

$H_0 : \mu = 0$

$H_A : \mu < 0$

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	0.2	0.054772254	29	3.6514838	0.9995

(6) FIT Let μ_C correspond to the mean child run time and μ_M the mean mother run time. What is the hypotheses of interest and the result of the test at the 5% significance level?

(A) We are testing $H_0 : \mu_C - \mu_M = 0$ against $H_A : \mu_C - \mu_M > 0$, the p-value is 0.05%. There is no evidence that children are getting slower.

(B) We are testing $H_0 : \mu_C - \mu_M = 0$ against $H_A : \mu_C - \mu_M < 0$, the p-value is greater than 99.95%, therefore there is no evidence that children and slower.

(C) We are testing $H_0 : \mu_C - \mu_M > 0$ against $H_A : \mu_C - \mu_M = 0$, the p-value is 0.05%. Therefore, there is evidence to suggest that children and mothers have the same running times.

(D) We are testing $H_0 : \mu_C - \mu_M = 0$ against $H_A : \mu_C - \mu_M > 0$, the p-value is 99.95%. Therefore, there is no evidence to suggest that children are slower.

(E) We are testing $H_0 : \mu_C - \mu_M = 0$ against $H_A : \mu_C - \mu_M > 0$, the p-value is 0.05%. Therefore, there is evidence to suggest that children are slower.

(7) FIT Construct a 90% confidence interval for the mean difference of mother and child runnning time.

(A) $[0.2 \pm 2.045 \times 0.054]$ (B) $[0.2 \pm 2.045 \times 3.065]$ (C) $[0.2 \pm 1.699 \times 0.054]$

(D) $[0.2 \pm 1.311 \times 0.054]$ (E) $[0.2 \pm 1.699 \times 0.054]$.

(8-11) ISOLATION University councillers are worried that a large number of students 'feel left out and isolated' when they start university. 530 students were interviewed. They answer yes or no to the question 'do you feel isolated?'. The following table breaks down the results by gender:

Contingency table results:
Rows: Isolation
Columns: None

Cell format
Count (Row percent) (Column percent) (Total percent)

	Male	Female	Total
Yes	50 (55.56%) (14.29%) (9.434%)	40 (44.44%) (22.22%) (7.547%)	90 (100.00%) (16.98%) (16.98%)
No	300 (68.18%) (85.71%) (56.6%)	140 (31.82%) (77.78%) (26.42%)	440 (100.00%) (83.02%) (83.02%)
Total	350 (66.04%) (100.00%) (66.04%)	180 (33.96%) (100.00%) (33.96%)	530 (100.00%) (100.00%) (100.00%)

(8) ISOLATION. From the data what can we say about gender and isolation?

(A) The proportion of males in the sample that feel isolated is 14.29%, the proportion of females that feel isolated is 22.24%. In this sample, females tend to feel more isolated than males.

(B) The proportion of males in the sample that feel isolated is 9.43%, the proportion of females that feel isolated is 7.54%. In this sample, males tend to feel more isolated than females.

(C) The proportion of males in the sample that feel isolated is 55.56%, the proportion of females that feel isolated is 44.44%. In this sample males tend to feel more isolated than females.

(D) A major drawback with this study is that the number of males and females are <u>not</u> the same. This means we cannot make a meaningful comparison of data between male and female isolation.

(E) C and D.

(9) ISOLATION What proportion of the sample are both female <u>and</u> feel isolated?

(A) 33.96% (B) 16.98% (C) 44.4% (D) 22.22% (E) 7.74%.

(10) ISOLATION Suppose there is NO association between gender and isolation. Based on the above data how many males out of 350 would you <u>expect</u> to feel isolated?

(A) It is impossible to say, as we do not have data available for this situation.

(B) $33.005 = 350 \times 0.0943$ (C) $50 = 350 \times 0.1429$

(D) $59.43 = 350 \times 0.1698$ (E) $4.715 = 50 \times 0.0943$

(11) ISOLATION To see whether there is any association between gender and isolation a chi-squared test was done. The chi-squared value is **5.3**. What is the result of the test at the 5% level.

(A) We test H_0 : Association between gender and isolation against H_A : There is no association between gender and isolation. Since $5.3 > 5.02$ the p-value is less than 2.5%, we reject the null. There is evidence to suggest there is <u>no association</u> between gender and isolation.

(B) We test H_0 : No association between gender and isolation against H_A : There is an association between gender and isolation. Since $5.3 > 5.02$ the p-value is less than 2.5% and we reject the null. There is evidence to suggests there <u>is an association</u>.

(C) We test H_0 : No association between gender and isolation against H_A : There is an association between gender and isolation. Since $5.3 > 3.48$ the p-value is greater than 5% and we cannot reject the null. There is <u>no evidence</u> to suggest there is an association between gender and isolation.

(D) We test H_0 : Association between gender and isolation against H_A : There is no association between gender and isolation. Since $5.3 > 3.48$ the p-value is greater than 5% and we cannot reject the null. There is <u>no evidence</u> to suggest there is association between gender and isolation.

(E) We test H_0 : Association between gender and isolation against H_A : There is no association between gender and isolation. Since $5.3 > 5.02$ the p-value is less than 2.5%, we reject the null. Therefore, Gender and Isolation are statistically independent events.

(12-14) NUTS Researchers at the National Institute of Health (NIH), have been following a group of 500 randomly selected volunteers over a 40 year period. At the start of the study the people were in their 50's and study ended when the the last person had died. At regular intervals, each volunteer was asked about the food they had eaten. Based on the food information that given by the volunteer, each volunteer was placed into either a LOW NUT CONSUMPTION GROUP or REGULAR CONSUMPTION GROUP. Let μ_L and μ_R denote the mean life expectany for those not taking regular nuts and those taking regular nuts respectively.

(12) NUTS. Suppose you want to investigate whether people who consume nuts on a regular basis have a different life expectany to those that have a low nut consumption. What are the hypotheses of interest?

(A) $H_0 : \mu_R - \mu_L \leq 0$ vs $H_A : \mu_H - \mu_R = 0$. (B) $H_0 : \mu_R - \mu_L = 0$ vs $H_A : \mu_R - \mu_L > 0$

(C) $H_0 : \mu_R - \mu_L \geq 0$ vs $H_A : \mu_R - \mu_L = 0$ (D) $H_0 : \mu_R - \mu_L = 0$ vs $H_A : \mu_R - \mu_L < 0$

(E) $H_0 : \mu_R - \mu_L = 0$ vs $H_A : \mu_R - \mu_L \neq 0$.

(13) NUTS. Based on this data, the average life time of those in REGULAR nut group was 84 years, whereas the average life time in the LOW nut group was 82.5 years. The researchers want to test the hypothesis that $H_0 : \mu_R - \mu_L = 0$ against $H_A : \mu_R - \mu_L > 0$. Using this data set they get the Statcrunch output:

Options

Hypothesis test results:
μ_1 : mean of population 1
μ_2 : mean of population 2
$\mu_1 - \mu_2$: mean difference
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 \neq 0$
(without pooled variances)

Difference	Sample Mean	Std. Err.	DF
$\mu_1 - \mu_2$	1.5	0.6549407	229.94469

probability	0.15	0.10	0.05	0.025	0.01	0.005
t^*	1.04	1.28	1.65	1.97	2.34	2.60

T-tables for a t-distribution with 229.94df

The test is done at the 5% level. Which statement gives the correct result?

(A) The t-transform is $t = 2.29$, the p-value is between 1-5%. There is no evidence that people who regularly consume nuts live longer than those who don't.

(B) The t-transform is $t = 2.29$, the p-value is between 1-5%. There is evidence to suggest that the people who regularly consume nuts tend to live longer than those who don't.

(C) The t-transform is $t = 2.29$, the p-value is between 1-5%. There is some evidence that the people who regularly consume nuts tend to have a shorter life expectancy than those who don't.

(D) The p-value is 0.0065%. There is no evidence that people who regularly consume nuts live longer than those who don't.

(E) The p-value is 0.0065%. There is some evidence that the people who regularly consume nuts tend to have a shorter life expectancy than those who don't.

(14) NUTS. Suppose that in the previous question the p-value was 0.00001. What can we say about nut consumption increasing life expectancy?

(A) The p-value is extremely small (smaller than most reasonable significance levels), this means that life expectancy increases substantially if nuts are consumed on a regular basis.

(B) The p-value is extremely small (smaller than most reasonable significance levels), this means that life expectancy increases by 0.00001 years if nuts are consumed on a regular basis.

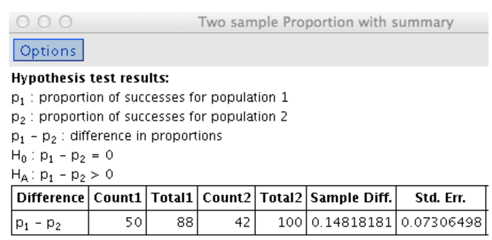
(C) The small p-value suggests that the people who consume nuts on a regular basis have a longer life expectancy. However, the statistical analysis did not take into account hidden factors, such as people who eat nut tend tend to have a healthy diet. Therefore, we cannot say whether regular nut consumption actually increases life expectancy.

(D) The small p-value means there is a very large chance that a type II error was made.

(E) (A) and (D).

(15) A drug for pain management has recently been developed. It is being compared to old drugs.

	New	Old
Pain reduced	53	42
Total	88	100


Two sample Proportion with summary

Options

Hypothesis test results:

p_1 : proportion of successes for population 1

p_2 : proportion of successes for population 2

$p_1 - p_2$: difference in proportions

$H_0 : p_1 - p_2 = 0$

$H_A : p_1 - p_2 > 0$

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.
$p_1 - p_2$	50	88	42	100	0.14818181	0.07306498

Let p_1 denote the change of improvement using the new drug and p_2 denote the chance of improvement using the old drug. We test $H_0 : p_1 - p_2 = 0$ against $H_A : p_1 - p_2 > 0$ (the output is given above). Which is the correct answer?

(A) There is no evidence to reject the null at the 5% level.

(B) There is evidence to reject the null at the 10% level, but not at the 5% level.

(C) There is evidence to reject the null at the 0.1% level.

(D) There is evidence to reject the null at the 5% level, but not at the 1% level. Therefore, there is evidence against the null, but it is not extremely strong.

(E) The p-value is greater than 95%, there is no evidence against the null.