TOTAL 150

185

Test 2 – Math 163
Fall, 2016

Name:

Show your work where indicated. Use 4 decimal places in your answers. Show all your work including the formulas you used or show what you entered on the calculator. You must state the test statistics and the test used to get full credit. You also must state the t, z, p, P value or any other pertinent numbers that you found to come to your conclusions.

Some probability rules for events A and B:

1. **If A and B are disjoint, $P(A \text{ or } B) = P(A) + P(B)$**
2. **A^C is the event that A does not occur. $P(A^C) = 1 - P(A)$**
3. **If A and B are independent, $P(A \text{ and } B) = P(A)P(B)$ and $P(B|A = P(B)$**
4. **For any two events A and B, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$**
5. **When $P(A) > 0$, the conditional probability of B given A is $P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$**

Calculator: $\text{normalcdf}(a, b, \mu, \sigma)$, $\text{invNorm}(p, \mu, \sigma)$, $\text{binompdf}(n, p, k)$, $\text{binomcdf}(n, p, k)$

1. The following table entries are counts of the number of college degrees (in thousands) earned in the US in 2009 – 2010 which was 3103 thousand degrees. For all the following questions, you need to clearly state the event and write the answer using the proper notation.

Degree	Female	Male	Total
Associate's	447	268	715
Bachelor's	945	651	1596
Master's	397	251	648
Professional	49	44	93
Doctorate	26	25	51
Total	1864	1239	3103

$$P(B|F) = \frac{P(B \text{ and } F)}{P(F)} = \frac{\frac{945}{3103}}{\frac{1864}{3103}} = \frac{945}{1864} = .507$$

4 If the individual that earned a college graduate is selected at random, what is the probability that the person received a Bachelor's degree given that that the person is a female? Do work in space to right of chart.

2. Fill in the blanks with the appropriate terms:

3 A. Statistical inference provides methods for drawing conclusions about a population from sample data.

3 B. The probability that the interval will capture the true parameter value in repeated samples is called a confidence interval

3 C. For a randomized experiment, the smaller the P – value, the stronger the evidence against H_0.

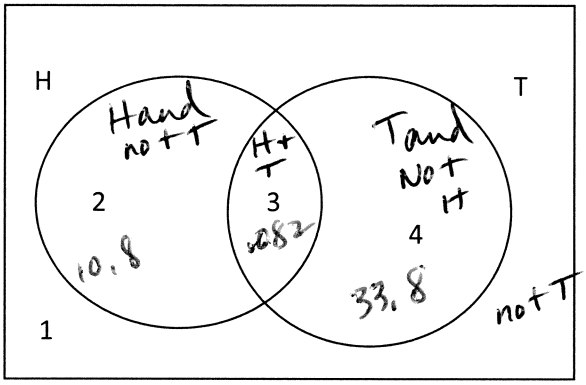
13

4 D. A population proportion is a parameter. A sample proportion is a statistic

3. In a survey of adults who own vehicles, 19% drove hybrids, 42% drove Toyotas, and 8.2% drove Toyota hybrids. Let H be the event that the person drives a hybrid, and let T be the event that the person drives a Toyota vehicle.

A. Fill in the Venn diagram. This should help you answer the questions below.

3



3 B. What is $P(T)$? = .42

3 C. What is $P(H)$? = .19

D. What does region 3 represent in the context of the question? What is the probability of region 3? T and H. Probability car is both a hybrid & a Toyota

4 $P(T \text{ and } H) = 0.82$

E. What is $P(H \text{ or } T)$? $= P(T) + P(H) - P(T \text{ and } H)$

4 $= .19 + .42 - .082 = .528$

F. Describe what region 1 represents in terms of probability. What is the probability of region 1?

4 $P(\text{not } T \text{ and not } H) = 1 - .082 - .108 - .338$

$= .472$

G. Describe what region 2 represents in terms of probability.

4 $P(H \text{ and not } T)$ prob. car is a hybrid but not a Toyota

H. Find the probability that the person drives a hybrid if it is known that the person drives a Toyota.

4 $P(H|T) = \dfrac{P(H \text{ and } T)}{P(T)} = \dfrac{.082}{.42} = .1952$

29/33

4. Lactose intolerance causes difficulty digesting dairy products that contain lactose.

A. In the US, 80% of the population is white, 15% is black and 5% are Asian. Let W be the event that a person is white, B be the event that a person is black and A be the event that a person is Asian. Using probability notation, write what these numbers are telling you. (This means what does 15% of the population being black mean in terms of probabilities).

$P(W) = .80 \quad P(B) = .15 \quad P(A) = .05$

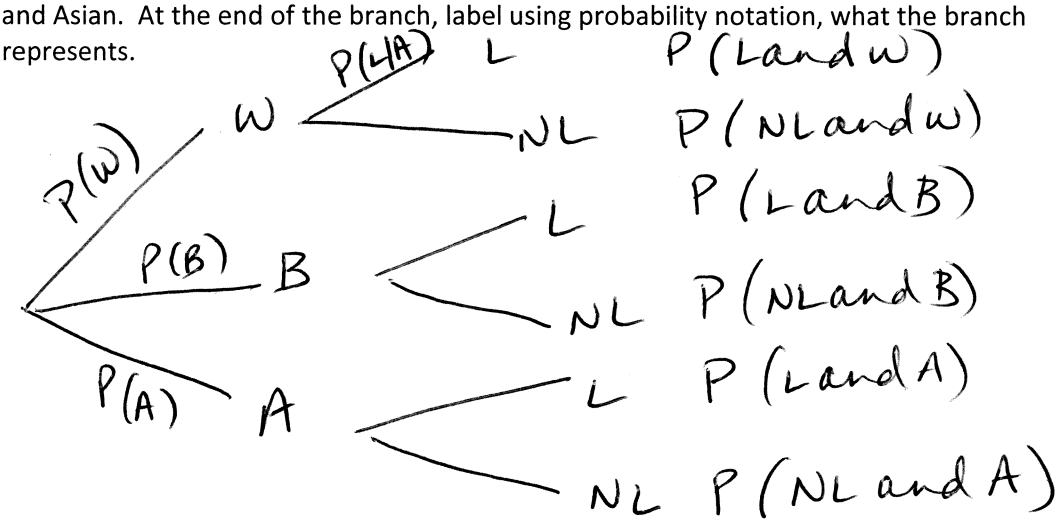
3

B. It is known that 15% of whites, 70% of blacks and 90% of Asians are lactose intolerant. Let L be the event that a person is lactose intolerant. Using probability notation, write what these numbers are telling you. Hint: these are conditional probabilities.

$P(L|W) = .15 \quad P(L|B) = .7$

3 $P(L|A) = .9$

C. Draw a tree diagram (labeling all branches) that would help you find the probability that a person is lactose intolerant and white, lactose intolerant and black and lactose intolerant and Asian. At the end of the branch, label using probability notation, what the branch represents.



5

D. Use the tree to find the probability of being lactose intolerant. Clearly show what branches you are referring to.

5

$P(L) = P(L \text{ and } W) + P(L \text{ and } B) + P(L \text{ and } A)$

$= (.8)(.15) + (.15)(.70) + (.05)(.90)$

$= .12 + .105 + .045 = .27$

16

5. A study found that 64% of all households in the United States subscribe to Netflix. You randomly selected 75 houses and ask them if they subscribe to Netflix.

A. State the 4 reasons that tell you this is a binomial setting.

4
① fixed no. of observations
② observations are independent
③ 2 possible outcomes - either get Netflix OR not
④ Prob of each the same $P(N) = .64$

B. What are the values for n and p?

2
$n = 75$ $p = .64$

C. If N is the number of households that subscribe to Netflix, what possible values can N take on?

3
$N = \{0, 1, \ldots 75\}$

D. What is the mean of this distribution? Explain what the mean tells you in the context of the problem.

4
$np = (.64)(75) \approx 48$
out of 75 households, on average, 48 will have Netflix

E. What is the standard deviation of this distribution?

2
$\sqrt{(75)(.64)(.36)} \approx 4.157$

F. Find the probability that $N = 63$. You should use the appropriate distribution on your calculator for this and parts G and H and show what you entered.

3
binompdf $(75, .64, 63) = .000076$

G. Find $P(N < 64)$. Show what you entered.

4
$P(N \leq 63) =$ binomcdf $(75, .64, 63)$
$= .99996$

H. Find $P(63 \leq N \leq 71)$. Show what you entered on the calculator.

4
$P(N \leq 71) - P(N \leq 62) =$ binomcdf $(75, .64, 71)$
$-$ binomcdf $(75, .64, 62) =$
$1 - .999888 = .000112$

26

6. The National Institute of Diabetes reports that the average cost of bariatric (weight loss) surgery is $22,500 with a standard deviation of $3015. You do not think this is correct. You randomly select 30 bariatric surgery patients and find that the average cost of their surgeries is $21,545.

A. State the null hypothesis and explain what it means.

3 $H_0: \mu = 22,500$ average cost of surgery is $22,500

B. State the alternative hypothesis and explain what it means.

3 $H_0: \mu \neq 22,500$ average cost is not $22,500 (two sided)

C. Would you use the z or t hypothesis test statistic to test your hypothesis? Explain your answer.

3 using population standard deviation

D. Find the z or t test statistic using the formula from the text and show your work. Explain what this number tells you.

4 $z = \frac{21,545 - 22,500}{3015/\sqrt{30}} = -1.749$

1.749 standard deviation below the mean

E. Using a 5% significance level, is there enough evidence to reject or accept the null hypothesis? We answered this question 2 different ways in class. Both are using calculator. Show how you get this answer using BOTH methods. Note: You will obviously get the same answer using both methods, so it is important that you show what you entered in your calculator for both methods. Explain what your conclusion means in the context of the problem. (19)

6 z test
$p = .0828$

Normal dist.
2 normalcdf (-100, -1.749, 0, 1)
OR 1 - normalcdf (-1.749, 1.749, 0, 1)

Since .0828 > .05 we do not have strong enough evidence that the sample mean is statistically different from the pop. mean So we accept the null hypothesis and conclude that the mean cost of bariatric surgery is $22,500.

7. Using the information from question #6, give a 95% confidence interval for the mean surgery cost. Explain what this interval tells you.

z interval (20,466, 22,624)

6 there is a 95% probability that the population mean falls in this interval. OR in 95% of collected in same way + of same size, pop. mean is in this interval.

8. A credit card watchdog group claims there is a difference in the mean credit card debt of households in New York and Texas. The results of a random survey of 250 households from each state are shown in the table below.

New York	Texas
$\overline{x_1} = \$4446.25$	$\overline{x_2} = \$4567.24$
$s_1 = \$1045.70$	$s_2 = \$1361.95$
$n_1 = 250$	$n_2 = 250$

A. State the null and alternative hypotheses and explain what they mean.

$H_0: \mu_{NY} = \mu_T$ mean debt is the same in both states.

5 $H_a: \mu_{NY} \neq \mu_T$ mean debt is different

B. Do the results support the group's claim to a 5% significance level? Explain what this means in the context of the problem.

2 sample t test

$t = -1.1141$

$p = .2658$

$.2658 > .05$ so we accept the null and conclude that the mean credit card debt is the same in New York + Texas.

5

16

9. Field workers for a health department compared driving mileage on a five day work schedule to a four day work schedule. Their goal was to see if the four day schedule reduced the amount of miles driven for a one year time period for 11 employees. The following chart gives the names of the 11 employees, the amounts they drove using the two different schedules and the difference in mileage.

Name	5-day mileage	4-day mileage	Difference
Colin	2798	2914	-116
Amanda	7724	6112	1612
Ian	7505	6177	1328
Gordon	838	1102	-264
Caitlin	4592	3281	1311
Noah	8107	4997	3110
Alina	1228	1695	-467
Andrew	8718	6606	2112
Harry	1097	1063	34
Lara	8089	6392	1697
Reid	3807	3362	445

A. To find if there is strong evidence to support the claim that the four day schedule reduces the driving time, the researcher uses a matched pair *t* test. Explain why this test is used.

4 using matched pair t test because you are comparing 2 different schedules for the same person

B. State the null and alternative hypotheses and explain what they mean.

5 $H_0: \mu_5 - \mu_4 = 0$ OR $\mu_5 = \mu_4$ Claim there is no difference in the amount of miles driven for each schedule

$H_a: \mu_5 > \mu_4$ OR $\mu_5 - \mu_4 > 0$ there is a difference

C. Find the sample mean and standard deviation of the difference.

2 $\bar{x} = 982$ $S = 1140$ (calculator)

(16)

D. Is there evidence that the four day schedule does reduce driving time to a 3% significance level? Explain what this means in the context of the problem and show your work.

using t test $t = 2.858$ $p = .0085$

5 $.0085 < .03$ so reject null and conclude that the 4 day schedule reduces the amount of miles driven

Not a 2 sample t test - doing test on difference in miles driven.

10. A sample survey on the drinking habits of Americans estimated the percentage of adults across the country who drink beer, wine, or hard liquor at least occasionally. Of the 1516 adults interviewed, 985 said they drink. Estimate with 98% confidence the proportion of all American adults who drink alcoholic drinks. Explain what the confidence interval tells you.

$\hat{p} = \frac{985}{1516} = .6497$ 1 prop z interval

(.62123, .67824)

6 98% confident (98% probability) that the population proportion of adults who drink occasionally is between 62.123% and 67.824%

11. In 2006, 9.8% of all U.S. families had incomes below the poverty level (U.S. Census Bureau). During the same year, of 400 randomly selected Wyoming families, 25 had incomes below the poverty level. The state of Wyoming wants to know if this sample gives evidence that, in 2006, that the percentage of families with incomes below the poverty level in Wyoming was different than among all U.S. families

A. Find $\hat{p}$ and explain what this number tells means.

4 $\hat{p} = \frac{25}{400} = .0625$ 6.25% of Wyoming residents in this sample live below the poverty level.

B. State the null and alternative hypotheses and explain what they mean.

5 $H_0: p = .098$ Wyoming has same proportion living in poverty as rest of US

$H_a: p \neq .098$ ↳ Wyoming has different proportion

C. Find the value of z and P.

3 $z = -2.388$ $p = .0169$ (22)

D. Decide if this sample gives good evidence to support the claim that the poverty level in Wyoming is different than the poverty level in the U.S to a 1% significance level. Explain in context of problem.

4 .0169 > .01 so accept null & conclude that the percentage of families living in Wyoming whose income is below the poverty level is the same as the rest of the US.

12. If you computed a t – score to be 4.53, would you accept or reject the null hypothesis? Would you need to know the P – value? Would the significance level be relevant? Explain.

$t = 4.53$ is a really big t score so the P-value would be very small. Look at calculator

5 $P = 1 - \text{normalcdf}(-4.53, 4.53, 0, 1) = .0000059\,04$

so reject null (α would never be that same)

13. To gauge the demand for vegetarian meals in restaurants 1181 U.S. adults were surveyed. The SRS consisted of 747 men and 434 women. Of those sampled, 276 men and 195 women said that they sometimes order a dish without meat, fish or fowl when they eat out. The study wished to use the data to see if the percentage of men who sometimes ordered vegetarian is smaller than the percentage of women who sometimes order vegetarian.

A. State the null and alternative hypothesis and explain what they mean.

5 $H_0: p_M = p_W$ percent who order vegetarian is same for men + women

$H_a: p_M < p_W$ percent for men is smaller

B. Find $\widehat{p_1}, \widehat{p_2}$, and $\hat{p}$ and explain what they mean.

men $\hat{p}_1 = .3695$ percent of men who sometimes order vegetarian

women $\hat{p}_2 = .4493$ " women "

4 both $\hat{p}_3 = .3988$ " of both "

C. Find the value of z and P and explain what these values tell you.

$z = -2.702$

$P = .00345$

2 proportion z test

3

(21)

D. Use your answer in part D to make a conclusion about the null and alternative hypotheses to a 5% significance level. Explain in context of the problem.

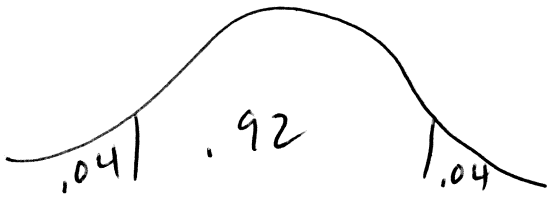
$.0034 < .05$ so we reject the null hypothesis

4 and conclude that the percent of men who sometimes order vegetarian is less than the percent of women who sometimes order vegetarian.

14. If you are asked to find a 92% confidence interval without using the z –interval on your calculator and instead are asked to use $InvNorm$.

A. Sketch a Normal distribution illustrating the relevant areas on the distribution.

3



B. What would you enter on your calculator to find the correct z score? What is the z score?

3 InvNorm (.96, 0, 1) = 1.7506

Remember inverse norm finds z score for entire area to the left of the z score

15. If you found $P = 0.314$ when the alternative hypothesis was one sided, what would the P – value be if the hypothesis was two – sided? Sketch the distribution and shade in the relevant portions.

2(.314) = .628

3

16. If you were given $z = 1.1$ and it was known that the alternative hypothesis was one – sided, does it make sense that $P = 0.40$? Explain you reasoning and include a sketch of the distribution. No

68% of area is ±1 st. dev. from mean

4

13

z value would be close to zero if P = .40

normal cdf (1.1, 100, 0, 1) = .1357

So area is 13.57%