Stat 2610 Test 3 Review Names Todd

For full credit circle answers and **show all your work.**

1. Binomial probability distributions depend on the number of trials *n* of a binomial experiment and the probability of success *p* on each trial. Under what conditions is it appropriate to use a normal approximation to the binomial? (Circle all that apply.)

np > 10 np > 5 nq > 10 nq > 5 n > 30 np < 5 nq < 5

2. In problem #3 below, determine if it is appropriate to use the normal approximation to the binomial. State why or why not.

$.11 \times 36 = 3.96 < 5$

$\therefore$ It is not appropriate to use the normal approximation.

3. Regardless of your answer from #2, use the normal distribution to estimate the requested probabilities. After a concerted effort was made to remove lead from the environment only 11% of children in the United States are at risk of high blood-lead levels. In a random sample (our classroom) of 36 students, what is the probability that 10 or more had high blood-lead levels? (Round your answer to three decimal places.)

normal cdf $(9.5, 36.5, .11 \times 36, \sqrt{36 \times .11 \times .89}) = .002$

4. Suppose the heights of 18-year-old men are approximately *normally distributed*, with mean 72 inches and standard deviation 1 inches. Calculate the probability that an 18-year-old man selected at random is between 71 and 73 inches tall? (Round your answer to three decimal places.)

normal cdf $(71, 73, 72, 1) = .683$

It is normally distributed, so not approximating - don't need to round!

5. Ocean fishing for billfish is very popular in the Cozumel region of Mexico. In the Cozumel region about 46% of strikes (while trolling) resulted in a catch. Suppose that on a given day a fleet of fishing boats got a total of 22 strikes. Find the following probabilities; if appropriate, use the normal distribution to approximate these binomial probabilities. (Round your answers to four decimal places.)

12 or fewer fish were caught normal cdf $(-.5, 12.5, 22 \times .46, \sqrt{22 \times .46 \times .54}) = .8457$

5 or more fish were caught normal cdf $(4.5, 9999, 22 \times .46, \sqrt{22 \times .46 \times .54}) = .9919$

between 5 and 12 fish were caught normal cdf $(4.5, 12.5, 22 \times .46, \sqrt{22 \times .46 \times .54}) = .8376$

6. Suppose Ryan sampled 72 jumbo perch. The average weight for these fish was $x = 1.76$ pounds. Based on previous studies, we can assume that the weights of jumbo perch have a normal distribution, with $\sigma = 0.37$ pounds. Find an 80% confidence interval for the average weights of Ryan's jumbo perch in the study region. Round your answers to two decimal places.

Z - Interval (σ known)

$\sigma = .37$

$\bar{x} = 1.76$

$n = 72$

(1.70, 1.82) is the interval, but I heard Ryan can't bait his own hook!

Calculate the margin of error in the jumbo perch calculation knowing $E = z_c(\frac{\sigma}{\sqrt{n}})$. Round your answer to two decimal places. $E = 1.28\left(\frac{.37}{\sqrt{72}}\right)$

$= .0558$

$\approx .06$

OR

$1.82 - 1.76 = .06$

OR

$1.76 - 1.70 = .06$

7. Let's say I want a sleeping bag that will keep me warm in temperatures from 20°F to 45°F. A random sample of prices ($) for sleeping bags in this temperature range is given below. Assume that the population of x values has an approximately normal distribution.

80	60	40	85	75	120	30	23	100	110
105	95	105	60	110	120	95	90	60	70

Find the sample mean price $\bar{x}$ and sample standard deviation s.

$\bar{x} =$ \$81.65

$s =$ \$28.88

Create a 90% confidence interval for (label and round answers appropriately.)

T-Interval Data
List L1
Freq 1
C-level: .9

(\$70.49, \$92.82)

Explain in words the 90% confidence interval. (I want you to tell me something about sleeping bags.)

I'm 90% certain the true population mean for 20° to 45°F sleeping bags is between \$70.49 and \$92.82. If I took 100 of these samples, 90 of them would capture the true population mean.

As you fondly recall from the homework: $E = z_c\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}$ and $E = t_c\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$ and confidence intervals are found by using: $(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$.

8. Case studies showed that out of 10,120 convicts who escaped from certain prisons, only 8005 were recaptured. Let p represent the proportion of all escaped convicts who will eventually be recaptured. Find a 99% confidence interval for p. (Round your answers to three decimal places.)

1-Prop Z-Interval
$x = 8005$
$n = 10120$
c-level = .99

(.781, .801)

9. Which is the best statement of the meaning of the confidence interval?

A. 99% of all confidence intervals would include the true proportion of recaptured escaped convicts.

B. 1% of all confidence intervals would include the true proportion of recaptured escaped convicts.

C. 1% of the confidence intervals created using this method would include the true proportion of recaptured escaped convicts.

D. 99% of the confidence intervals created using this method would include the true proportion of recaptured escaped convicts.

10. Inorganic phosphorous is a naturally occurring element in all plants and animals, with concentrations increasing progressively up the food chain (fruit < vegetables < cereals < nuts < corpse). Geochemical surveys take soil samples to determine phosphorous content (in ppm, parts per million). A high phosphorous content may or may not indicate an ancient burial site, food storage site, or even a garbage dump. The Hill of Tara is a very important archaeological site in Ireland. It is by legend the seat of Ireland's ancient high kings. Independent random samples from two regions in Tara gave the following phosphorous measurements (ppm). Assume the population distributions of phosphorous are mound-shaped and symmetric for these two regions.

Region I: x_1; n_1=12					
540	810	790	790	340	800
890	860	820	640	970	720

Region II: x_2; n_2=16							
750	870	700	810	965	350	895	850
635	955	710	890	520	650	280	993

Let μ_1 be the population mean for x_1 and let μ_2 be the population mean for x_2. Find a 95% confidence interval for $\mu_1 - \mu_2$. (Round your answers to one decimal place.) (-140.2, 157.3)

2 Sample T-Interval
Data
L1
L2 .95

11. The U.S. Geological Survey compiled historical data about Old Faithful Geyser (Yellowstone National Park) from 1870 to 1987. Let x_1 be a random variable that represents the time interval (in minutes) between Old Faithful eruptions for the years 1948 to 1952. Based on 8900 observations, the sample mean interval was x_1 = 61.6 minutes. Let x_2 be a random variable that represents the time interval in minutes between Old Faithful eruptions for the years 1983 to 1987. Based on 24,989 observations, the sample mean time interval was x_2 = 71.8 minutes. Historical data suggest that $\sigma_1 = 9.19$ minutes and $\sigma_2 = 11.64$ minutes. Let μ_1 be the population mean of x_1 and let μ_2 be the population mean of x_2. Compute a 95% confidence interval for $\mu_1 - \mu_2$. (Use 2 decimal places.) 2 Sample Z - Interval - stats (-10.44, -9.96)

$\sigma_1 = 9.19$ $n_1 = 8900$
$\sigma_2 = 11.64$ $n_2 = 24989$
$\bar{x}_1 = 61.6$
$\bar{x}_2 = 71.8$

From the confidence interval you created, make a statement about the timing between Old Faithful's eruptions comparing the 1940's and 1950's to the 1980's. I'm 95% sure Old faithful took between 9.96 and 10.44 minutes longer between eruptions in the 1983-1987 years than in the 1948 to 1952 years. If I did 100 samples, 95 of them would capture the true population mean.

12. At Community Hospital, the burn center is experimenting with a new plasma compress treatment. A random sample of $n_1 = 314$ patients with minor burns received the plasma compress treatment. Of these patients, it was found that 266 had no visible scars after treatment. Another random sample of $n_2 = 420$ patients with minor burns received no plasma compress treatment. For this group, it was found that 103 had no visible scars after treatment. Let p_1 be the population proportion of all patients with minor burns receiving the plasma compress treatment who have no visible scars. Let p_2 be the population proportion of all patients with minor burns not receiving the plasma compress treatment who have no visible scars.
Find a 95% confidence interval for $p_1 - p_2$. (Round your answers to three decimal places.)

2-Prop Z-Int (.545, .659)
$x_1 = 266$ $x_2 = 103$ C: .95
$n_1 = 314$ $n_2 = 420$

From the confidence interval you created, make a statement about which treatment is better, group one or group two. I'm 95% certain the plasma compress was between .545 and .659 BETTER than no plasma compress. This means if I did 100 samples then 95 of them would contain the true difference.

As you fondly recall from the homework: $E = z_c\sqrt{(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})}$ and $E = t_c\sqrt{(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2})}$ and confidence intervals are found by using: $(\bar{x}_1-\bar{x}_2)-E<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+E$. Yes, I recall it fondly. Ahhhh... the good old days.

13. Case studies showed that out of 7,423 students who escaped from certain statistics classes, only 2152 were recaptured. Let p represent the proportion of all escaped convicts who will eventually be recaptured. Find a 99% confidence interval for p. (Round your answers to three decimal places.)

(0.276, 0.303) I'm 99% sure that between 27.6% and 30.3% of escaped stats students are recaptured.

14. Which is the best statement of the meaning of the confidence interval?

A. 99% of the confidence intervals created using this method would include the true proportion of recaptured escaped statistics students.
B. 1% of all confidence intervals would include the true proportion of recaptured statistics students.
C. 1% of the confidence intervals created using this method would include the true proportion of recaptured statistics students.
D. 99% of all confidence intervals would include the true proportion of recaptured statistics students.

15. Muscle cars are known to have large engines producing lots of horsepower – more is better and too much is best. The 2014 Ford Shelby GT 500 claims to produce 662 hp while the Chevrolet Camaro ZL1 claims to produce 580 hp. Several cars were tested on a dyno to measure actual hp output at the rear wheels where we expect ~10% power loss due to drive train. Assume the population distributions of horsepower are mound-shaped and symmetric for these two regions.

Shelby GT500: x_1; n_1=12					
595	610	595	600	630	605
620	590	620	615	600	610

Camaro ZL1: x_2; n_2=16							
520	515	500	510	525	510	525	505
525	515	505	510	520	510	520	515

Let μ_1 be the population mean for x_1 and let μ_2 be the population mean for x_2. Find a 95% confidence interval for $\mu_1-\mu_2$ (round your answers to one decimal place) and interpret the meaning of this confidence interval.

2 Sample T-Interval (84.7, 101.6)

I'm 95% certain the Shelby GT 500 produces between 84.7 and 101.6 more horsepower than the Camaro. Soon the weather will be nice and people can get out their fun cars and drive around!

16. David E. Brown is an expert in wildlife conservation. In his book *The Wolf in the Southwest: The Making of an Endangered Species* (University of Arizona Press), he records the following weights of adult grey wolves from two regions in Old Mexico.

Chihuahua region: x_1 variable in pounds				
86	75	91	70	79
80	68	71	74	64

Durango region: x_2 variable in pounds								
68	72	79	68	77	89	62	55	68
68	59	63	66	58	54	71	59	67

Use a calculator to find: $\bar{x}_1$, s_1, $\bar{x}_2$, and s_2

$\bar{x}_1 =$ 75.8 $\bar{x}_2 =$ 66.8

$s_1 =$ 8.32 $s_2 =$ 8.87

(3.17, 14.77)

Let μ_1 be the mean weight of the population of all grey wolves in the Chihuahua region. Let μ_2 be the mean weight of the population of all grey wolves in the Durango region. Find a 90% confidence interval for $\mu_1-\mu_2$. (Use 2 decimal places.) Finally, make a concluding statement comparing the wolves in the two regions.

I'm 90% certain the average weight of wolves in the Chihuahua region is between 3.17 and 14.77 lbs more than wolves in Durango region. If 100 samples, 90 would capture true diff.

17. At Community Hospital, the burn center is experimenting with a new plasma compress treatment. A random sample of $n_1 = 514$ patients with minor burns received the plasma compress treatment. Of these patients, it was found that 266 had no visible scars after treatment. Another random sample of $n_2 = 420$ patients with minor burns received no plasma compress treatment. For this group, it was found that 217 had no visible scars after treatment. Let p_1 be the population proportion of all patients with minor burns receiving the plasma compress treatment who have no visible scars. Let p_2 be the population proportion of all patients with minor burns not receiving the plasma compress treatment who have no visible scars.
Find a 95% confidence interval for $p_1 - p_2$. (Round your answers to three decimal places.)

(- .064, .065)

From the confidence interval you created, make a statement about which treatment is better, group one or group two. I'm 95% sure these treatments are NOT any different.
If 100 samples were taken than 90 of them would capture the true population mean.

18. How much customers buy is a direct result of how much time they spend in the store. A study of average shopping times in a large national housewares store gave the following information (Source: *Why We Buy: The Science of Shopping* by P. Underhill).

Women with female companion: 8.3 min.

Women with male companion: 4.5 min.

Suppose you want to set up a statistical test to challenge the claim that a woman with a female friend spends an average of 8.3 minutes shopping in such a store. What would you use for the null and alternate hypotheses if you believe the average shopping time is less than 8.3 minutes?

A. H_o: $\mu < 8.3$; H_1: $\mu = 8.3$

B. H_o: $\mu = 8.3$; H_1: $\mu > 8.3$

C. H_o: $\mu = 8.3$; H_1: $\mu < 8.3$

D. H_o: $\mu = 8.3$; H_1: $\mu \neq 8.3$

19. Gentle Ben is a Morgan horse at a Colorado dude ranch. Over the past 8 weeks, a veterinarian took the following glucose readings from this horse (in mg/100 ml). 93, 81, 106, 99, 111, 82, 90, 93. The sample mean is $x \approx 94.375$. Let x be a random variable representing glucose readings taken from Gentle Ben. We may assume that x has a normal distribution, and we know from past experience that $\sigma = 12.5$. The mean glucose level for horses should be $\mu = 85$ mg/100 ml.† Do these data indicate that Gentle Ben has an overall average glucose level higher than 85? Use $\alpha = 0.05$.

H_0: $\mu = 85$

H_1: $\mu > 85$

What is the value of the sample test statistic? $z = 2.12$
(Round your answer to two decimal places.)

Find (or estimate) the P-value. $p = 0.0169$
(Round your answer to four decimal places.)

From the evidence above, make a statement about Gentle Ben's glucose level.

$p < \alpha$ ∴ I conclude Gentle Ben's glucose level really is higher than 85.
Reject the null hypothesis.

20. Let x be a random variable that represents hemoglobin count (HC) in grams per 100 milliliters of whole blood. Then x has a distribution that is approximately normal, with population mean of about 14 for healthy adult women. Suppose that a female patient has taken 10 laboratory blood tests during the past year. The HC data sent to the patient's doctor are as follows: 14, 18, 16, 18, 14, 13, 13, 17, 16, 11. I am wondering if this information indicates the population average HC for this patient is higher than 14 at $\alpha = 0.01$. State the null and alternative hypotheses and calculate the test statistic and P-value.

H_0: $\mu = 14$ H_1: $\mu > 14$

What is the value of the sample test statistic? $t = 1.34$
(Round your answer to two decimal places.)

Find (or estimate) the P-value. $p = 0.1063$
(Round your answer to four decimal places.)

From the evidence above, make a statement about the patient's hemoglobin level.
Fail to reject the null hypothesis. The data do not support the claim that she has a high hemoglobin.

21. The U.S. Department of Transportation, National Highway Traffic Safety Administration, reported that 74% of all fatally injured automobile drivers were intoxicated. A random sample of 52 records of automobile driver fatalities in a certain county showed that 43 involved an intoxicated driver. Do these data indicate that the population proportion of driver fatalities related to alcohol is greater than 74% in Beltrami County? Use α = 0.10.

H_0: $\mu = .74$ H_1: $\mu > .74$

1-Prop Z-Test

Find the value of the sample test statistic: $z = 1.429$
(Round your answer to two decimal places)

Find the P-value of the test statistic: $p = 0.0765$
(Round your answer to four decimal places)

22. Make a conclusion about the proportion of driver fatalities related to alcohol being greater than 74% in Beltrami County from the data. (Recall: this should be a four or six word answer then end with a statement about the situation.) Reject the null hypothesis. Data do support the claim that more than 74% of driver fatalities in Beltrami County involve alcohol.

23. In this problem, assume that the distribution of differences is approximately normal. At five weather stations on Trail Ridge Road in Rocky Mountain National Park, the peak wind gusts (in miles per hour) for January and April are recorded below. Want to show the wind blows faster in January than in April.

Weather Station	1	2	3	4	5	
January	135	139	138	64	81	L3
April	104	112	112	88	64	L4

L5 – Difference

Paired Sample

H_0: $\mu_1 = \mu_2$
Jan = Apr

H_1: $\mu_1 > \mu_2$
Jan > April

Find the value of the sample test statistic: $t = 1.5221$
(Round your answer to two decimal places)

Find the P-value of the test statistic: $p = 0.1012$
(Round your answer to four decimal places)

Fail to reject the null hypothesis. Data do not support the claim that the wind blows faster ~~than~~ in Jan. than April.

Make a conclusion about the wind blowing faster in January than April, use $\alpha = 0.01$.

Fail to reject the null hypothesis. Data do not support the claim that the wind blows faster in Jan than April.

24. A Michigan study concerning preference for outdoor activities used a questionnaire with a six-point Likert-type response in which 1 designated "not important" and 6 designated "extremely important." A random sample of $n_1 = 50$ adults were asked about fishing as an outdoor activity. The mean response was $x_1 = 4.9$. Another random sample of $n_2 = 53$ adults were asked about camping as an outdoor activity. For this group, the mean response was $x_2 = 4.2$. From previous studies, it is known that $\sigma_1 = 1.7$ and $\sigma_2 = 1.2$. Does this indicate a difference (either way) regarding preference for camping versus preference for fishing as an outdoor activity? Use a 5% level of significance.

State the null and alternative hypotheses:

H0: $\mu_1 = \mu_2$ H1: $\mu_1 \neq \mu_2$

Give the test statistic: $z = 2.40$ and p-value: $p = 0.0163$

From the data above, make a conclusion about Michiganders' preferences of camping vs fishing.

Reject the null hypothesis. The evidence supports the claim that Michiganders differ on their preferences for camping and fishing.

25. I found a piece of string that measured 19.3 cm. Beth immediately began tying knots in the string but Justine quickly interrupted indicating he really wanted to measure the string after each knot Beth tied. The data they found are in the table.

Knots	0	1	2	3	4
Length	19.3	18.9	18.5	18.1	17.7

From these data construct a linear regression model. Provide the model and correlation coefficient.

$y = 19.3 - .4x$

Explain what the numbers mean in the regression equation in terms of the string:

$-.4$ means for each knot the string

19.3 means the string, with no knots, is 19.3 cm long.

Explain what the correlation coefficient means for this model:

$r = -1$ means there is a perfect inverse relationship between number of knots and string length.