STATISTICS STAT2033 (H)

Review for TEST–4 **Name:––––––––––––––––––––––––––**

––

Answers are perovided to check numerical values, the actual content, style and fromats etc will be discussed in class.

Use the given sample data to construct the indicated confidence interval for the population mean.

1) A random sample of 30 households was selected from a particular neighborhood. The number of cars for each household is shown below. Estimate the mean number of cars per household for the population of households in this neighborhood. Give the 95% confidence interval.

2 0 1 2 3 2 1 0 1 4
1 3 2 0 1 1 2 3 1 2
1 0 4 1 0 2 2 1 0 2

Provide an appropriate response.

2) The distribution of the number of vacation days per year offered by different U.S. companies is skewed to the right. Find a 95% confidence interval for the mean number of vacation days offered by U.S. companies. Explain what "95% confidence" means in this context.

3) A researcher estimates that the mean systolic blood pressure for women aged between 18 and 24 is 118 mmHg. At the 95% confidence level, the margin of error is 2.4 mmHg. How would you interpret this margin of error?

Insurance companies track life expectancy information to assist in determining the cost of life insurance policies. The insurance company knows that, last year, the life expectancy of its policyholders was 77 years. They want to know if their clients this year have a longer life expectancy, on average, so the company randomly samples some of the recently paid policies to see if the mean life expectancy of policyholders has increased. The insurance company will only change their premium structure if there is evidence that people who buy their policies are living longer than before.

86	75	83	84	81	77	78	79	79	81
76	85	70	76	79	81	73	74	72	83

4) Does this sample indicate that the insurance company should change its premiums because life expectancy has increased? Test an appropriate hypothesis and state your conclusion.

5) For more accurate cost determination, the insurance companies want to estimate the life expectancy to within one year with 95% confidence. How many randomly selected records would they need to have?

Interpret the given confidence interval.

6) A grocery store is interested in determining whether or not a difference exists between the shelf life of Tasty Choice doughnuts and Sugar Twist doughnuts. A random sample of 100 boxes of each brand was selected and the mean shelf life in days was determined for each brand. A 90% confidence interval for the difference of the means, $\mu_{TC} - \mu_{ST}$, was determined to be (1.4, 2.5).

Test the given claim by using the P–value method of testing hypotheses. Assume that the sample is a simple random sample selected from a normally distributed population. Include the hypotheses, the test statistic, the p–value, and your conclusion.

7) A cereal company claims that the mean weight of the cereal in its packets is at least 14 oz. The weights (in ounces) of the cereal in a random sample of 8 of its cereal packets are listed below.

14.6 13.8 14.1 13.7 14.0 14.4 13.6 14.2

At the 0.01 significance level, test the claim that the mean weight is at least 14 ounces.

Construct the indicated confidence interval for the difference between the two population means. Assume that the assumptions and conditions for inference have been met.

8) A researcher was interested in comparing the GPAs of students at two different colleges. Independent random samples of 8 students from college A and 13 students from college B yielded the following GPAs.

College A	College B	
3.7	3.8	2.8
3.2	3.2	4.0
3.0	3.0	3.6
2.5	3.9	2.6
2.7	3.8	4.0
3.6	2.5	3.6
2.8	3.9	
3.4		

Determine a 95% confidence interval for the difference, $\mu_1 - \mu_2$, between the mean GPA of all college A students and the mean GPA of all college B students.

Assume that the assumptions and conditions for inference with a two–sample t–test are met. Test the indicated claim about the means of the two populations.

9) The Better Cookie Company claims its chocolate chip cookies have more chips than another chocolate chip cookie. 120 Better Cookies and 100 of the other type of cookie were randomly selected and the number of chips in each cookie was recorded. The results are as follows.

	Better	Another
Mean number of chips	7.6	6.9
Standard deviation	1.4	1.7

At the 2% level of significance, test the claim that the population of Better Cookies has a higher mean number of chips.

Use the paired t–interval procedure to obtain the required confidence interval for the mean difference. Assume that the conditions and assumptions for inference are satisfied.

10) A test of writing ability is given to a random sample of students before and after they completed a formal writing course. The results are given below. Construct a 99% confidence interval for the mean difference between the before and after scores .

Before	70	80	92	99	93	97	76	63	68	71	74
After	69	79	90	96	91	95	75	64	62	64	76

Use a paired t–test to perform the required hypothesis test for two population means. Assume that the conditions and assumptions for inference are satisfied.

11) Ten different families are tested for the number of gallons of water a day they use before and after viewing a conservation video. Do the data suggest that the mean amount after the viewing differs from the mean amount before the viewing? Perform a paired t–test at the 5% significance level.

Before	33	33	38	33	35	35	40	40	40	31
After	34	28	25	28	35	33	31	28	35	33

1) (1.1, 1.9)
2) Conditions:
* Randomization condition: We have a random sample of U.S. companies.
* 10% condition: The sample is less than 10% of the total number of U.S. companies.
* Nearly Normal condition: We know the distribution is skewed, but we have a large enough sample size to proceed..
We can find a t–interval for mean number of vacation days.

We know: $n = 60$, $\overline{y} = 22$, $s = 9$, and $SE(\overline{y}) = \frac{9}{\sqrt{60}} = 1.16$.

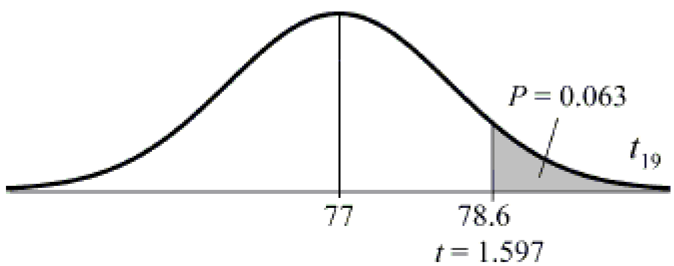
Our confidence interval has the form $\overline{y} \pm t^*_{n-1}\frac{s}{\sqrt{n}}$. We have $t^*_{59} = 2.009$ (we actually use the critical value for a t with 50 degrees of freedom). Our 95% confidence interval is then $22 \pm 2.009\frac{9}{\sqrt{60}} = 22 \pm 2.33$, or 19.7 to 24.3.

We are 95% confident that the interval 19.7 to 24.3 contains the true mean number of vacation days that are given by U.S. companies.
If many random samples of size 60 were taken, 95% of the confidence intervals produced would contain the actual mean number of vacation days offered by U.S. companies.
3) The researcher is 95% confident that the estimate of 118 mmHg differs from the true mean by at most 2.4 mmHg.
4) H_0: $\mu = 77$ years. The mean life life expectancy of the insurance company's patients is 77 years.
H_A: $\mu > 77$ years. The mean life expectancy of the insurance company's patients is greater than 77 years.
* Randomization condition: The records from the insurance company were randomly sampled.
* 10% Condition: 20 records represent less than 10% of the company's records.
* Nearly Normal condition: The histogram of the ages at death is unimodal and reasonably symmetric. This is close enough to Normal for our purposes

Under these conditions, the sampling distribution of the mean can be modeled by Student's t with $df = n - 1 = 20 - 1 = 19$.

We will use a one–sample t–test for the mean.
We know: $n = 20$, $\overline{y} = 78.6$ years , and $s = 4.48$ years. So, $SE(\overline{y}) = \frac{s}{\sqrt{n}} = \frac{4.48}{\sqrt{20}} = 1.002$ years.

$t = \frac{\overline{y} - \mu_0}{SE(\overline{y})} = \frac{78.6 - 77}{1.002} = 1.597$.
$P = P(t_{19} > 1.597) = 0.063$.



The P–value of 0.063 is fairly high, so we fail to reject the null hypothesis. The insurance company shouldn't need to increase their premiums because there is little evidence to indicate that people who buy their policies are living longer than before.

5) We wish to find the sample size, n, that would allow a 95% confidence level for the mean life expectancy of a policy holder from the insurance company to have a margin of error of only one year.

First estimate:
$ME = z^* \times SE(\overline{y})$
$1 = 1.96 \times \dfrac{4.48}{\sqrt{n}}$
$n = 77.1 \approx 78$
Although not necessary, since 78 is quite large, we could find a better estimate using $t^*_{75} = 1.992$, from Table T.
$ME = t^*_{75} \times SE(\overline{y})$
$1 = 1.992 \times \dfrac{4.48}{\sqrt{n}}$
$n = 79.6 \approx 80$

6) Based on this sample, we are 90% confident that Tasty Choice doughnuts will last on average between 1.4 and 2.5 days longer than Sugar Twist doughnuts.

7) H_0: $\mu = 14$ ounces
H_A: $\mu < 14$ ounces
Test statistic: $t = 0.41$
P–value = 0.6524
Fail to reject H_0: $\mu \geq 14$ ounces. There is not sufficient evidence to warrant rejection of the claim that the mean weight is at least 14 ounces.

8) (–0.81, 0.15)

9) H_0: $\mu_1 - \mu_2 = 0$ H_A: $\mu_1 - \mu_2 > 0$
Test statistic $t = 3.291$, P–value $= 5.94 \times 10^{-4}$, DF = 191.61
Reject the null hypothesis. There is sufficient evidence to support the claim that the population of Better Cookies has a higher mean number of chips.

10) (–0.5,4.5)

11) $H_0 : \mu_d = 0$
$H_A : \mu_d \neq 0$
Test statistic: $t = 2.894$
P–value = 0.0178
Reject H_0. At the 5% significance level, the data provide sufficient evidence to conclude that the mean amount after the viewing differs from the mean amount before the viewing.