STATISTICS STAT2033

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?

4) A government report on housing costs says that single–family home prices nationwide are skewed to the right, with a mean of $235,700 and standard deviation of $25,500. We collect price data from a random sample of 50 homes in Orange County, California. Suppose we hope improve our estimate by choosing a new sample. How many home prices must we survey to have 90% confidence of estimating the mean local price to within $2000?

Interpret the given confidence interval.

5) 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.

6) 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.

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.

7) 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

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.

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.

9) 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

Answer the question.

10) The human resources department of a large, well–known telecommunications firm would like to know the job satisfaction of the employees working at its company. This HR department hires an outside, impartial consultant to sample its management and its non–management employees separately. For this scenario, what is the best sampling method to use?

11) Management at a retail store is concerned about the possibility of drug abuse by people who work there. They decide to check on the extent of the problem by having a random sample of the employees undergo a drug test. Several plans for choosing the sample are proposed. Name the sampling strategy in each.

a. Randomly select an employee classification and test all the people who work in that classification– supervisors, full–time clerks, part–time clerks, and maintenance staff.
b. Choose the fourth person that arrives to work for each shift.
c. There are four employee classifications: supervisors, full–time clerks, part–time clerks, and maintenance staff. Randomly select ten people from each category.
d. Each employee has a three–digit identification number. Randomly choose 40 numbers.

12) Explain why the first plan suggested above, sampling an entire classification, might be biased. Be sure to name the kind(s) of bias you describe.

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 ± 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) $ME = t^*_{n-1} \times SE(\overline{y})$ or $2000 = (1.676)\left(\frac{25500}{\sqrt{n}}\right)$ or $\sqrt{n} = 21369$ or n = 456.6

We need to sample approximately 457 home prices. (Or 440 using $z^* = 1.645$)
5) 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.
6) H_0: μ = 14 ounces
H_A: μ < 14 ounces
Test statistic: t = 0.41
P–value = 0.6524
Fail to reject H_0: μ ≥ 14 ounces. There is not sufficient evidence to warrant rejection of the claim that the mean weight is at least 14 ounces.
7) (–0.5,4.5)
8) (–0.81, 0.15)
9) $H_0 : \mu_d = 0$
$H_A : \mu_d \neq 0$
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.
10) Stratified sampling
11) a. cluster
b. systematic
c. stratified
d. simple
12) Undercoverage might occur. The classification selected is not representative of all employees or may not represent the drug usage by employees at the retail store. Perhaps the pressure of being a supervisor leads to more drug use than other classifications.