100

Fall 2018 Math227 Test #4 (ch9, 10, 11) Name ARCADIA FLEEGE

4 DEC 2018

Show all work neatly and systematically for full credit. Total points:100.
Note: for hypothesis testing and confidence interval, make sure to show all steps. Make sure to write conclusion sentences.

Provide an appropriate response.

1) (10) At a local store, 65 female employees were randomly selected and it was found that their mean monthly income was $625 with a standard deviation of $121.50. Seventy-five male employees were also randomly selected and their mean monthly income was found to be $667 with a standard deviation of $168.70. Test the hypothesis that male employees have a higher monthly income than female employees. Use $\alpha = 0.01$.

Female	Male
$n_1 = 65$	$n_2 = 75$
$\bar{x}_1 = 625$	$\bar{x}_2 = 667$
$s_1 = 121.5$	$s_2 = 168.7$

Claim: $\mu_1 < \mu_2$

H_0 $\mu_1 = \mu_2$ H_1 $\mu_1 < \mu_2$

$\alpha = 0.01$ Left tailed

$t_{crit} = \text{invT}(0.01, 64) = -2.386$ ✓

TS $t = \dfrac{(625-667)}{\sqrt{\dfrac{121.5^2}{65} + \dfrac{168.7^2}{75}}} = -1.705$

$-1.705 > -2.386$ ✓

Fail to reject H_0

There is not sufficient evidence to support the claim that male employees have higher mean income than female employees.

2) (10) In a random sample of 300 women, 49% favored stricter gun control legislation. In a random sample of 200 men, 28% favored stricter gun control legislation. Construct a 98% confidence interval for the difference between the population proportions $p_1 - p_2$. Can you conclude that proportion of woemen favored stricter gun control legislation is higher than that for men?

Women	Men
$n_1 = 300$	$n_2 = 200$
$\hat{p}_1 = 0.49$	$\hat{p}_2 = 0.28$
$\hat{q}_1 = 0.51$	$\hat{q}_2 = 0.72$

✓

$CL = 98\%$ $\frac{\alpha}{2} = 0.01$

$z_{\frac{\alpha}{2}} = \text{invNorm}(0.01, 0, 1, R) = 2.326$

$E = 2.326\sqrt{\dfrac{(0.49)(0.51)}{300} + \dfrac{(0.28)(0.72)}{200}} = 0.0998$ ✓

$\hat{p}_1 - \hat{p}_2 = 0.21$

CI: $0.1102 < p_1 - p_2 < 0.310$

Because the interval is positive, we can be 98% confident that the proportion of women in favor of gun control is higher than the proportion of men in favor of gun control.

3) (10) A researcher was interested in comparing the amount of time spent watching television by women and by men. Independent simple random samples of 34 women and 37 men were selected, and each person was asked how many hours he or she had watched television during the previous week. The summary statistics are as follows.

Women	Men
$\bar{x}_1 = 12.5$ hrs	$\bar{x}_2 = 14.3$ hrs
$s_1 = 3.9$ hrs	$s_2 = 5.2$ hrs
$n_1 = 34$	$n_2 = 37$

Construct a 99% confidence interval for $\mu_1 - \mu_2$. Can you conclude that the mean amount of time spent watching television for women is less than the mean amount of time spent watching television for men.?

$\frac{\alpha}{2} = 0.005$

$t_{\frac{\alpha}{2}} = invT(0.005, 33) = -2.73$

$E = 2.73\sqrt{\frac{3.9^2}{34} + \frac{5.2^2}{37}} = 2.96$ ✓

$\bar{x}_1 - \bar{x}_2 = -1.8$ ✓

CI:

$-4.76 < \mu_1 - \mu_2 < 1.16$ ✓

Because the interval includes zero, we cannot conclude that the mean amount of time spent watching TV is different between men and women. ✓

4) (10) In a random sample of 500 people aged 20–24, 110 were smokers. In a random sample of 450 people aged 25–29, 65 were smokers. Test the claim that the proportion of smokers age 20 –24 is higher than the proportion of smokers age 25 –29. Use a significance level of 0.01.

20-24 y/o	25-29 y/o
$n_1 = 500$	$n_2 = 450$
$x_1 = 110$	$x_2 = 65$
$\hat{p}_1 = \frac{110}{500} = 0.22$	$\hat{p}_2 = \frac{65}{450} = 0.144$

$3.019 > 2.326$ Reject H_0 ✓

There is sufficient evidence to support the claim that the proportion of smokers 20-24 is higher than the proportion of smokers age 25-29.

Claim: $p_1 > p_2$ ✓

$H_0\ p_1 = p_2 \quad H_1\ p_1 > p_2$

Right tailed $\alpha = 0.01$

$Z_{crit} = invNorm(0.01, 0, 1, R) = 2.326$

TS $Z = \frac{0.22 - 0.144 - 0}{\sqrt{\frac{0.816 \times 0.184}{500} + \left(\frac{0.184 \times 0.816}{450}\right)}} = 3.019$ ✓

$\bar{p} = \frac{110 + 65}{500 + 450} = 0.184$ ✓

$\bar{q} = 1 - \bar{p} = 0.816$

5) (16 points) A coach uses a new technique in training middle distance runners. The times for 8 different athletes to run 800 meters before and after this training are shown below. Assume samples have been randomly selected from normally distributed populations.

Athlete	A	B	C	D	E	F	G	H
Time before training (seconds)	118.7	111.1	115.1	109.4	117.9	111.3	116.2	109
Time after training (seconds)	119.3	109.8	112.7	110.2	116.1	111.4	112.6	105.1

a. (10) Using a 0.05 level of significance, test the claim that the training helps to improve the athletes' times for the 800 meters.

$n=8$

$\bar{d}=1.4375$

$S_d = 1.826$

Claim: $\mu_d > 0$

$H_0\ \mu_d = 0 \quad H_1\ \mu_d > 0$ ✓

Right tailed $\alpha = 0.05$

ts: $t = \dfrac{1.4375-0}{\left(\frac{1.826}{\sqrt{8}}\right)} = 2.227$ ✓

$t_{crit} = \text{invT}(0.95, 7) = 1.895$ /

$2.23 > 1.895$ Reject H_0 ✓

There is sufficient evidence to support the claim that the training helps improve atheletes' times. ✓

b. (6) Construct a 90% confidence interval for the mean difference of the "before" minus "after" times.

$E = 1.895\left(\frac{1.826}{\sqrt{8}}\right) = 1.223$ ✓

CI : 1.4375 ± 1.223

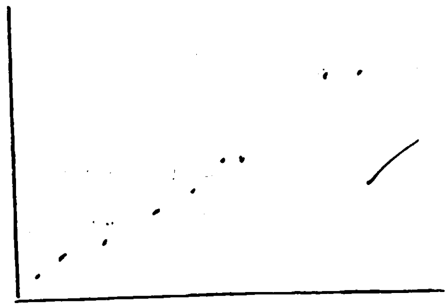
$0.214 < \mu_d < 2.661$ ✓

Because the confidence interval is positive, ✓ we can conclude that the training does reduce atheletes's times in the 800 meters dash. ✓

6) (16 points) The data below are the average one–way commute times (in minutes) of selected students during a summer literature class and the number of absences for those students for the term.

Commute time (min), x	70	83	89	88	86	96	73	98	78
Number of absences, y	7	11	14	14	12	19	8	19	9

a. (3) Draw a scatterplot.



b. (2) Find the linear correlation coefficient.

$r = 0.980$

c. (3) Test whether there is a linear correlation between the average one–way commute times (in minutes) and the number of absences.

$n = 9 \quad \alpha = 0.05 \quad r_{crit} = 0.666$

$|0.980| > 0.666$

There is a linear correlation between commute length & number of absences

d. (2) Find the regression equation.

-1 $\hat{y} = 0.490x - 25.37$

e. (2) Interpret the slope and the y–intercept, if possible.

$\frac{0.490}{1}$

For every 1 minute increase in commute time, number of absences is expected to increase by 0.490

0 is not a reasonable value of x

f. (2) Predict the number of absences for the summer literature class for a student who has one–way commute time of 80 minutes.

$\hat{y} = 0.49(80) - 25.37 = 13.83 \approx \cancel{14}$ absences

-1

g. (2) How many percent of variation in the number of absences can be explained by the linear relation between the one–way commute time and the number of absences? (The coefficient of determination.)

$r^2 = 0.960$

96% of variation can be explained by the linear relation

-2

7) (11) A teacher figures that final grades in the chemistry department are distributed as: A, 25%; B, 25%; C, 40%; D, 5%; F, 5%. At the end of a randomly selected semester, the following number of grades were recorded. Test the claim that the grade distribution for the department is different than expected. Use $\alpha = 0.01$.

Grade	A	B	C	D	F
Number	36	42	60	8	14

GRADE	O	P	E	$\frac{(O-E)^2}{E}$
A	36	0.25	40	0.4
B	42	0.25	40	0.1
C	60	0.4	64	0.25
D	8	0.05	8	0
F	14	0.05	8	4.5
Σ	160	1.0	160	5.25

$E = np$ TS $\chi^2 = \Sigma\left(\frac{(O-E)^2}{E}\right) = 5.25$

$df = 4$ $\chi^2_{crit} = 13.277 > 5.25$ fail to reject H_0

$\alpha = 0.01$

$H_0 = P(A) = 0.25$ $P(B) = 0.25$

$P(C) = 0.4$ $P(D) = P(F) = 0.05$

H_1 = At least one probability is different than expected

There is insufficient evidence to support the claim that the distribution of grades is different than expected.

(6) Use the given data to find the best predicted value of the response variable.

8) Based on the data from six students, the regression equation relating number of hours of preparation (x) and test score (y) is $\hat{y} = 67.3 + 1.07x$. The same data yield $r = 0.224$ and $\bar{y} = 75.2$. What is the best predicted test score for a student who spent 4 hours preparing for the test? (Note that you need to check wither there is a linear correlation between the variables first)

$r = 0.224$

$\alpha = 0.05$ $n = 6$

$r_{crit} = 0.811$

$|0.224| < 0.811$

There is no linear correlation between hours of preparation and test scores by this data.

The best predictor for the score of a student spending 4 hours to prepare is $\bar{y} = 75.2$.

9) (11) A medical researcher is interested in determining if there is a relationship between adults over 50 who exercise regularly and low, moderate, and high blood pressure. A random sample of 236 adults over 50 is selected and the results are given below. Test the claim that exercise and the blood pressure are independent. Use $\alpha = 0.01$.

H_0 exercise & blood pressure are independent

H_1 exercise & blood pressure are dependent

Blood Pressure	Low	Moderate	High	
Reg. Exercise	35 28.95	62 65.65	25 27.4	122
No Reg. Exercise	21 27.05	65 61.35	28 25.6	114
	56	127	53	236

$\alpha = 0.01$

$df = (2-1)\times(3-1) = 2$

$\chi^2_{crit} = 9.21$

$$TS\ \chi^2 = \frac{(35-28.95)^2}{28.95} + \frac{(62-65.65)^2}{65.65} + \frac{(25-27.4)^2}{27.4} + \frac{(21-27.05)^2}{27.05} + \frac{(65-61.35)^2}{61.35} + \frac{(28-25.6)^2}{25.6}$$

$$= 1.264 + 0.203 + 0.210 + 1.353 + 0.217 + 0.225 = 3.472$$

$3.472 < 9.21$ fail to reject H_0

There is sufficient evidence to support the claim that exercise and blood pressure are independent.