MATH 338

EXAM 2 – WRITTEN PORTION

TUESDAY, JULY 18, 2017

Your name: ______________________

Your scores (to be filled in by Dr. Wynne):

Problem 1:	____/10
Problem 2:	____/7.5
Problem 3:	____/9.5
Problem 4:	____/9
Total:	____/36

You have 60 minutes to complete this exam. This exam is closed book and closed notes with the exception of your formula sheet.

For full credit, show all work except for final numerical calculations (which can be done using a scientific calculator).

1. Do college men or women tend to experience more severe negative consequences if they engage in excessive drinking? Suls and Barton (2003) asked 111 male and 112 female college students at a large Midwestern university to indicate a numerical response to this question on a scale from 1 (women experience more negative consequences) to 7 (men experience more). The midpoint of the scale was 4.

On average, women rated the statement at 3.04, with a standard deviation of 1.39. Men's ratings averaged 3.45 and had a standard deviation of 1.33.

A) [6 pts] Evaluate the statistical validity of the following claims in the paper: "Men and women thought that women would suffer more [negative] consequences as a result of excessive drinking; men's mean responses differed significantly from the midpoint, $t(110) = 25.50$, $p < .001$, as did women's, $t(111) = 24.19$, $p < .001$."

1 pt men and women did both think that women would suffer more negative consequences because both sample means are closer to 1 (women experience more) than to 7 (men experience more)

1 pt given their real-world claims, the authors correctly performed two one-sample tests of H_0: $\mu = 4$ vs. H_a: $\mu \neq 4$ (I gave you credit for doing a two-sample test based on an incorrect reading of the claims)

2 pts both test statistics are wrong; the correct values are

For men: t = (3.45 – 4)/(1.33/sqrt(111)) = -4.35   For women: t = (3.04 – 4)/(1.39/sqrt(112)) = -7.31

1 pt the degrees of freedom are as stated in the claim (for men: 110, for women; 111)

1 pt $p < .001$ implies an alpha of 0.001, which gives a critical value of about 3.39 for a two-sided test. Since both t-statistics are less than -3.39, it is appropriate to say that the p-value is less than 0.001

1 pt therefore, the real-world claims that men's mean responses differed significantly from the midpoint and women's mean responses differed significantly from the midpoint are backed up by their statistical analysis

B) [2 pts] Check <u>all</u> assumptions necessary to perform the statistical test(s) done in part (A).

1 pt the samples are not SRS, but they are independent and representative of college-aged men and women in the US (or not representative of all men and women, I gave credit for either)

1 pt the sample size for the men is 111 > 40 and for women is 112 > 40, so both sample sizes are large enough to do one-sample t-tests; or, if you did a two-sample test, total sample size is 111 + 112 = 223 > 40, so the sample size is large enough to use this test

C) [2 pts] Based on parts (A) and (B), do you trust the authors' conclusions? Why or why not?

Argue that either it is appropriate to do a one sample t-test (based on part B) or give a reason why it is not. If it is appropriate, even though their test statistics are wrong, their p-value comparison to α is not, and their conclusions are consistent with this approach (and therefore trustworthy). If it is not appropriate, then we cannot trust their conclusions even though the math is (except for the t-statistic) appropriate.

2. A 2007 article in the Journal of Chemical Education suggests using a one-way ANOVA to compare three different methods of standardization used when determining the concentration of an analyte in an unknown solution. Students are asked to find the concentrations of three hydrocarbons in a single unknown vial using gas chromatography, standardizing their results using each of the three methods. A one-way ANOVA analysis is used to determine whether the concentration depends on the method.

A) [2 pts] Based on the description of this study, assuming that the uncertainty in concentration is normally distributed, is a one-way ANOVA appropriate to analyze the results? Why or why not?

1 pt No, it is not appropriate

1 pt One-way ANOVA requires independent cases and a single factor with more than two levels. However, the measurements are not independent (because we only have a single sample), and/or we have two different factors (the type of hydrocarbon and the method used).

B) [2.5 pts] Fill in the missing values in the example ANOVA table from the paper: 0.5 pts per value

Source	DF	Sum of Sq.	Mean Sq.	F	Pr (>F)
Between Groups	2	1784.7	892.35	19.50	0.0024
Within Groups	6	274.5	45.75		
Total	8	2059.2	257.4		

C) [3 pts] The paper makes the following statements about one-way ANOVA. For each statement, if it is statistically valid, write, “Valid.” Otherwise, change the statement to make it valid.

1 pt per statement

“The three quantitative methods should give similar values for the three analytes … This is the null hypothesis (H_0: $\bar{x}_1 = \bar{x}_2 = \bar{x}_3$).”

The null hypothesis should read $\mu_1 = \mu_2 = \mu_3$. It should not contain $\bar{x}$’s, which stand for sample means.

“If students find that their results do not support the null hypothesis then they use the Tukey multiple-comparison method for obtaining confidence intervals for the differences between means.”

This is valid. Alternatively, I gave credit for suggesting a different multiple-comparison method.

“The results in [the table in part (B)] show that the null hypothesis should be rejected because the calculated P value of 0.002 is less than the set α value of 0.05.”

This is valid. The p-value is reported correctly, and this is an appropriate decision given the p-value.

Problem 3. In 2009, a group of mice spent 3 months aboard the International Space Station. Three mice survived, and were compared to a group of three mice who spent those 3 months in equivalent cages, on Earth.

In one experiment, after all the mice were sacrificed, researchers measured the red blood cell count (in millions per microliter of blood) in the six mice. The table below summarizes the findings:

Group	Mean	Standard Deviation
Space	11.27	0.39
Earth	10.02	1.68

A) [6 pts] Assuming red blood cell counts are normally distributed, construct and interpret a 99% confidence interval for the difference of the population mean red blood cell counts for mice in space vs. mice on Earth.

1 pt use a two-sample t confidence interval

1 pt the point estimate is (11.27 – 10.02) = 1.25

1 pt the standard error is $\sqrt{\frac{0.39^2}{3} + \frac{1.68^2}{3}} = 0.996$

1 pt the degrees of freedom are min(n_1 – 1, n_2 – 1) = 2, so the t* critical value for 99% confidence is 9.925

1 pt the CI is 1.25 ± (9.925)(0.996) = (-8.63, 11.13)

1 pt We are 99% confident that the true difference between the population mean number of red blood cells in space mice vs. Earth mice is between -8.63 and 11.13 millions per microliter of blood.

B) [2 pts] Based on your answer to part (A), can you conclude (at the 1% significance level) that mice in space and mice on Earth have different population mean blood cell counts? Why or why not?

NO (0.5 pts) – based on our CI, the space mice might have 11.13 millions per microliter more, or the Earth mice might have 8.63 millions/microliter more, or they might have the same amount. Since the confidence interval contains 0, we cannot reject the null hypothesis H_0: $\mu_1 - \mu_2 = 0$ in favor of the two-sided alternative at the 1% significance level. (1.5 points for explanation)

C) [1.5 pts] If more space mice had survived (and assuming that the number of space mice would be compared with an equal number of control Earth mice), which of the following would have changed in your analysis from part (A)? Circle all that would change. 0.5 pts for each possible answer below

Confidence Level Degrees of Freedom Critical Value

4. Assume that when your lab's centrifuge is working, and you set it to 4000 revolutions per minute (RPM), the actual angular velocity is normally distributed with mean 4000 RPM and (theoretical) standard deviation 50 RPM. You take 6 independent measurements and perform a one-sided hypothesis test, using α = 0.05, to determine whether the centrifuge is slower than it should be.

A) [5 pts] What is the power of this hypothesis test to detect the specific alternative that the centrifuge is 1.5% too slow (i.e., that the true population mean velocity is 3940 RPM)?

1 pt the critical value for $\alpha = 0.05$ and a one-sided alternative is z = 1.645, according to the table, so the critical region on the z-scale is Z < -1.645

1.5 pts the critical region on the original scale is then $\bar{x} < -(1.645)\left(\frac{50}{\sqrt{6}}\right) + 4000$, or or 3966.42

1.5 pts we wish to find $P(\bar{x} < 3966.42)$ when $\bar{x} \sim N\left(3940, \frac{50}{\sqrt{6}}\right)$, which is $P\left(Z < \frac{3966.42 - 3940}{\frac{50}{\sqrt{6}}}\right)$, which we find to be $P(Z < 1.294)$, or just over 0.8997 based on the table.

1 pt therefore the power is approximately 90%

B) [2 pts] For this test, what would be a Type I Error? What is the probability of committing it?

1 pt a Type I Error would be claiming that the centrifuge is defective when it is working correctly

(half credit for giving the definition, rather than applying it to this specific test)

1 pt the probability of committing a Type I Error is $\alpha = 0.05$

C) [2 pts] For this test, what would be a Type II Error? What is the probability of committing it?

1 pt a Type II Error would be claiming that the centrifuge is working correctly when it is too slow

(half credit for giving the definition, rather than applying it to this specific test)

1 pt the probability of committing a Type II Error is $\beta = 1 - power$ = 0.1 or about 10%

Extra Space. The tables below show a number of critical values z for the standard normal variable $Z \sim N(0,1)$ and the corresponding cumulative proportions, corresponding to P(Z $\leq$ z).

z-score	Cumulative Proportion
-3.00	0.0013
-2.50	0.0062
-2.00	0.0228
-1.65	0.0495
-1.28	0.1003
-1.00	0.1587
-0.67	0.2514

z-score	Cumulative Proportion
0.67	0.7486
1.00	0.8413
1.28	0.8997
1.65	0.9505
2.00	0.9772
2.50	0.9938
3.00	0.9987

Refer to the following tables for t* and z* critical values for confidence intervals:

Df	C = 0.90 (90%)	C = 0.95 (95%)	C = 0.98 (98%)	C = 0.99 (99%)	C = 0.999 (99.9%)
1	6.314	12.71	31.82	63.66	636.62
2	2.920	4.303	6.965	9.925	31.60
3	2.353	3.182	4.541	5.841	12.92
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.869
6	1.943	2.447	3.143	3.707	5.959
10	1.812	2.228	2.764	3.169	4.587
≈30	1.697	2.042	2.457	2.750	3.646
≈50	1.676	2.009	2.403	2.678	3.496
≈100	1.660	1.984	2.364	2.626	3.390
≈1000	1.646	1.962	2.330	2.581	3.300

	C = 0.90 (90%)	C = 0.95 (95%)	C = 0.98 (98%)	C = 0.99 (99%)	C = 0.999 (99.9%)
z* values	1.645	1.960	2.326	2.576	3.291

For a two-sided hypothesis test, use the column corresponding to C = 1 – α

For a one-sided hypothesis test, use the column corresponding to C = 1 - 2α

The rest of this space to be used for extra work: