MATH 338

FINAL EXAM

CONCEPTUAL PORTION

MON/WED, MAY 14/16, 2018

Your name: ______________________

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

Problem 1:	____/3.5
Problem 2:	____/8.5
Problem 3:	____/5
Problem 4:	____/12
Total:	____/29

You have 110 minutes to complete this exam. This exam is closed book and closed notes with the exception of your two sheets of notes (front and back).

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

Problem 1 [0.5 pts each]. Circle the single correct answer to each multiple-choice question.

A) For a given sample size *n*, our study has a power of 0.81 at a significance level of $\alpha = 0.05$. What is the probability that we commit a Type II Error when we collect the data and run the hypothesis test?

a. 0.05

b. 0.19

c. 0.81

d. 0.95

B) Which of the following steps could we take to be guaranteed to increase the power of the study we are planning to perform in part (A)?

a. Decrease the sample size

b. Decrease the significance level

c. Decrease the standard deviation of the variable we are measuring

d. All of the above would increase the power of the study

C) Addison's disease is a rare disorder of the adrenal gland, often caused by either tuberculosis or autoimmune destruction. Researchers would like to test whether the cause of the disease (tuberculosis, TB, or autoimmune, AI) affects the mean concentration of aldosterone, an adrenal hormone, in the blood. Which of the following is the most appropriate null and alternative hypothesis pair for this test?

a. H_0: $\mu_{TB} - \mu_{AI} = 0$ vs. H_a: $\mu_{TB} - \mu_{AI} \neq 0$

b. H_0: $\mu_{TB} - \mu_{AI} \neq 0$ vs. H_a: $\mu_{TB} - \mu_{AI} = 0$

c. H_0: $\mu_{TB\text{-}AI} = 0$ vs. H_a: $\mu_{TB\text{-}AI} \neq 0$

d. H_0: $\mu_{TB\text{-}AI} \neq 0$ vs. H_a: $\mu_{TB\text{-}AI} = 0$

Note: there's nothing in here to indicate matching, but I gave you half-credit if you circled (c)

D) Suppose that a random sample of 70 Addison's disease patients in the US is taken, and in 64, the disease had an autoimmune cause. Which of the following methods would be most appropriate to estimate the proportion of Addison's disease patients in the US whose cause was autoimmune?

a. One sample t confidence interval

b. Large sample z confidence interval

c. Binomial exact confidence interval

d. All three methods are equally appropriate

Problem 1 (continued). Answer some more multiple-choice questions. Again, each question has a single correct answer that you should circle.

E) If a simple linear regression model reports an r^2 value of 0.81, which of the following is not true?

a. A linear relationship between the explanatory and response variables is appropriate

b. The correlation between the explanatory and response variables is either 0.9 or -0.9

c. 81% of the variation in the response variable is explained by variation in the explanatory variable

d. All of the above statements are true

Note: answer (a) is sometimes true and sometimes false. I gave you half-credit if you circled (d).

F) A 95% confidence interval for the slope in a simple linear regression model is reported as (-0.24, 0.62). Which of the following is not true?

a. The least-squares estimate of the slope is approximately 0.19

b. At the 5% significance level, the slope is not significantly different from 0

c. This interval is the same at all values of the explanatory variable in the sample

d. All of the above statements are true

G) We create a multiple linear regression model predicting our response variable y from two numerical variables, x_1 and x_2. The reported model takes the form of the equation:

$$\hat{y} = 0.32 + 0.537\, x_1 - 0.812\, x_2$$

Which of the following is a correct and complete interpretation of the value 0.537 in the above model?

a. For every one-unit increase in $\hat{y}$, x_1 is predicted to increase by 0.537 units

b. For every one-unit increase in x_1, y is predicted to increase by 0.537 units

c. The correlation between x_1 and $\hat{y}$ is 0.537

d. None of the above answers is a correct and complete interpretation of the value 0.537.

Problem 2. Suppose that we come up with a new method of computing 95% confidence intervals. We wish to calibrate our method using simulation; that is, we randomly generate *n* values from a population distribution with known μ and σ, and we repeat this process many times (like we did in Lab 6). Suppose we create 1000 independent randomly simulated samples and use each one to construct a 95% confidence interval.

A) [1 pt] Assuming our method truly does have a confidence level of 95%, how many of our 1000 confidence intervals do we expect to contain the true mean, μ?

We expect (95%) x 1000 = 950 CIs to contain the true mean

B) [2 pts] Again, assuming we have a true confidence level of 95%, find the approximate sampling distribution of $\hat{p}$, the proportion of our 1000 intervals that contain the true mean.

0.5 pts approximately normal and 0.5 pts centered at p = 0.95

1 pt sd = sqrt(0.95*0.05/1000) = 0.007

Or N(0.95, 0.007), or N(95%, 0.7%)

1 point total for giving me the correct binomial distribution for number of intervals instead

C) [3.5 pts] Using your answer to Part B (make up an answer if you have to), what is the approximate probability that fewer than 94.1% of our confidence intervals contain the population mean?

0.5 pts recognize that this is a cumulative proportions problem

2 pts P(p-hat $<$ 0.941) = P(z $<$ (0.941 – 0.95)/(0.007) = -1.31)

1 pt convert from z-score to cumulative proportion: roughly 0.1 or 10%

OR get exact probability from a calculator: 0.096 or 9.6%

D) [2 pts] Suppose that one of our simulated 95% confidence intervals is (2.125, 2.835). Can you find the proportion of population means that are between 2.125 and 2.835? If so, find it. If not, explain why not.

0.5 pts NO

1.5 pts we only have a single population mean. Either it is in the interval or it is not. Since this problem never gave you the value of the population mean, you cannot evaluate whether it is in the interval.

Problem 3. [5 pts] In a 2015 study, researchers investigated the performance of the BioRadNS1 Ag STRIP, a rapid diagnostic test for detecting dengue fever, and compared it to their own model that used only patient age, patient white blood cell count, and patient platelet count. This question asks only about the NS1 test, not the researchers' model.

1692 Vietnamese children were confirmed to have dengue fever by a composite of laboratory tests. Of those, 1192 were classified as dengue-positive by the NS1 test. Of the 4015 children suspected not to have dengue by the laboratory tests, 32 were classified as dengue-positive by the NS1 test. From this sample, estimate the sensitivity, specificity, positive predictive value, and negative predictive value of the NS1 test. Round all estimates to the nearest percent.

Sensitivity: 70% Specificity: 99%

PPV: 97% NPV: 89%

You may use the rest of this space to show work:

For calculating sensitivity, specificity, PPV and NPV, I am showing a table below because it is the most space-efficient way to do this. However, this was not the only way to find the values. 1 pt awarded for setting up any appropriate method (e.g., two-way table, tree diagram, Bayes's rule).

	Test +	Test -	Total
Dengue +	1192	500	1692
Dengue -	32	3983	4015
Total	1224	4483	5707

1 pt awarded – half for correct values for any of the four formulas, half for correct formula – for each blank filled in

Sensitivity = 1192/1692 = 70.45% or about 70%

Specificity = 3983/4015 = 99.20% or about 99%

PPV = 1192/1224 = 97.39% or about 97%

NPV = 3983/4483 = 88.85% or about 89%

Problem 4. A real scientific journal actually published a statistical analysis of Urban Dictionary. In this problem, you will interpret some of the study's methods and analysis.

For parts A-B, the following information may be useful.

The authors of the study sampled 4,465 definitions by dividing headwords (words/phrases that are being defined) into different groups depending on the number of unique definitions given, randomly sampling up to 200 headwords in each group, and then sampling up to 3 definitions for each headword.

A) [0.5 pts] True or False: The sample generated by the above description is a simple random sample?

FALSE

B) [1 pt] What is a case in the sample generated by the above description? How many cases are there?

A case is a single definition. There are 4,465 cases.

Or, the cases are the 4,465 definitions sampled by the authors.

For parts C-F, the following information and contingency table may be useful.

The authors crowdsourced ratings for whether each definition "describes the meaning of the word, expresses a personal opinion, [or] both" (variable TYPE, with levels meaning, opinion, and both) and whether the "word in the described [definition is appropriate to use with that definition] in a formal conversation" (variable FORMAL, with levels yes, unclear, and no).

		FORMAL			
TYPE		Yes	Unclear	No	Total
	Meaning	459	73	3039	3571
	Opinion	164	26	358	548
	Both	86	22	238	346
	Total	709	121	3635	4465

C) [1 pt] Use the table to find the marginal distribution of the variable FORMAL.

Marginal distribution = 709/4465 Yes (okay for a formal setting), 121/4465 Unclear (not sure if okay),

3635/4465 No (definitely not okay for a formal setting), or 15.9% Yes, 2.7% Unclear, 81.4% No

D) [2 pts] What proportion of opinion-only definitions would clearly not be appropriate for use in a formal conversation?

1 pt recognize that this is a conditional proportion given opinion-only

1 pt 358/548 = 0.653 or 65.3% of opinion-only definitions

Problem 4 (continued). The contingency table is reproduced below:

		FORMAL			
TYPE		Yes	Unclear	No	Total
	Meaning	459	73	3039	3571
	Opinion	164	26	358	548
	Both	86	22	238	346
	Total	709	121	3635	4465

E) [3 pts] If the authors were to perform a test to determine whether the type of definition and its appropriateness in a formal setting are independent, identify the contribution to the test statistic from definitions that express meaning and are definitely okay for use in formal conversation.

1.5 pts expected counts = (709)*(3571)/(4465) = 567.04

1.5 pts contribution = (O-E)^2/E = (459 – 567.04)^2/(567.04) = 20.585

F) [1 pt] How many degrees of freedom are in the distribution that test statistic comes from?

(r-1)(c-1) = 4

For parts G-I, the following excerpt from the paper may be useful.

"Table 4 reports the offensiveness scores separated by whether the definitions describe a meaning, opinion or both. An one-way ANOVA test indicates a slight significant difference ($F(2, 3963) = 2.766$, $p<0.1$)."

G) [1 pt] Is the offensiveness score a numerical or categorical response variable? How do you know?

Numerical, because the response variable in a one-way ANOVA has to be numerical, and TYPE is the categorical explanatory variable. We are assuming the authors performed the correct type of inference.

H) [1.5 pts] Not all of the 4,465 definitions were rated for offensiveness. How many of them were?

1 pt recognize that we can add the two degrees of freedom attached to the F statistic to get n-1

0.5 pts 2 + 3963 = 3965, so 3965 + 1 = 3966 definitions were rated

I) [1 pt] Explain in your own words what "p < 0.1" means.

If the null hypothesis is true (in this case that the three types of definitions are equally offensive, on average), then we would have gotten a test statistic this big or bigger in less than 10% of samples.

Extra Space. The tables below show a number of 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 table for χ^2 critical values:

Degrees of freedom	$\alpha = 0.05$	$\alpha = 0.01$
1	3.84	6.63
2	5.99	9.21
3	7.81	11.34
4	9.49	13.28

This space reserved for extra work: