MATH 338

FINAL EXAM

WRITTEN PORTION

MON/WED/THURS, DEC 11/13/14, 2017

Your name: ________________________

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

Problem 1:	____/7.5			
Problem 2:	____/4.5			
Problem 3:	____/4	Option:	Software	Written
Problem 4:	____/5	Option:	Software	Written
Total:	____/21			

You have the remainder of the exam period 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. In some strange alternate universe where Dr. Wynne teaches thousands of statistics students, he curves final grades to be normally distributed with mean 75 and standard deviation 10.

A. [2 pts] If the top 5% of students get A's, what curved final grade is necessary to obtain an A?

0.5 pts work in inverse normal proportions framework

0.5 pts: z score for top 5% is z-score for a cumulative proportion of 0.95, so about 1.65

1 pt $\sigma z+\mu = (10)(1.65) + 75 = 91.5$

Students must get a 91.5 or higher curved final grade to obtain an A

B. [1.5 pts] Suppose Dr. Wynne takes a random sample of 5 students and computes their average curved grade. Obtain the sampling distribution of the sample mean grade of these 5 students.

0.5 pts normal, 0.5 pts mean = population mean = 75, 0.5 pts sd = σ/sqrt(n) = 10/sqrt(5) = 4.47

Alternatively, N(75, 4.47) is full credit

C. [1 pt] Before curving, Dr. Wynne wants to test whether the mean raw grade in the class is 75. He takes a SRS of 5 students. Circle the correct answer to each prompt below:

Dr. Wynne uses a significance level of 0.01 and obtains a p-value of 0.044. What should he do?

Reject H_0 Accept H_0 Fail to Reject H_0 Nothing

If the true mean raw grade in the class is 70, Dr. Wynne makes a…

Correct Decision Type I Error Type II Error Sandwich

I am guaranteed to make a Type II Error. I am not guaranteed to make a sandwich afterwards.

E. [3 pts] If 40% of students who got an A after the curve are female, and 30% of students without A's are female, what proportion of female students in the class got an A?

0.5 pts recognize this as a Bayes's Rule problem

2 pts use a two-way table such as shown below, tree diagram, or Bayes's rule directly:

	Female	Not Female	Total
A	20	30	50
Not A	285	665	950
Total	305	695	1000

0.5 pts 20/305 or approximately 6.56% of female students in the class got an A

Problem 2. Dmitrieva and Burg (2014) investigated the effect of sodium concentration in blood serum on the levels of von Willebrand factor (a blood clotting protein) in blood plasma. The table below is taken from the paper's supplemental information:

Table S5. Multiple regression analysis of plasma level of vWF (transformed) with plasma Na^+ and glucose as predictor variables (ARIC Study, n = 14,679)

Independent variable	Regression coefficient b_j	SE of b_j	t	P
Intercept	3.799	0.188	20.2	<0.001
Na^+, mmol/L	0.005	0.0013	3.7	<0.001
Glucose, mmol/L	0.0344	0.0014	24.1	<0.001

Na^+ and glucose are significant predictors of vWF. $F(2, 14{,}676) = 289$; $P < 0.001$.

A. [1.5 pts] Write the multiple regression equation predicting von Willenbrand factor (vWF) level from sodium (Na^+) levels and glucose levels.

vWF = 3.799 + 0.005 (sodium) + 0.0344 (glucose)

B. [1.5 pts] Identify and interpret the slope corresponding to sodium (Na^+) level.

0.5 pts circle the number 0.005 or otherwise indicate that the slope is 0.005

0.5 pts When Na^+ level increases by 1 mmol/L, the expected plasma vWF level increases by 0.005 units…

0.5 pts …holding the glucose level constant (or after controlling/accounting for glucose level)

C. [1 pt] In the output above, circle the value of the test statistic for the ANOVA F test for this model. State below the type of distribution it comes from and the degrees of freedom for that distribution.

0.5 pts circle the number 289 in the output above

0.5 pts it comes from an F distribution with 2 and 14,676 degrees of freedom

D. [0.5 pts] Circle below the pair of variables that should be the *least* correlated, if our interpretation of the multiple linear regression model is statistically correct.

vWF and Na^+ vWF and glucose Na^+ and glucose

Problem 3. [4 pts] **<u>Circle on the title page whether to grade this problem or the software version of this problem.</u>**

A 2017 study investigated the prevalence of drunkenness at Swedish football (soccer) matches. The researchers defined a blood alcohol content of at least 0.1% to be "highly intoxicated." In a random sample of 4420 spectators, 395 had a blood alcohol content (BAC) of at least 0.1%. Construct and interpret a 95% confidence interval for the proportion of all Swedish football (soccer) match spectators who are highly intoxicated.

0.5 pts do a large sample one-proportion z confidence interval

0.5 pts identify the point estimate = $\hat{p} = \frac{395}{4420}$ = 0.089

0.5 pts compute the standard error = $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{\left(\frac{395}{4420}\right)(1-\frac{395}{4420})}{4420}}$ = 0.00429

0.5 pts identify that z* = 1.96…

0.5 pts …so the margin of error is (1.96)(0.00429) = 0.008

1.5 pts: We are 95% confident that the true proportion of highly intoxicated spectators at Swedish football (soccer) matches is between 0.081 and 0.097.

OR

We are 95% confident that between 8.1% and 9.7% of all Swedish football (soccer) spectators are highly intoxicated.

Problem 4. This problem expands on the gout problem in the Software portion of this exam. You do not need the data set to answer these questions. **Circle on the title page whether to grade this problem or the software version of this problem.**

A. [1.5 pts] Briefly describe how you would design a study to determine whether urate-lowering treatment causes the reduction in serum uric acid levels you observed in the software problem.

0.5 pts each for describing the terms below (whether the name is used is irrelevant):

1) Control: one group gets the treatment and one group doesn't

2) Randomization: study participants are assigned at random to one of the groups

0.5 pts for describing one or both of the terms below (whether the name is used is irrelevant):

3) Repetition: Sufficiently large sample size in each group to allow differences to "average out"

4) Blind or Double-blind: participants in the control group get a placebo and don't know which drug they are taking. For double-blind, doctors and other experimenters don't know either.

For parts B through D, refer to the table below. The row variable is the sex of the gout patient and the column variable shows whether the patient was on a urate-lowering treatment.

	Treatment	No Treatment	Total
Female	32	17	49
Male	384	93	477
Total	416	110	526

B. [1 pt] A chi-squared hypothesis test is performed to analyze the table. What is the null hypothesis for the appropriate chi-squared test?

1 pt H_0: Sex and Treatment are independent (or not associated)

C. [1 pt] Fill in the table below with the expected counts (round fractional counts to 2 decimal places) under the null hypothesis.

	Treatment	No Treatment	Total
Female	38.75	10.25	49
Male	377.25	99.75	477
Total	416	110	526

D. [1.5 pts] Compute the test statistic for this hypothesis test. Identify the type of distribution it comes from and the degrees of freedom for that distribution.

1 pt: $\chi^2 = (32 - 38.75)^2/38.75 + (17 - 10.25)^2/10.25 + (384 - 377.25)^2/377.25 + (93 - 99.75)^2/99.75 = 6.20$

0.5 pts: it comes from a χ^2 distribution with (2-1)*(2-1) = 1 degree of freedom

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 tables for t* and z* critical values for confidence intervals:

Degrees of freedom	C = 0.90 (90%)	C = 0.95 (95%)	C = 0.98 (98%)	C = 0.99 (99%)
1	6.314	12.71	31.82	63.66
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
≈30	1.697	2.042	2.457	2.750
≈50	1.676	2.009	2.403	2.678
≈100	1.660	1.984	2.364	2.626
≈1000	1.646	1.962	2.330	2.581
≈4000	1.645	1.961	2.327	2.577

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

For a two-sided hypothesis test, use the column corresponding to $C = 1 - \alpha$

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

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