Stat 201 Final

Student Name ____

Student Number ______________________________________

You have exactly 180 minutes to complete this exam. This test has 14 pages including this one.

Only calculators that are not communication devices are allowed for electronics. This means no translators and no phones. You are also allowed one double-sided aid sheet of letter (A4) size paper.

- Show your work whenever appropriate. It shows understanding, and that's what's being tested.

- Ask for extra scrap paper if space is an issue.

- If you get stuck on a part, don't abandon the question. Often later parts can be answered without earlier ones.

- Try not to panic, it rarely helps.

Question	1	2	3	4	5	Row Total	
Out of	6	6	7	6	7	32	

Question	6	7	8	9	10	Row Total	Total
Out of	6	8	8	8	5	35	67

Question 1:

Pr(G) = 0.5,   Pr(H) = 0.35,   and events G and H are **MUTUALLY EXCLUSIVE.** Find the following probabilities **(6 points)**

A) Pr(G and H) (2 points)

B) Pr(G or H)   (2 points)

C) Pr(G and NOT H) Hint: Try to visualize with a Venn Diagram. **(2 points)**

Question 2: There's a teaching method called Flipped Classroom in which 'homework' is done in class, and 'lectures' are studied from recordings at home. One class was taught under the traditional method, and the other under 'flipped. The results are below. **(6 points)**

Flipped Classroom grades	Traditional Classroom grades
$\bar{X}_{Flip} = 500$	$\bar{X}_{Trad} = 560$
$S_{Flip} = 300$	$S_{Trad} = 50$
$n_{Flip} = 100$	$n_{Trad} = 20$

A) Should this data be analyzed with a paired, or an independent test? What two pieces of information tell you this? (3 points)

B) Is pooling standard deviation appropriate for this data? Why or why not? (2 points)

c) How many degrees of freedom would you use for this t-test? (1 point)

Question 3: Consider the following summary data for an independent samples two-sample t-test. (7 points)

Group A	Group B
$\bar{X}_A = 44$	$\bar{X}_B = -23$
$S_{pooled} = 77$	
$n_A = 32$	$n_B = 18$

A) Find the test statistic against the two-tailed null that $\mu_B - \mu_A = 0$ (4 pts)

B) If the groups were balanced, what would the standard error be? (3 pts)

Question 4: Consider the regression output for daughter's heights as a function of their parents' heights. **(6 points total)**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-12.878	9.779		-1.317	[redacted]
	Average height of parents	1.193	.145	.800	8.230	

(1 pt) Write the regression equation.

(2 pts) Using the t-table and the regression output provided. Test the null hypothesis that the slope is zero using the significance level $\alpha = 0.05$. Explain your steps and include a conclusion.

(3 pts) Using the t-table provided. Find the 99% confidence interval of the slope.

Question 5: Consider the regression output for perceived stress during a task as a function of total mastery. Mastery scores for the population of interest are between 20 and 80. **(7 points total)**

Model		Unstandardized Coefficients		Standardized Coefficients	
		B	Std. Error	Beta	t
1	(Constant)	50.971	1.273		40.035
	Total Mastery	-.625	.061	-.424	-10.222

a. Dependent Variable: Total perceived stress

(1 pt) Write the regression equation.

(3 pts) Use the regression equation to predict the stress of someone with a mastery of 40. Is this prediction reasonable? Why or why not?

(3 pts) Use the regression equation to predict the stress of someone with a mastery of 140. Is this prediction reasonable? Why or why not?

Question 6: You're trying to see if race and smoking status are independent from a cross tab. Below are the results from the chi-squared test. **(6 points total)**

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3.112[a]	4	[redacted]
Likelihood Ratio	3.092	4	
N of Valid Cases	254		

a. 2 cells (20.0%) have expected count less than 5. The minimum expected count is 1.53.

a) Test if smoking status and race are independent at a = 0.01. Show work. **(2 pts)**

b) Give a range for the p-value of the chi-squared test. **(1 pt)**

c) If the differences between the observed values and expected values were larger such that the chi-squared value **<u>three times</u>** as large, would you reject the null hypothesis that they are independent at alpha = 0.01? Show work. **(2 pts)**

d) Are there are any indications of potential problems with the chi-square analysis from this output? Explain briefly. **(1 pt)**

Question 7: Angus is conducting a health survey to investigate the association between heavy smoking and heavy drinking. He surveys a simple random sample of 100 people and gets the following crosstab: **(8 points total)**

heavydrink * heavysmoke Crosstabulation

Count

		heavysmoke		
		No	Yes	Total
heavydrink	No	66	5	71
	Yes	23	[redacted]	29
Total		89	11	100

a) Under the assumption of independence between variables. Estimate the probability that someone from this population is both a heavy smoker AND a heavy drinker? **(2 pts)**

b) What is the EXPECTED number of respondents in the (heavydrink=YES, heavysmoke = YES) cell? What is the OBSERVED number of respondents? **(2 pts)**

b) What is the chi-squared contribution of the cell for people that are NOT heavy drinkers, and NOT heavy smokers? **(4 pts)**

Question 8: The BC government is trying to determine if there are significant differences in the average calcium in milk from different parts of the province. It has taken samples from 4 different parts of the province. **(8 points total)**

Calcium Level of All

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	29.501	3	9.834	.324	.808
Within Groups	1728.732	57	30.329		
Total	1758.233	60			

A) Is a t-test appropriate? Why or why not? **(2 pts)**

B) Is a regression appropriate? Why or why not? **(2 pts)**

C) Test the null hypothesis that the mean calcium levels are the same in each area at
alpha = 0.01. **(2 pts)**

D) How much of the variance in calcium can explained by the differences between the four areas? Show your work. **(2 pts)**

Question 9: Consider the ANOVA table of the responses from a randomized controlled trial **(8 points total)**

ANOVA

Time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	91.467	2	45.733	4.467	.021
Within Groups	276.400	27	10.237		
Total	367.867	29			

Draw the design diagram for this experiment. **(4 pts)**

From this table, is there evidence of **ANY** differences in the mean responses between the three treatment groups? Explain (2 pts)

From this table (ONLY use this table), can we determine that **ALL** three treatment groups have different mean responses? Explain. (2 pts)

Question 10: Consider this summary table of the three group means from a randomly controlled trial. **(5 points)**

Descriptives

Time

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Beginner	[redacted]	27.2000	3.04777	.96379	25.0198	29.3802	22.00	33.00
Intermediate	[redacted]	23.6000	3.30656	1.04563	21.2346	25.9654	18.00	29.00
Advanced	[redacted]	23.4000	3.23866	1.02415	21.0832	25.7168	18.00	29.00
Total	30	24.7333	3.56161	.65026	23.4034	26.0633	18.00	33.00

Assume the observations are independent, and that there is no pairing structure. Is it appropriate to use ANOVA to analyze this data? Explain. (Hint: What is the other assumption for ANOVA?) (2 points)

If this is a balanced design, what are the group sizes? (1 pt)

If we were only interested in whether there was a difference (in either direction) between the Beginner and the Advanced group, what kind of test would this be? (2 pts, 0.5 for each part of the description)

Scrap paper for extra work, or pictures of clouds, or whatever. You can tear off the page with the chi-squared and t tables.

Chi-squared table.

df	P = 0.05	P = 0.01	P = 0.001
1	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.35	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.13
9	16.92	21.67	27.88
10	18.31	23.21	29.59
11	19.68	24.73	31.26
12	21.03	26.22	32.91
13	22.36	27.69	34.53
14	23.69	29.14	36.12
15	25.00	30.58	37.70
16	26.30	32.00	39.25
17	27.59	33.41	40.79
18	28.87	34.81	42.31
19	30.14	36.19	43.82
20	31.41	37.57	45.32

Table T Critical Values of the t Distribution

df	One-Tail = .4	.25	.1	.05	.025	.01	.005	.0025	.001	.0005
	Two-Tail = .8	.5	.2	.1	.05	.02	.01	.005	.002	.001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Source: From *Biometrika Tables for Statisticians,* Vol. 1, Third Edition, edited by E. S. Pearson and H. O. Hartley, 1966, p. 146. Reprinted by permission of the Biometrika Trustees.