Stat 201 Final Exam

Student Name _____

Student Number _____

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

Only calculators (any kind) are allowed for electronics. That means no digital translators and no phones. You are allowed to bring a double-sided A4 size aid-sheet. You are also allowed to bring and use a fidget device and provided that you are not disturbing other students.

Protips:
- Show your work whenever appropriate. It shows understanding, and that's what's being tested.
- If you get stuck on a part, don't abandon the question. Often later parts can be answered without earlier ones.
- Ask about any words you don't know. If it's not related to statistics, you will likely get an answer.
- Trust yourself. You have studied; you are an expert in this.
- Even experts check their work.
- Try not to panic, it rarely helps.

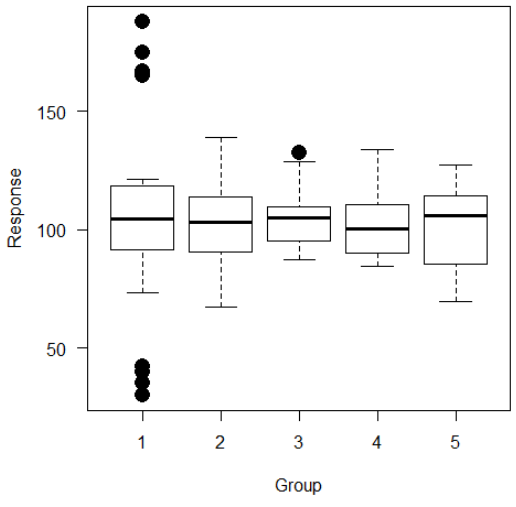
Good luck!

Question	1	2	3	4	5	6	TOTAL
Out of	5	6	7	9	5	4	36

Question	7	8	9	10	11	SUBTOTAL	TOTAL
Out of	7	7	9	8	8	39	75

Question 1 - ANOVA 1 (5 points)

Consider the summary information below.




```
> by(y,grp,mean)
grp: 1
[1] 103.9212
----------------
grp: 2
[1] 103.7821
----------------
grp: 3
[1] 105.1211
----------------
grp: 4
[1] 101.5719
----------------
grp: 5
[1] 101.6215
```

```
> by(y,grp,sd)
grp: 1
[1] 40.5822
----------------
grp: 2
[1] 18.15574
----------------
grp: 3
[1] 12.45372
----------------
grp: 4
[1] 13.29503
----------------
grp: 5
[1] 17.42002
```

A) What is the problem with conducting an ANOVA on this dataset? How could they be fixed? **(2 pts)**

B) If we wanted to conduct a t-test on each pair of means in this situation, a multiple testing correction would be necessary. If the experiment-wide alpha is set to 0.05, what would be the alpha for each pair using the Bonferroni correction? **(1 pt)**

C) If we also included the measurements of the five groups before each of the treatments, could we treat this as a one-way ANOVA with ten groups? Why or why not? **(2 pts)**

Question 2 - ANOVA 2 (6 points)

Consider the ANOVA table of four groups of data with sizes 8, 12, 9, and 14 respectively.

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	3	269.1	89.707	0.3957	0.7568
Residuals	[redacted]	8841.2	226.698		

A) State the null hypothesis in mathematical symbols. **(2 pts)**

B) Test the null hypothesis at significance level alpha = 0.01 **(1 pt)**

C) What proportion of the variance in the response can be explained by the groups? **(2 pts)**

D) How many df (for the residual / between groups) are there? (1 pts)

Question 3 - Regression 1 (7 points)

Consider this regression output for the number of prey in an area

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  404.670     21.532  18.794 [redacted]
predators     -4.403      1.295  -3.401 [redacted]
```

A) State in WORDS the null hypothesis being tested in the (intercept) line. **(2 pts)**

B) State in WORDS the null hypothesis being tested in the prey line **(2 pts)**

C) Test the prey line null hypothesis at alpha = 0.05. State your conclusion. **(2 pts)**

D) What can you say for certain about the sample correlation between predators and prey? **(1 pts)**

Question 4 - Regression 2 (9 points)

Consider the following regression output, which describes the amount of rage in an area (in gigafuries / square km) as a function of the number of Canada geese around.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   31.913     19.618   1.627     0.11
geese         10.419      1.192   8.739 1.72e-11 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.8 on 48 degrees of freedom
Multiple R-squared:  0.6141,    Adjusted R-squared:  0.606
```

A) What proportion of variance in rage is explained by amount of geese? **(1 pts)**

B) Predict the rage value for a new area with 12 geese. **(2 pts)**

C) Is the prediction in Part B reasonable in real-world terms? Why or why not? **(2 pts)**

D) Estimate the response value for the average of all areas with 20 geese. **(2 pts)**

E) Between the values in parts B and D, which prediction / estimate will have a wider confidence interval, assuming the same level of confidence alpha, and that both values of geese are near the middle of the observations? Why? **(2 pts)**

Question 5 - Crosstabs 1 – (5 Points)

Consider the crosstab below.

OBSERVED				Total
	100	81	64	245
			144	274
		62	72	213
Total	260	192	280	732

A) Fill in the missing observed values. **(1 pt)**

EXPECTED				Total
	87.02	64.26		245
	97.32	71.87		274
	75.66	55.87		213
Total	260	192	280	732

B) Fill in the missing expected values. **Use the expected count formula, show work below (2 pts)**

Chi-Squared Contribution		
		9.42
2.74	7.28	14.65
0.15	0.67	1.10

C) Fill in the missing chi-squared contributions. **(2 pts)**

Question 6 - Crosstabs 2 – (4 Points)

Consider the table of chi-squared contributions below

Chi-Squared Contribution		
0.04	1.33	2.41
0.07	1.31	2.11

A) Find the chi-squared statistic and the degrees of freedom. (2 pts)

B) Use the table below to give a range for the p-value of the chi-squared test (2pts)

DEGREES OF FREEDOM	**PROBABILITY**										
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
				Nonsignificant						**Significant**	

Source: R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed., Table IV, Oliver & Boyd, Ltd., Edinburgh, 1963, by permission of the authors and publishers.

Question 7 - Crosstabs 3 – (7 points)

Consider the 3x3 crosstab below.

OBSERVED	Bulbasaur	Squirtle	Charmander	Total
Age 19 or less	3	7	4	14
Age 20-29	4	3	1	8
Age 30 or more	5	4	6	15
Total	12	14	11	37

EXPECTED	Bulbasaur	Squirtle	Charmander	Total
Age 19 or less	4.54	5.30	4.16	14
Age 20-29	2.59	3.03	2.38	8
Age 30 or more	4.86	5.68	4.46	15
Total	12	14	11	37

A) Is it appropriate to apply the chi-squared procedure? Why or why not? **(2 pts)**

B) What could be done to fix the problem? Describe your solution in words. **(2 pts)**

C) Use this solution produce a new observed table that fixes the problem. **(3 pts)**

Question 8 - Probability – (7 points)

A) Find Pr(A and (not B)), if A and B are **independent**. (2 pts)
Pr(A) = 2/5 Pr(B) = 1/3

B) Find Pr(A and (not B)), if A and B are **mutually exclusive.** (2 pts)
Pr(A) = 2/5 Pr(B) = 1/3

C) Find Pr(A **OR** (not B)), if A and B are **independent**. (2 pts)
Pr(A) = 0.25 Pr(B) = 0.3

D) Pr(A or B), if the set {A, B} is **exhaustive.** (1 pts)
Pr(A) not given Pr(B) not given

Question 9 – T-Tests: Two ponds with ducks are in town, but in one of the ponds only natural food is allowed to be given to the ducks. At the other pond it's mostly bread. We catch some ducks from each pond and weigh them. Assume by some miracle that the ducks we catch are a simple random sample. See below.**(9 points)**

Ducks with natural food	Ducks with bread
$\bar{x}_{Nat}$= 1530 grams	$\bar{x}_{Bread}$ = 1403 grams
S_{nat} = 120	S_{bread} = 182
n_{nat} = 9	n_{bread} = 14

A) Name a formal hypothesis test you could conduct to test the assumption of normality within each group. (**1 pts)**

B) Find the test statistic for the appropriate t-test. Show all your work. **(8 pts)**

Question 10 - Descriptives: Consider the frequency table below. Some parts are left blank as a workspace for you **(8 points)**

x	Frequency (n)	Cumulative Frequency
-4	47	
0	31	
5	62	
20	135	
99	25	
TOTAL		

A) Find the sample mean, $\bar{X}$ **(2 points)**

B) Find the 80^{th} percentile **(2 points)**

C) Determine if the x = 99 values be considered outliers. **(4 points)**

Question 11 – Experimental Design: Consider a randomized block design for a comparative random experiment. **(8 points)**

A) Draw the design diagram for such a design when the blocking variable has 3 categories, and the treatment variable has 2 categories. The blocks are of size 60, 120, and 200 respectively, and the treatment assignment is balanced. Make sure to label everything. **(6 pts)**

B) What is the major advantage of setting up a randomly assigned experiment over the convenience of an observational study? **(1 pt)**

C) What is the typical advantage of balancing the treatment assignments? **(1 pt)**

This page left obnoxiously blank (hey, look, the finish line!)

Table T Critical Values of the *t* Distribution

df	One-Tail = .4 Two-Tail = .8	.25 .5	.1 .2	.05 .1	.025 .05	.01 .02	.005 .01	.0025 .005	.001 .002	.0005 .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