Statistics 302 Final Exam

Name: _____

Student Number: _____

Instructions

Fill in the information requested above.

Calculators and one 8.5×11 inch formula sheet are allowed. Write your answers **directly on the exam paper**. If you run out of space, you may write your answers on the back of the page, but must **clearly indicate** that you have done so.

Caution: In accordance with the University's Code of Academic Integrity and Good Conduct (S10), academic dishonesty in any form will not be tolerated. Prohibited acts include, but are not limited to the following:

- the unauthorized sharing of material such as notes or calculators;
- concealing information pertaining to the examination in the examination room, or in washrooms or other places in the vicinity of the examination room;
- using course notes or any other aids not approved by an Instructor during an examination;
- the unauthorized possession or use of an examination question sheet, an examination answer book, or a completed examination;
- using, or attempting to use, another student's answers;
- providing answers to other students;
- failing to take reasonable measures to protect answers from use by other students; or
- impersonating a candidate or being impersonated in an examination.

1. (1 mark) Which of the following summary statistics measures the spread of a distribution: first quartile, third quartile, inter-quartile range, Median.

2. (1 mark) Briefly, define the sampling distribution of a statistic.

3. (15 marks) An individual's critical flicker frequency is the highest frequency at which the flicker in a flickering light source can be detected. At frequencies above the critical frequency, the light source appears to be continuous even though it is actually flickering. This investigation recorded critical flicker frequency and iris colour of the eye for 19 subjects. A summary of the flicker frequencies by group (eye colour) is as follows:

Colour	mean	sd	n
Blue	28.16667	1.527962	6
Brown	25.58750	1.365323	8
Green	26.92000	1.843095	5

   (a) (2 mark) Would an analysis of these data be a fixed or random effects ANOVA? Justify your answer.

   (b) (1 mark) Is this a balanced or unbalanced design?

(c) (1 mark) Based on the above summary, comment on the homoscedasticity assumption for one-way ANOVA.

(d) (3 marks) Write down the model for a fixed effects one-way ANOVA for these data. Define any variables you use.

(e) (2 marks) Fill in the missing information in the following ANOVA table to four digits. You may write your answers in the blanks provided.

	Df	Sum Sq	Mean Sq	F value
Colour	__	23.00	_____	_____
Residuals	__	38.31	2.394	

(f) (3 marks) Use Table A4 at the back of the exam to test the significance of the eye colour effect at the 5% level. Report the critical value you used and the conclusion of the test.

(g) (3 marks) Use Bonferroni's method to find 95% confidence intervals for the pair-wise differences between means. Report your answers to four digits. For part marks report the critical value from Table A2 that you used and the standard errors for each comparison. The summary statistics are repeated here for convenience.

Colour	mean	sd	n
Blue	28.16667	1.527962	6
Brown	25.58750	1.365323	8
Green	26.92000	1.843095	5

4. (4 marks) A football team records data on the length of punts made by 13 players at a tryout for the team. The distance measure for each punter is the average of ten punts. They also record the following information on each player:

Variable	Description
Hang	Time in air in seconds
RStrength	Right leg strength in pounds
LStrength	Left leg strength in pounds
RFlexibility	Right leg flexibility in degrees
LFlexibility	Left leg flexibility in degrees
OStrength	Overall leg strength in pounds

You perform stepwise regression with the BIC criterion to build a predictive model of distance. The largest model in your search includes all main effects; the smallest model includes only an intercept.

(a) (1 mark) What is the BIC penalty term for model selection in this example? Report your answer to four digits.

(b) (2 marks) After several iterations of stepwise selection you obtain a model that includes RStrength, RFlexibility, LFlexibility and OStrength. Here is a summary of the next application of ADD1 and DROP1:

```
Distance ~ RStrength + RFlexibility + LFlexibility + OStrength

               Df Sum of Sq    RSS    BIC
- RStrength     1    221.49 1728.3 73.829
- RFlexibility  1    228.58 1735.3 73.882
- LFlexibility  1     64.68 1571.5 72.592
- OStrength     1    660.99 2167.8 76.774
<none>                     1506.8 74.611
+ Hang          1      5.81 1501.0 77.126
+ LStrength     1      4.45 1502.3 77.137
```

What action would you take next? Justify your answer.

(c) (1 mark) After finishing stepwise selection we obtain the following model:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.7676    24.9926   0.511   0.6205
RStrength    0.5563     0.2104   2.644   0.0246
OStrength    0.2717     0.1003   2.709   0.0220
```

From this model, what is the predicted Distance for a punter with RStrength 170 and OStrength 266? Report your answer to four digits.

5. (24 marks) A softdrink vendor collects data on the relationship between the time in minutes a delivery takes, and two explanatory variables: (i) the number of cases delivered (Cases), and (ii) the walking distance in feet to make the delivery (Distance). The following summaries are obtained for 25 deliveries:

```
      Time            Cases          Distance
 Min.   : 8.00   Min.   : 2.00   Min.   :  36.0
 1st Qu.:13.75   1st Qu.: 4.00   1st Qu.: 150.0
 Median :18.11   Median : 7.00   Median : 330.0
 Mean   :22.38   Mean   : 8.76   Mean   : 409.3
 3rd Qu.:21.50   3rd Qu.:10.00   3rd Qu.: 605.0
 Max.   :79.24   Max.   :30.00   Max.   :1460.0
```

(a) (1 mark) How would you describe the shape of the Distance variable?

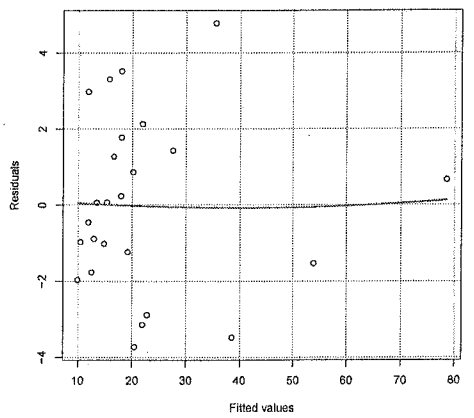
(b) (2 marks) Write out a linear model for mean Time that includes interaction between Cases and Distance. Define any notation you use.

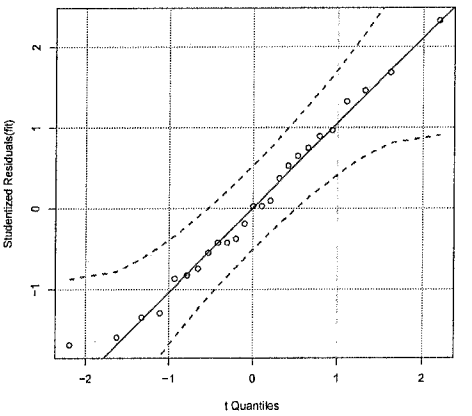
(c) (2 marks) In terms of your model from the previous question, state formal hypotheses for testing for statistical interaction.

(d) (2 marks) Computer software reports the following VIFs. Do you have any concerns about collinearity? If so, why? If not, why not?

```
     Cases      Distance Cases:Distance
  6.932817      4.842433      9.765414
```

(e) (6 marks) The model is fit and yields the following residual plots. For each plot, state the assumptions being checked, and give your opinion about whether or not each assumption is plausible:





(f) (2 marks) A summary of the hat values for the fitted model is as follows:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.04540 0.07806 0.11740 0.16000 0.15540 0.92160
```

Are there any observations that have noteworthy leverage? If so, why? If not, why not?

(g) (2 marks) The Cook's Distance values are:

```
     1      2      3      4      5      6      7      8      9     10     11     12
 0.087  0.001  0.004  0.058  0.004  0.000  0.035  0.005  2.758  0.172  0.192  0.006
    13     14     15     16     17     18     19     20     21     22     23     24
 0.000  0.013  0.013  0.028  0.000  0.054  0.010  0.106  0.017  0.134  0.029  0.070
    25
 0.014
```

Are there any observations with noteworthy influence? If so, why? If not, why not?

(h) (2 marks) A summary of the fitted model is as follows:

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.1390846  1.3997413   5.100 4.73e-05
Cases          1.0144063  0.1912517   5.304 2.93e-05
Distance       0.0058273  0.0033825   1.723 0.099622
Cases:Distance 0.0007419  0.0001750   4.240 0.000366
```

Test for interaction at the 5% level and write a sentence to report your conclusion.

(i) (2 marks) Using the fitted model summary from the previous question and Table A2 at the back of the exam report a 95% confidence interval for the interaction term. Report your answers to 4 digits. The model summary is repeated for convenience.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.1390846	1.3997413	5.100	4.73e-05
Cases	1.0144063	0.1912517	5.304	2.93e-05
Distance	0.0058273	0.0033825	1.723	0.099622
Cases:Distance	0.0007419	0.0001750	4.240	0.000366

(j) (3 marks) Write a sentence to interpret the effect of increasing Distance by 100 feet for a delivery of 10 cases. Report the effect size to four digits.

6. (13 marks) Maple tree seeds look like spinning helicopters when they fall from the tree. A forest scientist studied the relationship between how fast they fall (Velocity) and their size (Load), taking a total of 35 measurements on three trees (12 on two of them and 11 on the third). We will analyze the relationship between Velocity and Load, allowing for different relationships in the three trees.

   (a) (1 mark) Define two dummy variables $X_{2.1}$ and $X_{2.2}$ for the Tree variable, using baseline coding with Tree 1 as the baseline category.

   (b) (2 marks) Using notation from class, state a model in terms of your dummy variables that allows a different linear relationship between Velocity and Load for each of the three trees.

   (c) (1 mark) In terms of your model, state the null hypothesis for the test of parallelism.

   (d) (1 mark) In terms of your model, state the null hypothesis for the test of coincidence.

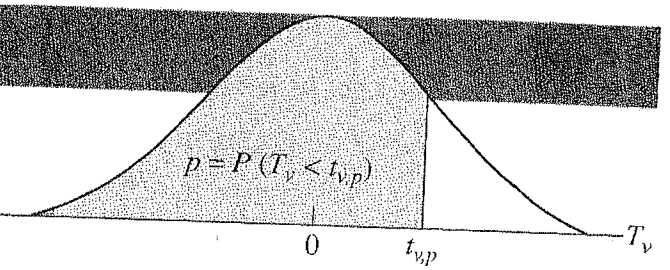
(e) (4 marks) You are given the following sums of squares for models with interaction between Load and Tree, with main effects for Load and Tree and with a main effect for Load only:

Model	SSM	SSE
interactions	0.8929081	0.1654919
main effects	0.8549587	0.2034413
Load only	0.8436365	0.2147635

Using Table A4 at the back of the exam, perform the test of parallelism at the 5% level. Report the extra sum of squares and SSE for your F-statistic to four digits, the value of the F-statistic itself to four digits, the critical value you compare to and the conclusion of your test.

(f) (4 marks) Using Table A4 and the sums of squares from the previous question, perform the test of coincidence at the 5% level. Report the extra sum of squares and SSE for your F-statistic to four digits, the value of the F-statistic itself to four digits, the critical value you compare to and the conclusion of your test.

TABLE A.2 Percentiles of the t Distribution



Student's t distribution

df \ 100p	55	65	75	85	90	95	97.5	99	99.5	99.95
1	0.158	0.510	1.000	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.142	0.445	0.816	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.137	0.424	0.765	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.134	0.414	0.741	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.408	0.727	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.404	0.718	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.402	0.711	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.130	0.399	0.706	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.398	0.703	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.397	0.700	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.396	0.697	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.395	0.695	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.394	0.694	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.393	0.692	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.393	0.691	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.392	0.690	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.392	0.689	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.392	0.688	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.391	0.688	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.391	0.687	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.127	0.391	0.686	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.390	0.686	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.390	0.685	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.127	0.390	0.685	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.390	0.684	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.390	0.684	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.389	0.684	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.127	0.389	0.683	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.389	0.683	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.127	0.389	0.683	1.055	1.310	1.697	2.042	2.457	2.750	3.646
35	0.127	0.388	0.682	1.052	1.306	1.690	2.030	2.438	2.724	3.591
40	0.126	0.388	0.681	1.050	1.303	1.684	2.021	2.423	2.704	3.551
45	0.126	0.388	0.680	1.049	1.301	1.679	2.014	2.412	2.690	3.520
50	0.126	0.388	0.679	1.047	1.299	1.676	2.009	2.403	2.678	3.496
60	0.126	0.387	0.679	1.045	1.296	1.671	2.000	2.390	2.660	3.460
70	0.126	0.387	0.678	1.044	1.294	1.667	1.994	2.381	2.648	3.435
80	0.126	0.387	0.678	1.043	1.292	1.664	1.990	2.374	2.639	3.416
90	0.126	0.387	0.677	1.042	1.291	1.662	1.987	2.368	2.632	3.402
00	0.126	0.386	0.677	1.042	1.290	1.660	1.984	2.364	2.626	3.390
20	0.126	0.386	0.677	1.041	1.289	1.658	1.980	2.358	2.617	3.373
[illegible]	0.126	0.386	0.676	1.040	1.288	1.656	1.977	2.353	2.611	3.361
30	0.126	0.386	0.676	1.040	1.287	1.654	1.975	2.350	2.607	3.352
00	0.126	0.386	0.676	1.039	1.286	1.653	1.973	2.547	2.603	3.345
∞	0.126	0.386	0.676	1.039	1.286	1.653	1.972	2.345	2.601	3.340
	0.126	0.385	0.674	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Table A4. Upper percentage points of the F distribution with ν_1, ν_2 df

(a) 5% points

ν_2	ν_1=df for the numerator 1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882	243.906	245.950	248.013	249.052	250.095	251.143	252.196	253.253	254.314
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5493	8.5265
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0659
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1273	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7331
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6377
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0033