Faculty of Science
FINAL EXAMINATION
MATH 204B
Principles of Statistics 2

Examiner: Professor K.J. Worsley
Associate Examiner: Professor R. Steele

Date: Friday, April 28, 2005
Time: 14:00 - 17:00 hours

INSTRUCTIONS: Answer all questions. An 8.5×11 inch piece of paper, written on both sides, may be brought into the exam. Calculators are permitted. Computer printout, figures and tables are provided at the end of the exam.
Use of a regular and or translation dictionary is permitted.
Each part of each question is worth approximately equal marks.

This exam comprises this cover, 3 pages of questions, 8 pages of computer printout, and 3 pages of tables (15 pages in all).

1. The article "Responsiveness of Food Sales to Shelf Space Changes in Supermarkets" (*J. Mktg. Res.* (1964): 63-67) described an experiment to assess the effect of allotted shelf space sales of "Tang". Six stores were used in the experiment, and six different shelf-space allotments of 6, 9, 12, 15, 18 and 21 feet were tried for 1 week. The author speculated that changes in shelf space would affect sales. Data on the number of containers of Tang sold is given in the table below.

	Shelf space (feet)					
Store	6	9	12	15	18	21
1	30	35	25	27	38	31
2	47	59	43	62	65	48
3	47	55	48	54	36	54
4	29	19	41	27	33	39
5	17	11	25	23	24	26
6	22	9	19	18	25	22

Choose appropriate pairs of models, or T-statistics, to answer the following.

(a) Is there any effect of shelf space (as a factor) on sales, allowing for different sales in different stores? Test this.

(b) The table on the printout labelled "Multiple Comparisons" tests for differences between all pairs of shelf space, allowing for different sales in different stores. Are there any differences?

(c) Is there any effect of shelf space (as a linear covariate) on sales, allowing for different sales in different stores? Test this.

(d) Explain carefully why there is a discrepancy in your answers to (a), (b) and (c).

(e) Does the effect of shelf space (as a linear covariate) differ from one store to another? Test this.

(f) Store 1 is thinking of devoting 10 feet of shelf space to Tang. Use what you think is the most appropriate model to predict the weekly sales of Tang. Give an approximate standard deviation for your prediction.

(g) Tricky question: Why can we *not* test that the effect shelf space (as a factor) differs from one store to another?

2. The so-called fog index is intended to measure how difficult it is to read a piece of writing. It is calculated as $0.4\times$ (average number of words per sentence) + (percentage of words with three or more syllables). Shruptine and McVicker (1981) made a study of the readability of magazine advertisements in which they obtained fog indices from a random sample of six advertisements from *Scientific American* and *Newsweek*, obtaining the results in the table below:

	Fog indices					
Scientific American	15.75	11.55	8.20	9.23	11.16	9.92
Newsweek	9.66	7.67	10.21	5.12	3.12	4.88

Carry out a suitable non-parametric test to see if there is any difference in fog indices between *Scientific American* and *Newsweek*. **There is no printout for this question.**

3. The table below gives the number of patients with malignant melanoma cross classified by Type (the type of abnormal cells making up the melanoma) and Site (where on the body the cancer appeared).

	Site		
Type	Extremities	Head and Neck	Trunk
Hutchinson's melanomic freckle	10	22	2
Intermediate	28	11	17
Nodular	73	19	33
Superficial spreading melanoma	115	16	54

   (a) Estimate the expected number of cases of superficial spreading melanoma in the trunk, assuming that Type and Site are independent (this cell is empty on the printout).

   (b) Is there any evidence that the proportion of types of melanoma is different at different sites? Test this.

4. The data for this question originated from a car's journey log kept by a brother of statistics professor Jack Robinson at Waterloo over a 12 month period. Each time the car was filled with petrol, five variables were measured:

   - Dist, the distance run since the last fill-up, in kilometres.
   - KmPerL, the fuel efficiency, in kilometres per liter.
   - Days, the number of days siince the last fill-up.
   - Month, the month of the year (January=1, ..., December=12).
   - Temp, the average monthly temperature in Waterloo.

   There were 86 measurements in total, and the first 20 appear on the printout. Our aim is to predict the fuel efficiency, KmPerL, as a function of the other variables. Choose appropriate pairs of models, or T-statistics, to test the following.

   (a) Ignoring Temp for the moment, a model was fitted with Month as a factor and Dist and Days as covariates. Is there any effect of evidence of an effect of

      i. Dist?
      ii. Days?
      iii. Month?

      allowing for an effect of the other two.

   (b) Month was replaced by Temp in the model. Is there an effect of Temp, allowing for Dist and Days?

(c) Month was added back into the model, together with Dist, Days and Temp. Does the mean monthly temperature explain all the effect of month, or putting it another way, can the effect of Month be explained solely by a linear effect of Temp?

(d) The graph is a plot of residuals from the model fitted in (b), against the fitted values. Do you think the assumptions of normality and equal variance are satisfied?

(e) What proportion of the variance of KmPerL is explained by the model in (b)?

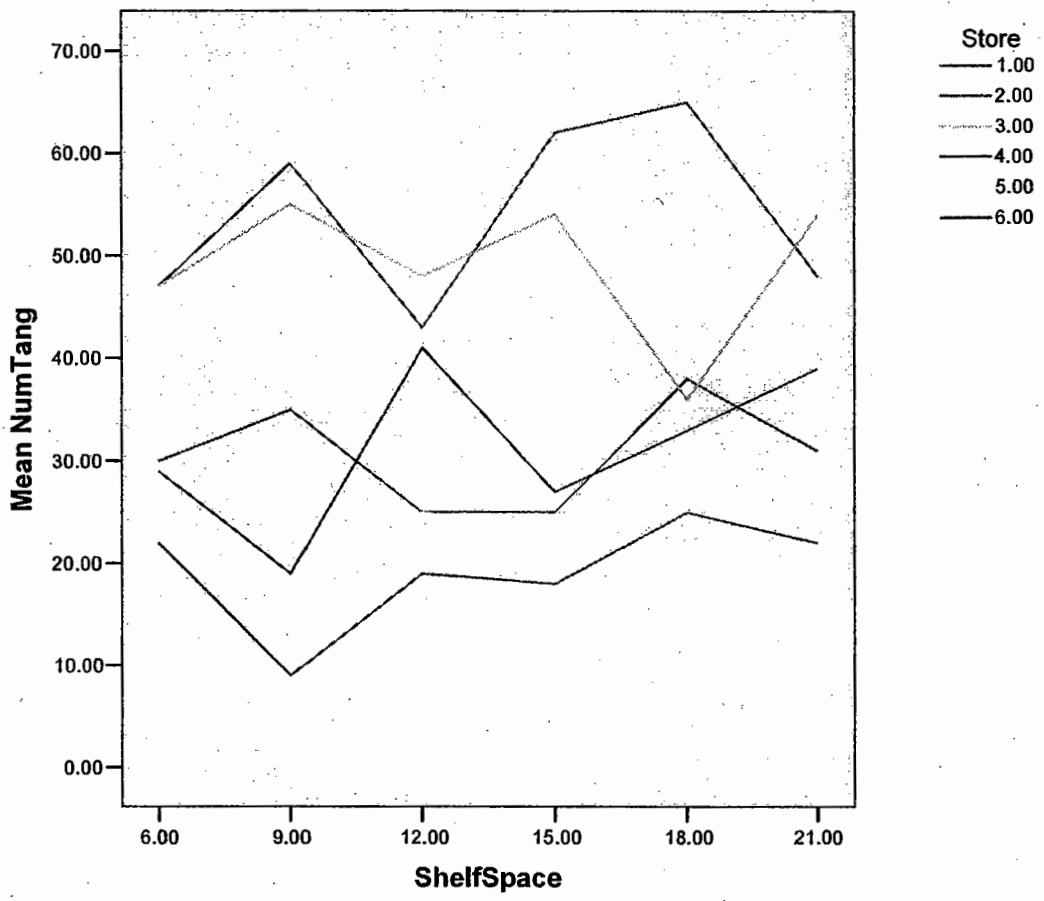
(f) Use a suitable model to predict the fuel efficiency of a long trip of 300 km with 0 days between fill-ups in April (Month=4) when the average temperature is 5.80 degrees. Give an approximate standard deviation for your prediction.

(g) Tricky question: In the last model fitted, explain exactly why [Month=11.00] has a coefficient of 0 and is redundant.

Question 1. Tang

Case Summaries

	NumTang	Store	ShelfSpace
1	30.00	1.00	6.00
2	47.00	2.00	6.00
3	47.00	3.00	6.00
4	29.00	4.00	6.00
5	17.00	5.00	6.00
6	22.00	6.00	6.00
7	35.00	1.00	9.00
8	59.00	2.00	9.00
9	55.00	3.00	9.00
10	19.00	4.00	9.00
11	11.00	5.00	9.00
12	9.00	6.00	9.00
13	25.00	1.00	12.00
14	43.00	2.00	12.00
15	48.00	3.00	12.00
16	41.00	4.00	12.00
17	25.00	5.00	12.00
18	19.00	6.00	12.00
19	25.00	1.00	15.00
20	62.00	2.00	15.00
21	54.00	3.00	15.00
22	27.00	4.00	15.00
23	23.00	5.00	15.00
24	18.00	6.00	15.00
25	38.00	1.00	18.00
26	65.00	2.00	18.00
27	36.00	3.00	18.00
28	33.00	4.00	18.00
29	24.00	5.00	18.00
30	25.00	6.00	18.00
31	31.00	1.00	21.00
32	48.00	2.00	21.00
33	54.00	3.00	21.00
34	39.00	4.00	21.00
35	26.00	5.00	21.00
36	22.00	6.00	21.00
Total N	36	36	36



Error SS = 7661.639, df = 35

Dependent Variable: NumTang

Parameter	B	Std. Error	t	Sig.
Intercept	34.194	2.466	13.867	.000

Error SS = 1469.50, df = 30

Dependent Variable: NumTang

Parameter	B	Std. Error	t	Sig.
Intercept	19.167	2.857	6.708	.000
[Store=1.00]	11.500	4.041	2.846	.008
[Store=2.00]	34.833	4.041	8.620	.000
[Store=3.00]	29.833	4.041	7.383	.000
[Store=4.00]	12.167	4.041	3.011	.005
[Store=5.00]	1.833	4.041	.454	.653
[Store=6.00]	0(a)	.	.	.

a This parameter is set to zero because it is redundant.

Error SS = 1307.694, df = 25

Dependent Variable: NumTang

Parameter	B	Std. Error	t	Sig.
Intercept	21.639	3.998	5.413	.000
[Store=1.00]	11.500	4.176	2.754	.011
[Store=2.00]	34.833	4.176	8.342	.000
[Store=3.00]	29.833	4.176	7.145	.000
[Store=4.00]	12.167	4.176	2.914	.007
[Store=5.00]	1.833	4.176	.439	.664
[Store=6.00]	0(a)	.	.	.
[ShelfSpace=6.00]	-4.667	4.176	-1.118	.274
[ShelfSpace=9.00]	-5.333	4.176	-1.277	.213
[ShelfSpace=12.00]	-3.167	4.176	-.758	.455
[ShelfSpace=15.00]	-1.833	4.176	-.439	.664
[ShelfSpace=18.00]	.167	4.176	.040	.968
[ShelfSpace=21.00]	0(a)	.	.	.

a This parameter is set to zero because it is redundant.

Multiple Comparisons

Dependent Variable: NumTang
LSD

(I) ShelfSpace	(J) ShelfSpace	Mean Difference (I-J)	Std. Error	Sig.
6.00	9.00	.6667	4.17563	.874
	12.00	-1.5000	4.17563	.722
	15.00	-2.8333	4.17563	.504
	18.00	-4.8333	4.17563	.258
	21.00	-4.6667	4.17563	.274
9.00	6.00	-.6667	4.17563	.874
	12.00	-2.1667	4.17563	.608
	15.00	-3.5000	4.17563	.410
	18.00	-5.5000	4.17563	.200
	21.00	-5.3333	4.17563	.213
12.00	6.00	1.5000	4.17563	.722
	9.00	2.1667	4.17563	.608
	15.00	-1.3333	4.17563	.752
	18.00	-3.3333	4.17563	.432
	21.00	-3.1667	4.17563	.455
15.00	6.00	2.8333	4.17563	.504
	9.00	3.5000	4.17563	.410
	12.00	1.3333	4.17563	.752
	18.00	-2.0000	4.17563	.636
	21.00	-1.8333	4.17563	.664
18.00	6.00	4.8333	4.17563	.258
	9.00	5.5000	4.17563	.200
	12.00	3.3333	4.17563	.432

	15.00	2.0000	4.17563	.636
	21.00	.1667	4.17563	.968
21.00	6.00	4.6667	4.17563	.274
	9.00	5.3333	4.17563	.213
	12.00	3.1667	4.17563	.455
	15.00	1.8333	4.17563	.664
	18.00	-.1667	4.17563	.968

Based on observed means.

Error SS = 7516.379, df = 34

Dependent Variable: NumTang

Parameter	B	Std. Error	t	Sig.
Intercept	28.902	6.984	4.138	.000
ShelfSpace	.392	.484	.811	.423

Error SS = 1324.24, df = 29

Dependent Variable: NumTang

Parameter	B	Std. Error	t	Sig.
Intercept	13.874	4.052	3.424	.002
[Store=1.00]	11.500	3.901	2.948	.006
[Store=2.00]	34.833	3.901	8.928	.000
[Store=3.00]	29.833	3.901	7.647	.000
[Store=4.00]	12.167	3.901	3.119	.004
[Store=5.00]	1.833	3.901	.470	.642
[Store=6.00]	0(a)	.	.	.
ShelfSpace	.392	.220	1.784	.085

a This parameter is set to zero because it is redundant.

Error SS = 1223.314, df = 24

Dependent Variable: NumTang

Parameter	B	Std. Error	t	Sig.
Intercept	13.124	8.214	1.598	.123
[Store=1.00]	15.743	11.617	1.355	.188
[Store=2.00]	35.476	11.617	3.054	.005
[Store=3.00]	37.933	11.617	3.265	.003
[Store=4.00]	8.181	11.617	.704	.488
[Store=5.00]	-2.667	11.617	-.230	.820
[Store=6.00]	0(a)	.	.	.
ShelfSpace	.448	.569	.787	.439
[Store=1.00] * ShelfSpace	-.314	.805	-.391	.700
[Store=2.00] * ShelfSpace	-.048	.805	-.059	.953
[Store=3.00] * ShelfSpace	-.600	.805	-.746	.463
[Store=4.00] * ShelfSpace	.295	.805	.367	.717
[Store=5.00] * ShelfSpace	.333	.805	.414	.682
[Store=6.00] * ShelfSpace	0(a)	.	.	.

a This parameter is set to zero because it is redundant.

Question 2. Melanoma

Type * Site Crosstabulation

			Site			Total
			Extremities	Head and Neck	Trunk	
Type	Hutchinson's	Count	10	22	2	34
		Expected Count	19.2	5.8	9.0	34.0
	Intermediate	Count	28	11	17	56
		Expected Count	31.6	9.5	14.8	56.0
	Nodular	Count	73	19	33	125
		Expected Count	70.6	21.3	33.1	125.0
	Superficial	Count	115	16	54	185
		Expected Count	104.5	31.5		185.0
Total		Count	226	68	106	400
		Expected Count	226.0	68.0	106.0	400.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	65.813		

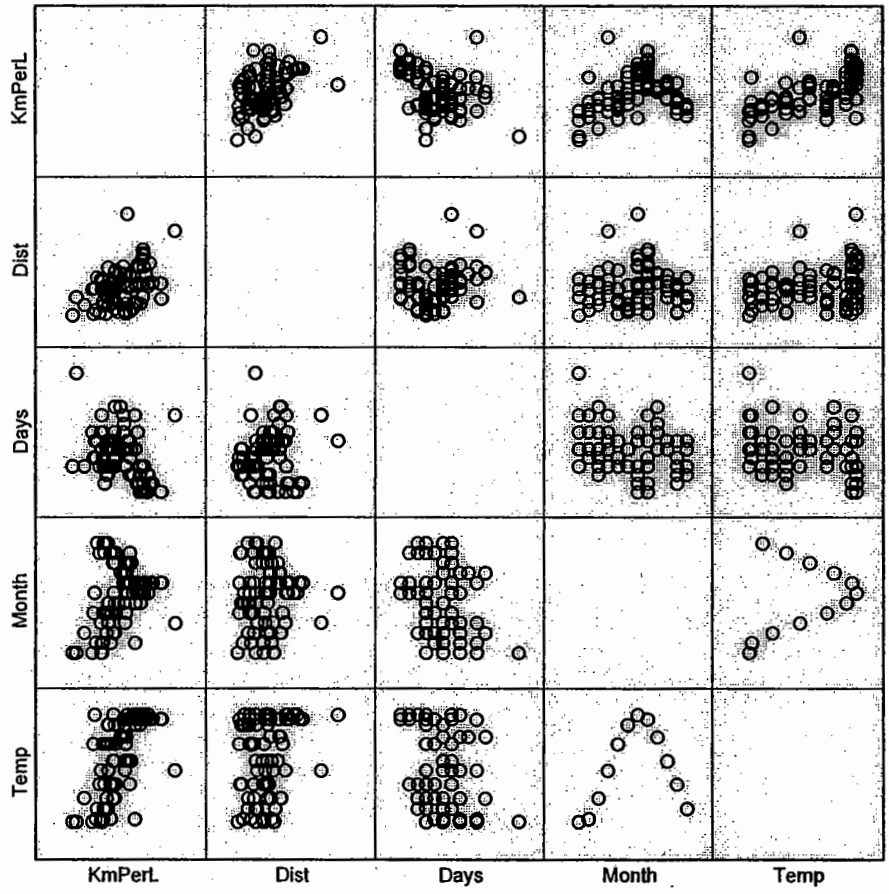
Question 4. Fuel

Case Summaries(a)

	Dist	KmPerL	Days	Month	Temp
1	313.00	6.01	7.00	4.00	5.80
2	374.00	6.55	6.00	4.00	5.80
3	272.00	6.59	5.00	4.00	5.80
4	232.00	7.81	3.00	4.00	5.80
5	268.00	6.57	6.00	5.00	12.50
6	141.00	6.38	3.00	5.00	12.50
7	361.00	7.11	6.00	5.00	12.50
8	141.00	5.44	4.00	5.00	12.50
9	203.00	6.15	4.00	5.00	12.50
10	193.00	7.88	3.00	5.00	12.50
11	214.00	6.79	3.00	5.00	12.50
12	156.00	7.78	5.00	6.00	17.30
13	229.00	6.54	3.00	6.00	17.30
14	187.00	6.58	3.00	6.00	17.30

15	193.00	8.98	3.00	6.00	17.30
16	286.00	8.61	5.00	6.00	17.30
17	167.00	7.05	3.00	6.00	17.30
18	151.00	7.31	5.00	6.00	17.30
19	388.00	8.78	7.00	7.00	19.80
20	195.00	6.96	3.00	7.00	19.80
Total N	20	20	20	20	20

a Limited to first 20 cases.



Error SS = 184.796, df = 85

Dependent Variable: KmPerL

Parameter	B	Std. Error	t	Sig.
Intercept	7.539	.159	47.416	.000

Error SS = 106.080, df = 83

Dependent Variable: KmPerL

Parameter	B	Std. Error	t	Sig.
Intercept	6.409	.415	15.460	.000
Dist	.008	.001	5.883	.000
Days	-.253	.044	-5.757	.000

Error SS = 65.417, df = 72

Dependent Variable: KmPerL

Parameter	B	Std. Error	t	Sig.
Intercept	5.678	.544	10.432	.000
Dist	.004	.001	3.234	.002
Days	-.111	.051	-2.192	.032
[Month=1.00]	-.878	.623	-1.408	.163
[Month=2.00]	.549	.608	.902	.370
[Month=3.00]	.191	.580	.329	.743
[Month=4.00]	1.131	.622	1.820	.073
[Month=5.00]	.423	.559	.756	.452
[Month=6.00]	1.421	.561	2.533	.013
[Month=7.00]	1.696	.529	3.209	.002
[Month=8.00]	2.201	.535	4.118	.000
[Month=9.00]	1.371	.662	2.071	.042
[Month=10.00]	1.370	.581	2.357	.021
[Month=11.00]	.575	.559	1.028	.308
[Month=12.00]	0(a)	.	.	.

a This parameter is set to zero because it is redundant.

Error SS = 79.313, df = 82

Dependent Variable: KmPerL

Parameter	B	Std. Error	t	Sig.
Intercept	5.849	.376	15.556	.000
Dist	.006	.001	4.655	.000
Days	-.135	.044	-3.057	.003
Temp	.069	.013	5.261	.000

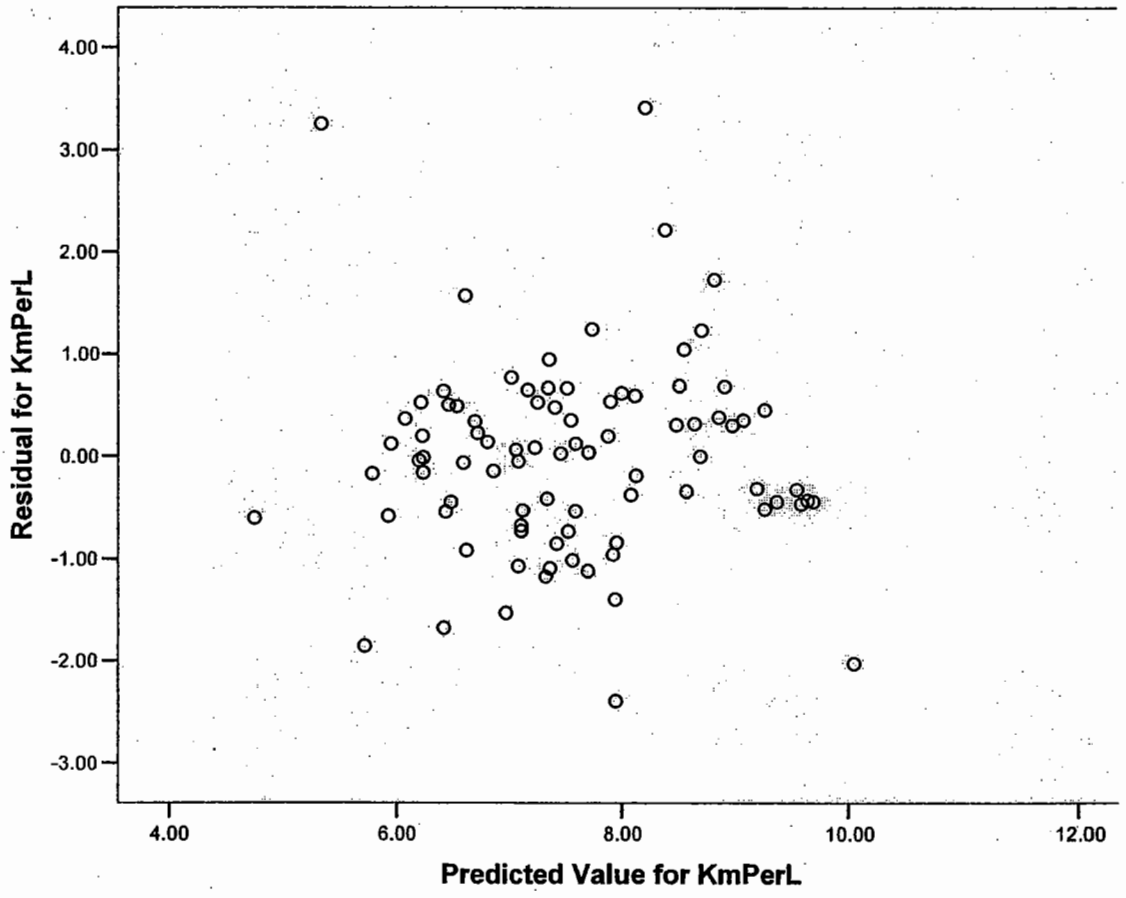
Error SS = 65.417, df = 72

Dependent Variable: KmPerL

Parameter	B	Std. Error	t	Sig.
Intercept	6.036	.427	14.131	.000
Dist	.004	.001	3.234	.002
Days	-.111	.051	-2.192	.032
Temp	.094	.092	1.028	.308
[Month=1.00]	-.567	.819	-.692	.491
[Month=2.00]	.794	.761	1.043	.301
[Month=3.00]	-.054	.489	-.111	.912

[Month=4.00]	.227	.772	.294	.770
[Month=5.00]	-1.113	1.253	-.888	.378
[Month=6.00]	-.566	1.671	-.339	.736
[Month=7.00]	-.527	1.877	-.281	.780
[Month=8.00]	.082	1.781	.046	.963
[Month=9.00]	-.334	1.463	-.228	.820
[Month=10.00]	.240	.916	.262	.794
[Month=11.00]	0(a)	.	.	.
[Month=12.00]	0(a)	.	.	.

a This parameter is set to zero because it is redundant.



Quantiles of the χ^2 distribution

Degrees of	Upper tail probability									
freedom	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
35	17.19	18.51	20.57	22.47	24.80	46.06	49.80	53.20	57.34	60.27
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
45	24.31	25.90	28.37	30.61	33.35	57.51	61.66	65.41	69.96	73.17
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
55	31.73	33.57	36.40	38.96	42.06	68.80	73.31	77.38	82.29	85.75
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17
120	83.85	86.92	91.57	95.70	100.62	140.23	146.57	152.21	158.95	163.65

Quantiles of the F distribution, $P = 0.05$

Denominator degrees of freedom	Numerator degrees of freedom									
	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.7	224.6	230.2	233.0	236.8	238.9	240.5	241.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Critical values of T_L and T_U for the Wilcoxon rank sum test: Independent samples. Test statistic is the rank sum associated with the smaller sample.

$P = 0.025$, one tailed, $P = 0.05$ two tailed.

	n_1															
	3		4		5		6		7		8		9		10	
n_2	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

$P = 0.05$, one tailed, $P = 0.10$ two tailed.

	n_1															
	3		4		5		6		7		8		9		10	
n_2	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	6	15	7	17	7	20	8	22	9	24	9	27	10	29	11	31
4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42
5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	54
6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67
7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80
8	9	27	16	36	24	46	32	58	41	71	52	84	54	90	57	95
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	69	111
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

Critical values of T_0 for the Wilcoxon paired difference signed rank test.

		n															
One-tailed	Two-tailed	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$P = 0.05$	$P = 0.10$	1	2	4	6	8	11	14	17	21	26	30	36	41	47	54	60
$P = 0.025$	$P = 0.05$		1	2	4	6	8	11	14	17	21	25	30	35	40	46	52

End of exam

MATH 204 EXAMINATION 2006

SOLUTIONS

1. For this question I will use the following notation:

`store`	:	for the store factor predictor
`shelf`	:	for the shelf space as a factor predictor
`shelfx`	:	for the shelf space treated as a continuous covariate

Using this notation, the SPSS output corresponds to the following models (in the same order as the output starting at the bottom of page 6). The quantity k is the total number of parameters fitted in the model apart from the intercept, and EDF is the error degrees of freedom. Here $n = 36$, and note that

$$\text{EDF} = n - k - 1 \qquad \therefore \qquad k = n - \text{EDF} - 1$$

	Model	k	EDF	SSE
M0	Null	0	35	7661.639
M1	`store`	5	30	1469.500
M2	`store` + `shelf`	10	25	1307.694
M3	`shelfx`	1	34	7516.379
M4	`store` + `shelfx`	6	29	1324.240
M5	`store` + `shelfx` + `store .shelfx`	11	24	1223.314

For this question, we have no ANOVA tables, so we will use the method of ANOVA-F testing for nested models, and the test statistic

$$F = \frac{(SSE_R - SSE_C)/(k-g)}{SSE_C/(n-k-1)}$$

where SSE_R is the error sum of squares for the **Reduced Model**, specified using $g+1$ parameters including the intercept, and SSE_C is the error sum of squares for the **Complete Model**, specified using $k+1$ parameters including the intercept.

If the reduced model is an adequate simplification of the complete model, then

$$F \sim \text{Fisher-F}(k-g, n-k-1)$$

Note here that

$$k - g = (n - g - 1) - (n - k - 1) = \text{EDF}_R - \text{EDF}_C$$

so the $k - g$ quantities can be deduced directly from the EDFs.

(a) We wish to assess the effect of `shelf` (ie shelf-space as a factor predictor), allowing for the effect of `store` . This suggests that we should compare the nested models M1 and M2, with M2 as the Complete model. Therefore we compare

$$F = \frac{(1469.500 - 1307.694)/(10-5)}{1307.694/25} = 0.619$$

with the Fisher-F$(k-g, n-k-1) \equiv$ Fisher-F$(5, 25)$ distribution. From the tables on page 14, Fisher-F$_{0.05}(5, 25) = 2.60 > 0.619$. Therefore we have no reason to reject M1 as an adequate simplification of M2, and conclude that there is **no evidence** that shelf space (when fitted as a factor predictor) has an effect on sales.

(b) The list of post-hoc test p-values (marked as "Sig." in the final column in the Multiple Comparisons table on page 7) reveals that there are **no** significant differences between any of the levels - all p-values are greater than 0.05.

(c) We wish to assess the effect of `shelfx` (ie shelf-space as a covariate), allowing for the effect of `store` . This suggests that we should compare the nested models M1 and M4, with M4 as the Complete model. Therefore we compare

$$F = \frac{(1469.500 - 1324.240)/(6-5)}{1324.240/29} = 3.181$$

with the Fisher-F$(1, 29)$. From the tables on page 14, Fisher-F$_{0.05}(1, 29) = 4.18 > 3.181$. Therefore we have no reason to reject M1 as an adequate simplification of M4, which implies that there is **no evidence** that shelf-space (when fitted as a continuous covariate) has an effect on sales. For completeness we also compare the nested models M1 and M5 using

$$F = \frac{(1469.500 - 1223.314)/(11-5)}{1223.314/24} = 0.805$$

but as Fisher-F$_{0.05}(6, 24) = 2.51 > 0.805$, again, M1 is an adequate simplification.

Thus it appears that shelf-space is not a useful variable for explaining sales, whether it is fitted as a factor predictor or a covariate.

(d) There is no discrepancy; all parts give the same conclusion. For model M4, the table on page, we see a p-value of 0.085 attached to the coefficient of `shelfx` , which is almost significant at $\alpha = 0.05$, but not quite. The results indicate that a linear relationship between sales and shelf-space is implied, but is not statistically significant. Note that we cannot easily compare M2 with M4, as the models are not nested.

Note: I do not know what the official solution to this question is; I cannot find any evidence in the data when analyzed as above. Clearly the model where a shelf-space is fitted as a covariate produces an almost significant result - fewer parameters are being fitted - but the result is still not significant at $\alpha = 0.05$

(e) To answer the question of whether the effect of shelf-space as a linear covariate changes from store to store, we compare models M4 and M5. Therefore we compare

$$F = \frac{(1324.240 - 1223.314)/(11-6)}{1223.314/24} = 0.396$$

with the Fisher-F$(5, 24)$. From the tables on page 14, Fisher-F$_{0.05}(5, 24) = 2.62 > 0.396$. Therefore we have no reason to reject M4 as an adequate simplification of M5, which implies that there is **no evidence** that shelf-space (when fitted as a continuous covariate) has a different effect on sales in different stores

(f) In my opinion, the most appropriate model is model M1, which fits a **constant but different mean level for each store**. The table on page 6, because of the SPSS contrast parameterization that is adopted, we combine the estimates for the intercept and for store 1 to get the prediction, that is, predicted sales are $19.167 + 11.500 = 30.667$.

The estimated standard deviation, s, yields the standard error of the prediction **within group** as $s/\sqrt{6}$; in the general linear model,

$$s^2 = \frac{SSE}{EDF} = \frac{1469.500}{30} = 48.983$$

so the standard error of the prediction is $\sqrt{48.983/6} = 2.857$; note that this can be deduced by using the arbitrariness of the store labelling - in this balanced case, the standard error of the prediction for any one group mean must be the same as for all other group means.

(g) It it not possible to fit the interaction between `store` and `shelf` as we do not have sufficient data - the number of replicates in each of the factor-level combinations is one. Therefore the interaction model if fitted will yield a sum of squared errors and error degrees of freedom equal to zero.

2. The design of the experiment means that a **two independent samples** analysis is required, as a random sample of advertisements is selected from each publication. Given the tables provided, we will carry out a two sample **Mann-Whitney-Wilcoxon** test; specifically, we will use the small sample version based on the Wilcoxon statistic, W, the sum of the ranks for observation from publication 2. Here the sample sizes are even, so we can label the samples arbitrarily; we take the *Newsweek* data as observations from population 2.

   Thus $n_1 = n_2 = 6$, so $n = 12$. The ranks are computed as below

Group	1	1	1	1	1	1	2	2	2	2	2	2
y	8.20	9.23	9.92	11.16	11.55	15.75	3.12	4.88	5.12	7.67	9.66	10.21
Rank	5	6	8	10	11	12	1	2	3	4	7	9

   Hence $R_1 = 52$, $R_2 = 26$. Thus, according to our convention, $W = 26$.

   Using the tables provided, we find the critical values

$$T_L = 26 \qquad T_U = 52.$$

   and thus $W = T_L$, and W lies within the lower segment of the rejection region. For the hypothesis

$$H_0 \ : \ \eta_1 = \eta_2$$

   we

$$\begin{aligned} \textbf{Reject } H_0 \text{ against } H_a \ &: \ \eta_1 > \eta_2 \quad \text{as} \quad W \leq T_L \\ \textbf{Do not reject } H_0 \text{ against } H_a \ &: \ \eta_1 < \eta_2 \quad \text{as} \quad W < T_U \\ \textbf{Reject } H_0 \text{ against } H_a \ &: \ \eta_1 \neq \eta_2 \quad \text{as} \quad W \leq T_L \end{aligned}$$

   at the $\alpha = 0.025, 0.025$ and 0.05 significance levels respectively.

   Thus we reject the hypothesis of equal population medians, or more generally the hypothesis of equal distributions, in favour of the hypothesis that advertisements in *Scientific American* are more difficult to read than those in *Newsweek*.

3. This question addresses the analysis of categorical data in a contingency table, and the use of a Chi-squared test for independence.

   (a) The expected number of cases under the assumption of independence is

   $$\widehat{n}_{ij} = \frac{n_{i.} n_{.j}}{n}$$

   where $i = 4, j = 3$, and $n_{i.}$ and $n_{.j}$ are the row and column sums for row i and column j respectively. Here, from the SPSS output on page 9, $n_{4.} = 185, n_{.3} = 106$, and $n = 400$, so

   $$\widehat{n}_{43} = \frac{185 \times 106}{400} = 49.025.$$

   Note that the table on the output is rounded to one decimal place; note also that the expected counts totals in rows and columns must equal the actual row and column totals. Hence to this level of approximation, we could find $\widehat{n}_{43}$ using the third row as

   $$\widehat{n}_{43} = 185 - 104.5 - 31.5 = 49.0$$

   (b) The degrees of freedom missing from the table is

   $$(r-1)(c-1) = 3 \times 2 = 6.$$

   Using the table on page 13, the $\alpha = 0.0005$ tail quantile for the Chisquared(6) distribution is 18.55. Therefore we **reject** the null hypothesis of independence, and conclude that there is a difference between the incidence of melanoma at the different sites.

4. (a) Using the first table on page 11, we can conclude by inspecting the t-statistics and p-values that both `Dist` and `Days` are significant in the model. There seems to be no significant multicollinearity between these variables. Thus the simplest model we will consider is

$$\texttt{Dist} + \texttt{Days}$$

To test whether `Month` should be added, we use the ANOVA-F testing procedure for nested models; as

$$k - g = (n - g - 1) - (n - k - 1)$$

we have

$$F = \frac{(106.080 - 65.417)/(83 - 72)}{65.417/72} = 4.069$$

to be compared with the Fisher-F$(k - g, n - k - 1) \equiv$ Fisher-F$(11, 72)$ distribution. From the tables on page 14, Fisher-F$_{0.05}(11, 72) \doteq$ Fisher-F$_{0.05}(10, 80) = 1.95 < 4.069$. Therefore we reject the reduced model as an adequate simplification, and conclude that `Month` should be included, that is, the simplest adequate model checked so far is

$$\texttt{Dist} + \texttt{Days} + \texttt{Month}.$$

We call this model $\mathrm{M_a}$.

(b) When `Temp` replaces `Month` , the same method of comparison gives

$$F = \frac{(106.080 - 78.313)/(83 - 82)}{78.313/82} = 29.074$$

to be compared with the Fisher-F$(k - g, n - k - 1) \equiv$ Fisher-F$(1, 82)$ distribution. From the tables on page 14, Fisher-F$_{0.05}(1, 82) \doteq$ Fisher-F$_{0.05}(1, 80) = 3.96 < 29.074$. Therefore we reject the reduced model as an adequate simplification, and conclude that `Temp` should be included, that is, the model

$$\texttt{Dist} + \texttt{Days} + \texttt{Temp}$$

also fits better than the model `Dist + Days` . This is also confirmed by the p-value for `Temp` in the regression model ($p < 0.001$). We denote this model $\mathrm{M_b}$.

(c) Looking at the original data table on page 9, it is clear that the average monthly temperature provides **identical** information to month number. That is, `Temp` is an exact (linear) function of month. Looking at the sums of squared errors, we see that for both the models

$$\texttt{Dist} + \texttt{Days} + \texttt{Month}$$

and

$$\texttt{Dist} + \texttt{Days} + \texttt{Temp} + \texttt{Month}$$

have identical SSE values (65.417) for each model. Thus `Temp` explains all the effect of `Month` ; we have that the covariates are perfectly multicollinear - this also shows up in the scatterplot on page 10.

(d) It is hard to assess this graph, but it seems that there might be (i) some outliers (extremely large in magnitude residuals) and (ii) a slight positive incline in the plot of predicted values versus residuals. Thus the assumption of Normality might be inappropriate, and the assumption of constant variance might also be inappropriate. It also seems that the residuals are patterned, in that they seem more likely to be negative when `KmPerl` is small, and positive when `KmPerl` is large, which indicates a deficiency in the model, although not necessarily in the assumptions about the random error.

(e) The proportion of variation explained by model $\mathrm{M_b}$ is

$$R^2 = \frac{SSE_{\text{Null}} - SSE_{\mathrm{M_b}}}{SSE_{\text{Null}}} = \frac{184.798 - 65.417}{184.798} = 0.646$$

(f) The best fitting model is either $\mathrm{M_a}$ or $\mathrm{M_b}$. We assess this by inspecting the adjusted R^2 quantities

$$\text{Adj } R^2_{\mathrm{M}} = 1 - \frac{SSE_{\mathrm{M}}/(n-k-1)}{SSE_{\text{Null}}/(n-1)}$$

For Model $\mathrm{M_a}$:

$$\text{Adj } R^2_{\mathrm{M_a}} = 1 - \frac{SSE_{\mathrm{M_a}}/(n-k-1)}{SSE_{\text{Null}}/(n-1)} = 1 - \frac{65.417/72}{184.796/85} = 0.582$$

For Model $\mathrm{M_b}$:

$$\text{Adj } R^2_{\mathrm{M_b}} = 1 - \frac{SSE_{\mathrm{M_b}}/(n-k-1)}{SSE_{\text{Null}}/(n-1)} = 1 - \frac{79.313/82}{184.796/85} = 0.555$$

Thus the prediction is found by using the formula with estimates from model $\mathrm{M_a}$

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \texttt{Dist} + \widehat{\beta}_2 \texttt{Days} + \widehat{\beta}_3 \texttt{Month}$$

where `Month` is month 4, and $\widehat{\beta}_3$ is the coefficient estimated for that level of the factor predictor. From the table on page 11, we have

$$5.676 + (0.004 \times 300) + (-0.111 \times 0) + 1.131 = 8.007$$

In reality, it is **not possible** to report a standard error for this prediction. For **independent** random variables $X_1, \ldots, X_k$ and constants $a_1, \ldots, a_k$, the formula

$$\text{Var}\left[\sum_{i=1}^{k} a_k X_k\right] = \sum_{i=1}^{k} a_k^2 \text{Var}[X_k]$$

gives the variance of the sum in terms of the sum of the variables. However, this result does not work for **dependent** variables, so the calculation

$$\text{s.e.}(\widehat{y}) = \sqrt{0.544^2 + 300^2 \times 0.001^2 + 0.622^2} = 0.879$$

that takes the sums of the squared standard errors appropriately scaled **is not valid** here, as the coefficients $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$ are probabilistically dependent.

(g) There are two redundant parameters in the estimation of `Month` factor level coefficients.

- The first is redundant as it is essentially reported as the "Intercept" in the first row of the tables. SPSS sets the baseline group as the highest labelled factor level, and uses a contrast parameterization, that is, estimates differences from baseline. The intercept is reporting the mean level of the baseline group.
- The second parameter is redundant as `Month` is perfectly multicollinear with `Temp` . There are only twelve different parameters to represent the effect of the twelve months; the model

  $$\texttt{Dist} + \texttt{Days} + \texttt{Temp} + \texttt{Month}$$

  tries to fit fourteen (one intercept, one for `Temp` and twelve for month). Thus the software is forced to set two parameters to zero.

McGill University

Faculty of Science

MATH 204

PRINCIPLES OF STATISTICS II

Final Examination

Date: 23rd April 2007 Time: 2pm-5pm

Examiner: Dr. David A. Stephens
Associate Examiner: Dr. Russell Steele

Please write your answers in the answer booklets provided.

This paper contains six questions. Each question carries 25 marks. Credit will be given for all questions attempted. The total mark available is 150 but rescaling of the final mark may occur

Candidates may take one double-sided sheet of Letter-sized (216 $\times$ 279 mm) or A4-sized (210 $\times$ 297mm) paper with handwritten notes into the examination room.

Calculators may be used. Relevant statistical tables are provided.

Dictionaries and Translation dictionaries are permitted.

© 2007 McGill University

1. In an experimental study of nutrition, laboratory animals were allocated at random to one of four different diets, A,B,C and D. The response measurement was the weight gain (in grammes) of each animal over one week.

   The data recorded are tabulated below; entries in the rows of the table are the weight gains for animals allocated to each diet.

Diet							
A	0.54	1.98	0.65	0.52	1.92	1.48	0.97
B	1.24	1.82	1.39	1.25	1.29		
C	2.05	2.18	1.94	2.50	1.98	2.17	1.83
D	1.88	6.23	3.51	3.77	1.25	0.72	

   (a) What is the fundamental difference between an *experimental* and an *observational* scientific study ?

   4 MARKS

   (b) In the experimental study described above, what kind of design is being used ?

   2 MARKS

   (c) Using the data, an ANOVA analysis is to be carried out. The ANOVA table below contains some missing entries marked by the notation $\star$.

SOURCE	DF	SS	MS	F
TREATMENTS	$\star$	11.355	$\star$	$\star$
ERROR	21	$\star$	$\star$	
TOTAL	$\star$	35.053		

   Write out the ANOVA table in full, filling in the missing values **using the information already given in the table**.

   10 MARKS

   (d) What is the conclusion of the ANOVA analysis ? State clearly the null and alternative hypothesis, the test statistic, the null distribution, and the conclusion.
   A table of the Fisher-F distribution is provided on page 13. Entries in the table are the 0.05 tail quantile of the Fisher-F(ν_1, ν_2) distribution, for different values of ν_1 and ν_2.

   5 MARKS

   (e) What other information is required to confirm that the conclusion of the test is valid ? Justify your answer in terms of the assumptions underlying the ANOVA analysis.

   4 MARKS

2. (a) Explain the structure of a *balanced complete randomized block design with replication* for one treatment factor and one blocking factor. Explain why this kind of design differs from a *factorial design* with two treatment factors.

6 MARKS

(b) In a study of the heart drug enalaprilat, nine patients with congestive heart disease were treated with four different dose levels of the drug (labelled 0 to 3), and their active heart rate (in beats per minute) recorded. The objective of the study was to find out whether there are any differences between the heart rates at different dose levels. In the analysis, the patients are to be treated as levels of a (fixed) blocking factor.

Reference: CS Maskin, S Ocken, B Chadwick, and TH LeJemtel (1985), Comparative systemic and renal effects of dopamine and angiotensin - converting enzyme inhibition with enalaprilat in patients with heart failure *Circulation*, **72**, pp 846-852 (adapted)

The data from the study are displayed in the table below.

	Dose Level			
Patient	0	1	2	3
1	96	92	86	92
2	110	106	108	114
3	89	86	85	83
4	95	78	78	83
5	128	124	118	118
6	100	98	199	94
7	72	68	67	71
8	79	75	74	74
9	100	106	104	102

The data were analyzed in SPSS; the output is included on page 8. The variable **id** denotes the different patients, whereas **dose** denotes dose level.

(i) Explain the results labelled Analysis 1, the model that was fitted, the hypotheses tested, and the conclusions that are indicated.

9 MARKS

(ii) Explain the results labelled Analysis 2; in particular, explain what has gone wrong and why.

6 MARKS

(c) If the assumptions underlying the ANOVA tests were not believed to hold, briefly describe one alternative approach that could be used to test for differences between responses at different dose levels, taking into account the blocking structure.

4 MARKS

3. Measurements of bacteria counts following the culturing of the bacterium *Staphylococcus aureus* at three temperatures (27, 35 or 43 Celsius, labelled T1 to T3 respectively), for five different concentrations of the nutrient tryptone (0.6, 0.8, 1.0, 1.2 or 1.4, labelled C1 to C5 respectively) were recorded. Two independent replicate experiments were made at each of the 3×5 factor level combinations, with the objective of finding out which combination promoted the highest level of bacteria growth.

(a) What is the name of this kind of design ?

3 MARKS

(b) List the five (main effects and interactions) models that can be fitted to these data if the two variables are treated as factors. Use the standard model notation, where, for two factors A and B, the main effects plus interaction model is denoted

$$A + B + A.B$$

For this example, use the notation

C for the concentration factor

T for the temperature factor

5 MARKS

(c) Output from the SPSS analysis of the response data (on the log scale) is given on page 9. In the output, the variable **conc** denotes concentration, and **temp** denotes temperature.

Summarize in detail what the output illustrates. Comment specifically on

(i) The statistical evidence for including or excluding an interaction term.

(ii) The most appropriate model in ANOVA-F terms, using **backward selection**. Recall that the ANOVA-F test for comparing nested models is based on the statistic

$$F = \frac{(SSE_R - SSE_C)/(k-g)}{SSE_C/(n-k-1)}$$

where

- SSE_R is the error sum of squares for the **Reduced Model**, specified using $g+1$ parameters including the intercept.
- SSE_C is the error sum of squares for the **Complete Model**, specified using $k+1$ parameters including the intercept.

If the reduced model is an adequate simplification of the complete model, then

$$F \sim \text{Fisher-F}(k-g, n-k-1)$$

(iii) Overall model fit.

12 MARKS

(d) In the analysis described in (c), concentration is treated as a **factor predictor** taking five levels. However, it is possible instead to treat concentration as a **continuous covariate**.

How many **fewer** parameters does the **main effects only** model

$$C + T$$

contain if concentration is treated as a continuous covariate rather than a factor predictor ? Justify your answer.

5 MARKS

4. For lung transplantation it is desirable for the donor's lungs to have similar total capacity to those of the recipient. Total lung capacity (TLC) is difficult to measure, so a predictive model for TLC in terms of Age, Sex and Height of the donor is sought.

Measurements of TLC were collected from 32 patients pre-transplant, and their age, sex and height recorded, and analyzed using a general linear regression model.

Reference: Otulana BA, Higenbottam TW, Scott JP, Clelland C, Hutter, J and Wallwork J. (1989) Pulmonary function monitoring allows diagnosis of rejection in heart-lung transplant recipients. *Transplant Proceedings*, 1989;21:25834.

Output from an SPSS analysis is recorded on pages 11 to 12. Analyses of different models are presented under different sub-headings. In the analysis, sex is a factor predictor, and is coded as 1 for female, 2 for male.

(a) Given the results of all the analyses presented, decide on the best model to explain the observed variation in the response, total lung capacity. Justify your answer.

12 MARKS

(b) Explain the apparently contradictory results of Analyses 2 and 4.

4 MARKS

(c) Using analyses available, predict the total lung capacity for a 30-year-old female who is 165 cm tall. Recall that to predict in a linear regression model of the form

$$y = \beta_0 + \beta_1 x_1 + \cdots \beta_k x_k + \epsilon$$

the prediction is

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots + \widehat{\beta}_k x_k$$

where $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$ are the estimated coefficients.

6 MARKS

(d) In Analysis 2 and Analysis 4, some of the parameter estimates are given the note

This parameter is set to zero because it is redundant.

Briefly explain this note.

3 MARKS

5. The following data relate to a study of dental enamel erosion by swimming pool chlorine, in a case-control observational study of swimmers; 49 cases of dental enamel erosion, and 245 controls, were recruited, and asked if they swam for greater than or equal to six hours, or less than six hours per week.

Reference: Centerwall BS *et al.* (1986), Erosion of dental enamel among competitive swimmers at a gas-chlorinated swimming pool, *American Journal of Epidemiology*, **123**(4) pp 641-647.

Swim	Erosion of Enamel		
≥ 6 hrs.	Yes	No	Total
Yes	32	118	150
No	17	127	144
Total	49	245	294

For this $r \times c$ table $(r = 2, c = 2)$, we wish to test the null hypothesis of **independence** using a Chi-squared test, and the odds-ratio statistic.

(a) Form the table of **expected values** under the null hypothesis with entries $\widehat{n}_{ij}$ given by the formula

$$\widehat{n}_{ij} = \frac{n_{i.} n_{.j}}{n} \qquad i = 1, 2,\ j = 1, 2.$$

where

$n_{i.}$ is the row total for row i $\qquad$ $n_{.j}$ is the column total for column j.

6 MARKS

(b) Compute the Chi-squared statistic

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(n_{ij} - \widehat{n}_{ij})^2}{\widehat{n}_{ij}}$$

4 MARKS

(c) Complete the test at the $\alpha = 0.05$ significance level of the null hypothesis, recalling that if the independence hypothesis is true, $X^2 \approx$ Chisquared$((r-1)(c-1))$.

The table on page 14 contains the 0.05 and 0.01 tail quantiles of the Chisquared(ν) distribution, for $\nu = 1$ to 20.

4 MARKS

(d) Are the required conditions for the Chi-squared test met for this analysis ? Justify your answer.

3 MARKS

(e) Using the log odds ratio, its standard error and the test statistic Z

$$\log \widehat{\psi} = \log \left(\frac{n_{11}\, n_{22}}{n_{12}\, n_{21}} \right) \qquad \text{s.e.}(\log \widehat{\psi}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

$$Z = \frac{\log \widehat{\psi}}{\text{s.e.}(\log \widehat{\psi})}$$

test for an association between the factor and disease status. Use the result that under the hypothesis of no association, $Z \approx N(0, 1)$, so that the critical values in a two-tailed test are ± 1.96.

8 MARKS

6. (a) Explain briefly what each of the following named tests is used for:

(i) Levene's Test

(ii) Friedman's Test

(iii) Fisher's Exact Test

9 MARKS

(b) Using the Kruskal-Wallis Test, assess the statistical significance of the following data. The data correspond to the comparison of three treatments (1-Control, 2-Relaxation, 3-Dietary) for chronic headache sufferers; data in the table are the improvement measures (y) and corresponding ranks (r) for the 18 patients in this randomized experimental study.

Reference: Fentress, DW, Masek, BJ, Mehegan, JE, and Benson, H, (1986). Biofeedback and relaxation-response training in the treatment of pediatric migraine. *Developmental Medicine and Child Neurology*, **28**, 138146.

Group	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3
y	62	74	86	74	91	37	69	43	100	94	100	98	50	-120	100	-288	4	-76
r	11	8.5	7	8.5	6	14	10	13	2	5	2	4	12	17	2	18	15	16

Recall that the Kruskal-Wallis test statistic is

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

where R_j is the rank sum for group $j = 1, 2, 3$, and k is the number of groups being compared. If the relevant null hypothesis, H_0, is **true**, then for large n,

$$H \approx \text{Chisquared}(k-1).$$

The table on page 14 contains the 0.05 and 0.01 tail quantiles of the Chisquared(ν) distribution, for $\nu = 1$ to 20.

12 MARKS

(c) If the data y were assumed to be Normally distributed, what kind of test might be used to analyze the data ? List any extra assumptions that need to be made.

4 MARKS

Output for Question 2

Analysis 1:

Tests of Between-Subjects Effects

Dependent Variable: y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12673.278[a]	11	1152.116	3.625	.004
Intercept	331008.444	1	331008.444	1041.415	.000
id	12107.056	8	1513.382	4.761	.001
dose	566.222	3	188.741	.594	.625
Error	7628.278	24	317.845		
Total	351310.000	36			
Corrected Total	20301.556	35			

a. R Squared = .624 (Adjusted R Squared = .452)

Parameter Estimates

Dependent Variable: y

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	99.444	10.293	9.661	.000	78.200	120.688
[id=1]	-11.500	12.606	-.912	.371	-37.518	14.518
[id=2]	6.500	12.606	.516	.611	-19.518	32.518
[id=3]	-17.250	12.606	-1.368	.184	-43.268	8.768
[id=4]	-19.500	12.606	-1.547	.135	-45.518	6.518
[id=5]	19.000	12.606	1.507	.145	-7.018	45.018
[id=6]	19.750	12.606	1.567	.130	-6.268	45.768
[id=7]	-33.500	12.606	-2.657	.014	-59.518	-7.482
[id=8]	-27.500	12.606	-2.181	.039	-53.518	-1.482
[id=9]	0[a]	.	.	.	.	.
[dose=1]	4.222	8.404	.502	.620	-13.123	21.568
[dose=2]	.222	8.404	.026	.979	-17.123	17.568
[dose=3]	9.778	8.404	1.163	.256	-7.568	27.123
[dose=4]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Analysis 2:

Tests of Between-Subjects Effects

Dependent Variable: y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	20301.556[a]	35	580.044	.	.
Intercept	331008.444	1	331008.444	.	.
id	12107.056	8	1513.382	.	.
dose	566.222	3	188.741	.	.
id * dose	7628.278	24	317.845	.	.
Error	.000	0	.		
Total	351310.000	36			
Corrected Total	20301.556	35			

a. R Squared = 1.000 (Adjusted R Squared = .)

Output for Question 3

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9.356	14	.668	.917	.562
Intercept	636.336	1	636.336	873.407	.000
conc	2.658	4	.665	.912	.482
temp	6.378	2	3.189	4.377	.032
conc * temp	.320	8	.040	.055	1.000
Error	10.929	15	.729		
Total	656.621	30			
Corrected Total	20.284	29			

R Squared = .461 (Adjusted R Squared = -.042)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9.036	6	1.506	3.080	.023
Intercept	636.336	1	636.336	1301.179	.000
conc	2.658	4	.665	1.359	.279
temp	6.378	2	3.189	6.521	.006
Error	11.248	23	.489		
Total	656.621	30			
Corrected Total	20.284	29			

R Squared = .445 (Adjusted R Squared = .301)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.378	2	3.189	6.191	.006
Intercept	636.336	1	636.336	1235.470	.000
temp	6.378	2	3.189	6.191	.006
Error	13.907	27	.515		
Total	656.621	30			
Corrected Total	20.284	29			

R Squared = .314 (Adjusted R Squared = .264)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.658	4	.665	.943	.456
Intercept	636.336	1	636.336	902.557	.000
conc	2.658	4	.665	.943	.456
Error	17.626	25	.705		
Total	656.621	30			
Corrected Total	20.284	29			

R Squared = .131 (Adjusted R Squared = -.008)

Output for Question 4

Analysis 1:

Tests of Between-Subjects Effects

Dependent Variable: Total Lung Capacity

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.411[a]	1	4.411	1.712	.201
Intercept	96.398	1	96.398	37.412	.000
Age	4.411	1	4.411	1.712	.201
Error	77.301	30	2.577		
Total	1267.557	32			
Corrected Total	81.712	31			

a. R Squared = .054 (Adjusted R Squared = .022)

Analysis 2:

Tests of Between-Subjects Effects

Dependent Variable: Total Lung Capacity

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	25.312[a]	1	25.312	13.464	.001
Intercept	1185.845	1	1185.845	630.764	.000
sex	25.312	1	25.312	13.464	.001
Error	56.400	30	1.880		
Total	1267.557	32			
Corrected Total	81.712	31			

a. R Squared = .310 (Adjusted R Squared = .287)

Parameter Estimates

Dependent Variable: Total Lung Capacity

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	6.977	.343	20.354	.000	6.277	7.677
[sex=1]	-1.779	.485	-3.669	.001	-2.769	-.789
[sex=2]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Output for Question 4

Analysis 3:

Tests of Between-Subjects Effects

Dependent Variable: Total Lung Capacity

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39.549[a]	1	39.549	28.140	.000
Intercept	14.904	1	14.904	10.604	.003
Height	39.549	1	39.549	28.140	.000
Error	42.163	30	1.405		
Total	1267.557	32			
Corrected Total	81.712	31			

a. R Squared = .484 (Adjusted R Squared = .467)

Parameter Estimates

Dependent Variable: Total Lung Capacity

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	-9.740	2.991	-3.256	.003	-15.849	-3.632
Height	.095	.018	5.305	.000	.058	.131

Analysis 4:

Tests of Between-Subjects Effects

Dependent Variable: Total Lung Capacity

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	42.818[a]	3	14.273	10.275	.000
Intercept	4.771	1	4.771	3.435	.074
sex	.080	1	.080	.057	.813
Height	17.361	1	17.361	12.498	.001
sex * Height	.025	1	.025	.018	.895
Error	38.894	28	1.389		
Total	1267.557	32			
Corrected Total	81.712	31			

a. R Squared = .524 (Adjusted R Squared = .473)

Parameter Estimates

Dependent Variable: Total Lung Capacity

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	-5.828	4.977	-1.171	.251	-16.023	4.367
[sex=1]	-1.728	7.221	-.239	.813	-16.520	13.064
[sex=2]	0[a]	.	.	.	.	.
Height	.074	.029	2.577	.016	.015	.132
[sex=1] * Height	.006	.043	.133	.895	-.083	.094
[sex=2] * Height	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Output for Question 4

Analysis 5:

Tests of Between-Subjects Effects

Dependent Variable: Total Lung Capacity

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	47.718[a]	6	7.953	5.849	.001
Intercept	7.568	1	7.568	5.566	.026
sex	.348	1	.348	.256	.617
Age	3.105	1	3.105	2.284	.143
Height	12.936	1	12.936	9.514	.005
sex * Age	1.150	1	1.150	.846	.367
sex * Height	.359	1	.359	.264	.612
Age * Height	3.389	1	3.389	2.492	.127
Error	33.994	25	1.360		
Total	1267.557	32			
Corrected Total	81.712	31			

a. R Squared = .584 (Adjusted R Squared = .484)

Analysis 6:

Tests of Between-Subjects Effects

Dependent Variable: Total Lung Capacity

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	50.492[a]	7	7.213	5.545	.001
Intercept	8.781	1	8.781	6.750	.016
sex	1.476	1	1.476	1.135	.297
Age	4.933	1	4.933	3.792	.063
Height	13.967	1	13.967	10.737	.003
sex * Age	3.033	1	3.033	2.331	.140
sex * Height	1.321	1	1.321	1.016	.324
Age * Height	5.050	1	5.050	3.882	.060
sex * Age * Height	2.774	1	2.774	2.132	.157
Error	31.220	24	1.301		
Total	1267.557	32			
Corrected Total	81.712	31			

a. R Squared = .618 (Adjusted R Squared = .506)

Table of the Fisher-F distribution

Entries in table are the $\alpha = 0.05$ tail quantile of Fisher-F(ν_1, ν_2) distribution
ν_1 given in columns, ν_2 given in rows.

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91	244.69	245.36	245.95	246.46	246.92	247.32	247.69	248.01	248.31	248.58	248.83	249.05	249.26
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43	19.43	19.44	19.44	19.44	19.45	19.45	19.45	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67	8.67	8.66	8.65	8.65	8.64	8.64	8.63
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82	5.81	5.80	5.79	5.79	5.78	5.77	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58	4.57	4.56	4.55	4.54	4.53	4.53	4.52
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90	3.88	3.87	3.86	3.86	3.85	3.84	3.83
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47	3.46	3.44	3.43	3.43	3.42	3.41	3.40
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17	3.16	3.15	3.14	3.13	3.12	3.12	3.11
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94	2.93	2.92	2.91	2.90	2.89
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79	2.77	2.76	2.75	2.75	2.74	2.73
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66	2.65	2.64	2.63	2.62	2.61	2.60
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56	2.54	2.53	2.52	2.51	2.51	2.50
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47	2.46	2.45	2.44	2.43	2.42	2.41
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40	2.39	2.38	2.37	2.36	2.35	2.34
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34	2.33	2.32	2.31	2.30	2.29	2.28
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29	2.28	2.26	2.25	2.24	2.24	2.23
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24	2.23	2.22	2.21	2.20	2.19	2.18
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20	2.19	2.18	2.17	2.16	2.15	2.14
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17	2.16	2.14	2.13	2.12	2.11	2.11
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15	2.14	2.12	2.11	2.10	2.09	2.08	2.07
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.11	2.10	2.08	2.07	2.06	2.05	2.05
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10	2.08	2.07	2.06	2.05	2.04	2.03	2.02
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15	2.13	2.11	2.09	2.08	2.06	2.05	2.04	2.02	2.01	2.01	2.00
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11	2.09	2.07	2.05	2.04	2.03	2.01	2.00	1.99	1.98	1.97
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04	2.02	2.01	2.00	1.98	1.97	1.96	1.96
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02	2.00	1.99	1.98	1.97	1.96	1.95	1.94
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.10	2.08	2.06	2.04	2.02	2.00	1.99	1.97	1.96	1.95	1.94	1.93	1.92
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99	1.97	1.96	1.95	1.93	1.92	1.91	1.91
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08	2.05	2.03	2.01	1.99	1.97	1.96	1.94	1.93	1.92	1.91	1.90	1.89
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96	1.95	1.93	1.92	1.91	1.90	1.89	1.88
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	2.11	2.08	2.05	2.03	2.00	1.98	1.96	1.95	1.93	1.92	1.91	1.90	1.88	1.88	1.87
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07	2.04	2.01	1.99	1.97	1.95	1.94	1.92	1.91	1.90	1.88	1.87	1.86	1.85

Table of the **Chisquared**(ν) distribution

Entries in table are the $\alpha = 0.05$ and $\alpha = 0.01$ tail quantiles

ν	1	2	3	4	5	6	7	8	9	10
$\alpha = 0.05$	3.841	5.991	7.815	9.488	11.070	12.592	14.067	15.507	16.919	18.307
$\alpha = 0.01$	6.635	9.210	11.345	13.277	15.086	16.812	18.475	20.090	21.666	23.209

ν	11	12	13	14	15	16	17	18	19	20
$\alpha = 0.05$	19.675	21.026	22.362	23.685	24.996	26.296	27.587	28.869	30.144	31.410
$\alpha = 0.01$	24.725	26.217	27.688	29.141	30.578	32.000	33.409	34.805	36.191	37.566

McGill University
Faculty of Science

Department of Mathematics and Statistics

April 2007

MATH 204

PRINCIPLES OF STATISTICS II

SOLUTIONS

1. (a) In an experimental study, the treatment is **assigned** to the experimental units by the experimenter; in an observational study, the experimenter merely **records or observes** the treatments received.

4 MARKS

(b) This is a **completely randomized design**.

2 MARKS

(c) The ANOVA table is as follows:

SOURCE	DF	SS	MS	F
TREATMENTS	3	11.355	3.785	3.355
ERROR	21	23.698	1.128	
TOTAL	24	35.053		

10 MARKS

(d) Null hypothesis is

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \cdots = \mu_k \\ H_a &: \text{At least one pair of } \mu\text{s different.} \end{aligned}$$

The test statistic is $F = 3.355$ and if H_0 is true,

$$F \sim \text{Fisher-F}(k-1, n-k) \equiv \text{Fisher-F}(3, 21)$$

Using the tables, for $\alpha = 0.05$, we have

$$\text{F}_\alpha(3, 21) = 3.07 < 3.355$$

so we **reject** the null hypothesis at $\alpha = 0.05$.

5 MARKS

(e) We need to assume **independence** and **normality** of the random errors, and to have **equal population variances** in the four groups.

4 MARKS

2. (a) In a balanced complete randomized block design with replication, we have r replicates for each of the $b \times k$ blocking factor level/treatment factor level combinations. For each level of the blocking factor, we randomly select kr experimental units, and then allocated r at random to each of the treatment levels. In a factorial design, no distinction is made between the two factors in terms of population substructure; we do not block according to the blocking factor.

6 MARKS

(b) (i) The model fitted is

$$\textbf{id} + \textbf{dose}$$

that is, a main effects for blocking factor patient and dose.

1 MARK

In total, two hypothesis are tested;

- the first concerns the differences between levels of **id** (differences between patients); this test produces a p-value of 0.001, which is significant at $\alpha = 0.05$ level. Hence we **reject** the hypothesis that the responses at different levels of **id** are equal, confirming that the blocking by **id** is necessary,

4 MARKS

- the second concerns the differences between levels of **dose**; this test produces a p-value of 0.625, which is not significant at $\alpha = 0.05$ level. Hence we **do not reject** the hypothesis that the responses at different levels of **dose** are equal.

4 MARKS

(ii) In this analysis, the model with interaction

$$\textbf{id} + \textbf{dose} + \textbf{id} \,.\, \textbf{dose}$$

However, with no replication, we cannot test this hypothesis, as the number of parameters equals the number of data points, leading to the result $SSE = EDF = 0$. Thus the F and p-values cannot be computed.

6 MARKS

(c) You could use the **non-parametric** procedure, Friedman's test, which takes into account the blocking structure. Alternatively, you could use a randomization/permutation procedure, taking care to preserve block structure when randomizing.

4 MARKS

3. (a) This is a balanced complete factorial design.

3 MARKS

(b) The five models are (in the order analyzed)

$$\begin{array}{ll} M_1 & C+T+C.T \\ M_2 & C+T \\ M_3 & T \\ M_4 & C \\ M_0 & \text{Null} \end{array}$$

5 MARKS

(c) The analysis fits the models in the above sequence; using inspection of the p-values in this **balanced design**, it appears that the backward selection sequence should be

$$M_1 \longrightarrow M_2 \longrightarrow M_3 \longrightarrow M_0$$

The sums of squares in each comparison is listed below

Complete	SSE_C	k	Reduced	SSE_R	g	F	$k-g$	$n-k-1$	F_α	Conclusion
M_1	10.929	14	M_2	11.248	6	0.055	8	15	2.64	Do not reject
M_2	11.248	6	M_3	13.907	2	1.359	4	23	2.80	Do not reject
M_3	13.907	2	M_0	20.284	0	6.190	2	27	3.35	Reject

Hence

(i) The interaction should be omitted

(ii) The most appropriate model is model M_3 that contains the main effect T only, that is, there is a different response for the different temperature levels, but no effect of concentration.

(iii) For this model, the R^2 and adjusted R^2 quantities are 0.314 and 0.264 respectively. Hence the explanatory power overall is not very high. We cannot comment on the appropriateness of the normality assumptions, or the presence of outliers, which also need to be checked.

Note that most of these conclusions can be deduced from the original ANOVA tables, as the design is balanced and complete.

12 MARKS

(d) The model

$$C+T$$

contains **six** non-intercept parameters in total if C is fitted as a factor predictor. Of these, four parameters correspond to the $5-1=4$ contrasts from baseline due to concentration. If concentration is fitted as a covariate, then there is only a single parameter for concentration, and the number of non-intercept parameters is $2+1=3$. Hence the difference is **three**.

5 MARKS

4. (a) By inspection, it seems that **Height** is the only variable that should be in the model; from Analysis 1, it seems that **Age** is not significant in the model, and despite the results of Analysis 2, **Sex** is also not significant in the presence of **Height** .

We compare the models

$$\textbf{Height} \qquad \text{and} \qquad \textbf{Height} + \textbf{Sex} + \textbf{Height} \,.\, \textbf{Sex}$$

using the F-test. The test statistic is

$$F = \frac{(42.163 - 38.894)/(3-1)}{38.894/28} = 1.389$$

which we compare against the Fisher-F$(2, 28)$ distribution. From tables, $\mathrm{F}_{\alpha}(2, 28) = 3.34$, so we do not reject this hypothesis that the reduced model is an adequate simplification of the complete model.

We cannot assess whether **Height** would not be significant in the presence of **Age** as this output is not recorded, but this is unlikely as **Age** and **Height** are unlikely to be dependent in adults. Also Analyses 1, 5 and 6 indicate that **Age** is not significant.

The R^2 and Adjusted R^2 for the preferred model are 0.484 and 0.467, so the explanatory power is only moderate.

12 MARKS

(b) There is dependence between **Sex** and **Height** and hence their apparent effects are **confounded**. Simply, we might suspect that height predicts sex well, that is, taller individuals are likely to be men.

4 MARKS

(c) Using the results from Analysis 3, the prediction is

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_{\textbf{Height}} \times 165 = -9.740 + 0.095 \times 165 = 5.935$$

6 MARKS

(d) This note is given as the corresponding parameter is estimated as the *Intercept*, as is always the case for factor predictors.

3 MARKS

5. (a) We have for the fitted values

Swim	Erosion of Enamel		
≥ 6 hrs.	Yes	No	Total
Yes	25	125	150
No	24	120	144
Total	49	245	294

6 MARKS

(b)
$$X^2 = \frac{(32-25)^2}{25} + \frac{(17-24)^2}{24} + \frac{(118-125)^2}{125} + \frac{(127-120)^2}{120} = 4.802$$

4 MARKS

(c) The $\alpha = 0.05$ quantile for Chisquared(1) distribution is 3.841. Thus we **reject** the hypothesis of independence.

4 MARKS

(d) The assumptions behind the chi-squared test may in general be violated for a case control study, as we do not have independent multinomial sampling overall; actually in this case the assumptions are met. Also, all expected cell entries are greater than five, so the chi-squared approximation seems to be valid.

3 MARKS

(e) Here we have
$$\log \widehat{\psi} = \log\left(\frac{n_{11}\, n_{22}}{n_{12}\, n_{21}}\right) = \log\left(\frac{32 \times 127}{118 \times 17}\right) = 0.706$$

and
$$\text{s.e.}(\log \widehat{\psi}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} = \sqrt{\frac{1}{32} + \frac{1}{118} + \frac{1}{17} + \frac{1}{127}} = 0.326$$

so that
$$Z = \frac{\log \widehat{\psi}}{\text{s.e.}(\log \widehat{\psi})} = \frac{0.706}{0.326} = 2.166$$

Given the critical values ± 1.96, we conclude that the log odds ratio is significantly different from zero.

8 MARKS

6. (a) (i) Levene's Test: Testing the equality of variances, for example in a one-way or two-way layout, as a precursor for an ANOVA test.

(ii) Friedman's Test: Non-parametric equivalent to ANOVA for the randomized block design.

(iii) Fisher's Exact Test: Exact test for independence in a 2×2 table.

9 MARKS

(b) We have

$$R_1 = 55 \qquad R_2 = 36 \qquad R_3 = 80$$

and hence

$$H = 5.696$$

We compare this with the Chisquared$(k-1)$ = Chisquared(2) distribution; Chisq$_{0.05}(2) = 5.991$. Thus the test result suggests that **we do not reject** the null hypothesis of different locations in the three groups.

12 MARKS

(c) With a normality assumption, we may use one-way ANOVA, provided that the variances in the three groups could be proved to be equal using Levene's Test.

4 MARKS

McGill University

Faculty of Science

MATH 204

PRINCIPLES OF STATISTICS II

Final Examination

Date: 17th April 2008 Time: 2pm-5pm

Examiner: Dr. David A. Stephens
Associate Examiner: Dr. Russell Steele

Please write your answers in the answer booklets provided.

This paper contains six questions. Each question carries 25 marks. Credit will be given for all questions attempted. The total mark available is 150 but rescaling of the final mark may occur.

Candidates may take one double-sided sheet of Letter-sized (216 $\times$ 279 mm) or A4-sized (210 $\times$ 297mm) paper with handwritten notes into the examination room.

Calculators may be used. Relevant statistical tables are provided.

Dictionaries and Translation dictionaries are permitted.

© 2008 McGill University

1. In an experimental study of the efficacy of treatments for depression, patients attending a clinic were allocated at random to one of a number different drug treatments and other therapies. The response measurement was the decrease in Hamilton Rating Depression Score (HRDS) over a three month course of treatment. The data recorded are tabulated below; entries in the rows of the table are the decrease in HRDS for each treatment. The final group was treated using an inactive treatment (or placebo).

Group					
Therapy 1	5.20	3.65	7.03	5.63	4.57
Therapy 2	7.05	7.91	1.46	4.11	4.18
Drug A	4.71	3.75	2.85	4.89	6.09
Drug B	6.82	6.77	8.31	6.57	6.55
Placebo	5.13	6.07	3.55	5.69	3.97

(a) What kind of design is being used in this study ? Identify the *factor(s)* being investigated, and for each factor state the number of levels that that factor has. Is the study *balanced* ? Is it *complete* ? Justify your answers.

6 MARKS

(b) Using the data, an ANOVA analysis is to be carried out. The ANOVA table below contains some missing entries marked by the notation $\star$. Write out the ANOVA table in full, filling in the missing values **using the information already given in the table**.

SOURCE	DF	SS	MS	F
GROUP	$\star$	19.613	$\star$	$\star$
ERROR	$\star$	$\star$	2.295	
TOTAL	$\star$	$\star$		

10 MARKS

(c) What is the conclusion of the ANOVA analysis ? State clearly the null and alternative hypothesis, the test statistic, the null distribution, and the conclusion.

A table of the Fisher-F distribution is provided on page 15. Entries in the table are the 0.05 tail quantile of the Fisher-F(ν_1, ν_2) distribution, for different values of ν_1 and ν_2.

5 MARKS

(d) The following table was also produced during an SPSS analysis of the data.

Test of Homogeneity of Variances

Y

Levene Statistic	df1	df2	Sig.
2.800	4	20	.054

Explain the relevance of this table in relation to the ANOVA. What steps would you carry out to confirm that the conclusion of the ANOVA is valid ?

4 MARKS

2. (a) Explain how to construct a complete balanced randomized block design to investigate the effect of treatment TMT (a factor taking k levels) in the presence of a blocking factor B (a factor taking b levels).

4 MARKS

(b) Using standard model notation, where, for example,

$$\text{TMT} + \text{B}$$

indicates the main effects only model with factors TMT and B, list the four models that can be fitted if the design does not have replication. Give the **total** number of parameters that each of the four models contains.

4 MARKS

(c) The efficacy of a blood pressure reduction treatment is to be investigated in an experimental study. Four dose levels (labelled 1,2,3,4) were investigated in three groups of patients (labelled 1,2,3, corresponding to 40-49 years, 50-59 years, over 60 years respectively); it was strongly believed before the study was carried out that the efficacy of the treatments would be different depending on the age of the patient. The recorded response was the reduction in blood pressure over four weeks of treatment.

The data that were collected in this complete and balanced design were analyzed in SPSS; the output is included on page 8. The variable **agegp** denotes the different age groups, whereas **tmt** denotes dose level.

(i) Explain the results presented in the ANOVA table.

6 MARKS

(ii) In terms of mean responses for the different subgroups of patients, give an interpretation of the interaction term, **agegp*tmt**. Use a sketch if necessary.

6 MARKS

(d) Using the Parameter Estimates table given on page 8, report an estimate of the difference in expected response between patients in Age-Group 40-49 at dose level 1 and patients in Age Group 60+ at dose level 4.

Remember that SPSS uses a **contrast parameterization**, with the highest level of a factor determining the baseline category.

5 MARKS

3. An experiment was carried out to investigate whether two drugs, levorphanol and epinephrine, can be used to reduce stress in adult and juvenile mice. Injections of one, other or both treatments were given to four adult and four juvenile animals, and the cortical sterone level (which reflects the stress-level) was measured. The data recorded are given in the table below. Note that measurements on three animals could not be made and are omitted from the table.

	Adult		Juvenile	
	Epinephrine		Epinephrine	
Levorphanol	No	Yes	No	Yes
No	1.11,0.84,1.47,4.45	1.67,1.53,2.84	3.24,1.89,2.17,2.25	3.02,1.74,3.62
Yes	0.17,1.41,-0.18,-0.17	1.54,2.60,2.22,2.63	0.39,-0.43,0.67,1.13	2.23,3.42,2.42

Output from the SPSS analysis of the response data is given on pages 9 and 10.

(a) What is the name of this kind of design ?

3 MARKS

(b) Using standard model notation, list the five models that were fitted in the SPSS analysis. Use L,E and A for the main effects of factor levorphanol, epinephrine and adult respectively.

6 MARKS

(c) Using the SPSS output, find the most appropriate model in ANOVA-F terms using **backward selection**. Recall that the ANOVA-F test for comparing nested models is based on the statistic

$$F = \frac{(\text{SSE}_R - \text{SSE}_C)/(k-g)}{\text{SSE}_C/(n-k-1)}$$

where

- SSE_R is the error sum of squares for the **Reduced Model**, specified using $g+1$ parameters including the intercept.
- SSE_C is the error sum of squares for the **Complete Model**, specified using $k+1$ parameters including the intercept.

If the reduced model is an adequate simplification of the complete model, then

$$F \sim \text{Fisher-F}(k-g, n-k-1)$$

Comment on the overall model fit of your selected model.

10 MARKS

(d) Explain why the p-values given within the ANOVA tables cannot be used formally to decide whether terms should be omitted from or included in the model.

2 MARKS

(e) Explain how to reanalyze the data for the **full factorial** model using a *permutation* (or *randomization*) procedure that would overcome the issue raised in part (d). Recall that in such a procedure, the permuted treatment sample sizes should be identical to those in the original experiment.

4 MARKS

4. The gas consumption of 56 houses was measured over a month, and the average outdoor temperature at the houses was also measured. Two types of houses were studied; 26 did not have cavity wall insulation, and 30 did have insulation.

Output from an SPSS analysis is recorded on pages 11 and 12. In the analysis, gas consumption is the response, **Insul** is a factor predictor for insulation (coded as 0 for no insulation, 1 for insulation) and **Temp** is a continuous covariate.

(a) Explain, using sketches, the five different models that can be fitted to the response that involve main effects and interactions involving **Insul** and **Temp**.

5 MARKS

(b) Using the results presented in the SPSS output, explain the most appropriate model to explain the observed response. Comment on the global fit of the model.

6 MARKS

(c) Comment on the residual plot given on page 12 obtained from the first model fitted. How does this residual plot assist in the analysis ? What other plots would you need ?

6 MARKS

(d) Basing your answer on the fit of the model

$$\textbf{Insul} + \textbf{Temp} + \textbf{Insul.Temp} \qquad (1)$$

compute an estimate, s^2, of the error variance, σ^2 in the regression model.

4 MARKS

(e) Basing your answer on the fit of the model in equation (1) predict the expected gas consumption for an uninsulated house in a month where the average temperature is 5 Celsius. By how much does the prediction change if the temperature is 6 Celsius ?

4 MARKS

5. (a) Explain how to carry out a Chi-squared test for **independence** of assignment of two factors, A and B, each of which takes two levels, using the statistic

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(n_{ij} - \widehat{n}_{ij})^2}{\widehat{n}_{ij}}$$

for a sample size of n. Explain what the terms in X^2 are, how to complete the test using tables of the Chi-squared distribution, and any assumptions that are required to hold in order for the test result to be valid.

8 MARKS

(b) Carry out a Chi-squared test of independence for the following 2×2 table that records the counts in a case-control study carried out to examine a hypothesized link between excema and hay fever in a sample of 11 year old children.

	Hay Fever		
Excema	Yes	No	Total
Yes	141	420	561
No	928	13525	14453
Total	1069	13945	15014

The table on page 16 contains the 0.05 and 0.01 tail quantiles of the Chisquared(ν) distribution, for $\nu = 1$ to 20.

6 MARKS

(c) In a first phase of a health study in a city, a random sample of size 2000 is to be obtained. The city is comprised (broadly) of five different ethnic subpopulations that make up 40 %, 30 %, 10 %, 10 % and 10% of the city population respectively.

A commercial company is employed to obtain the random sample, with the instruction that the sample should reflect the ethnic composition of the city. The sample they return is summarized in the following table.

	Ethnic Subpopulation				
	1	2	3	4	5
Expected Proportion (%)	40	30	10	10	10
Number in Sample	822	638	210	157	173

Using a Chi-squared test for this one-way layout, comment on whether the company have fulfilled their remit to produce a sample that reflects the ethnic composition of the city.

6 MARKS

(d) Using the usual guidelines under which the Chi-squared test is regarded to be valid, what is the smallest total sample size for which the test in part (c) could be considered ? Justify your answer.

5 MARKS

6. (a) Outline the differences between *paired* and *unpaired* (also known as *dependent* and *independent*) two sample tests in terms of the method of data collection.

4 MARKS

(b) In the table below, the measured central venous oxygen, SvO_2, in a group of 10 intensive care unit (ICU) patients, and 12 healthy controls is recorded. The study design involved measuring SvO_2 for the ICU patients at the time of admission (Time 0) and then six hours later (Time 1); for the healthy controls, two measurements were taken six hours apart. For all patients, the difference between the two measurements was computed.

ICU Patients			Healthy Controls		
Time 0	Time 1	Diff	Time 0	Time 1	Diff
59.1	56.7	-2.4	67.3	63.0	-4.3
58.2	60.7	2.5	64.8	64.3	-0.5
56.0	59.5	3.5	67.4	65.5	-1.9
65.3	59.8	-5.5	65.4	68.9	3.5
56.1	61.9	5.8	67.6	66.0	-1.6
60.6	67.7	7.1	63.7	65.1	1.4
37.8	50.0	12.2	64.5	68.2	3.7
39.7	52.9	13.2	68.2	63.3	-4.9
57.7	71.4	13.7	69.8	65.9	-3.9
33.6	51.3	17.7	64.3	66.6	2.3
			68.9	68.7	-0.2
			70.5	65.8	-4.7

After reading the SPSS output on pages 13 and 14 **carefully**, give answers to the following questions, and give supporting details of the statistical test results. In each case, indicate whether the test is **paired** or **unpaired**. Note that in most cases, the **asymptotic** test result only is quoted. **Note also that some of the SPSS output corresponds to inappropriate tests**.

(i) Do patients in the ICU exhibit a significant **increase** in SvO_2 level over the six hour period ?

5 MARKS

(ii) Is the amount by which SvO_2 changes over the six hour period **different for ICU patients and compared to healthy controls** ?

5 MARKS

(iii) Does SvO_2 level **change** significantly in healthy control subjects over a six hour period ?

5 MARKS

(c) Suppose that, in a new experimental study, three ICUs were to be compared to discover whether they were equally adept at increasing the SvO_2 measurement of patients in their care. Identify (by name) the type of design such a study would require, and identify one parametric test, and one non-parametric test, that could be used to assess the hypothesis of interest.

6 MARKS

Output for Question 2

Tests of Between-Subjects Effects

Dependent Variable:Y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	197.112[a]	11	17.919	4.357	.000
Intercept	1346.978	1	1346.978	327.517	.000
agegp	63.231	2	31.616	7.687	.001
tmt	57.541	3	19.180	4.664	.006
agegp * tmt	76.340	6	12.723	3.094	.012
Error	197.409	48	4.113		
Total	1741.500	60			
Corrected Total	394.521	59			

a. R Squared = .500 (Adjusted R Squared = .385)

Parameter Estimates

Dependent Variable:Y

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	8.780	.907	9.680	.000	6.956	10.603
[agegp=1]	-4.602	1.283	-3.588	.001	-7.181	-2.024
[agegp=2]	-3.732	1.283	-2.910	.005	-6.311	-1.153
[agegp=3]	0[a]	.	.	.	.	.
[tmt=1]	-3.240	1.283	-2.526	.015	-5.819	-.662
[tmt=2]	-4.315	1.283	-3.364	.002	-6.894	-1.736
[tmt=3]	-2.829	1.283	-2.206	.032	-5.408	-.251
[tmt=4]	0[a]	.	.	.	.	.
[agegp=1] * [tmt=1]	5.785	1.814	3.189	.003	2.138	9.432
[agegp=1] * [tmt=2]	2.133	1.814	1.176	.245	-1.514	5.780
[agegp=1] * [tmt=3]	1.343	1.814	.740	.463	-2.304	4.990
[agegp=1] * [tmt=4]	0[a]	.	.	.	.	.
[agegp=2] * [tmt=1]	1.199	1.814	.661	.512	-2.448	4.847
[agegp=2] * [tmt=2]	2.686	1.814	1.481	.145	-.961	6.333
[agegp=2] * [tmt=3]	2.848	1.814	1.570	.123	-.799	6.495
[agegp=2] * [tmt=4]	0[a]	.	.	.	.	.
[agegp=3] * [tmt=1]	0[a]	.	.	.	.	.
[agegp=3] * [tmt=2]	0[a]	.	.	.	.	.
[agegp=3] * [tmt=3]	0[a]	.	.	.	.	.
[agegp=3] * [tmt=4]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Output for Question 3: Part I

Tests of Between-Subjects Effects

Dependent Variable:Stress Reduction

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	24.099[a]	7	3.443	4.280	.004
Intercept	97.952	1	97.952	121.773	.000
lev	5.380	1	5.380	6.688	.017
epi	9.578	1	9.578	11.907	.002
adult	1.398	1	1.398	1.739	.202
epi * adult	.196	1	.196	.243	.627
lev * adult	.177	1	.177	.220	.644
lev * epi	6.224	1	6.224	7.738	.011
lev * epi * adult	.001	1	.001	.001	.973
Error	16.892	21	.804		
Total	133.830	29			
Corrected Total	40.991	28			

a. R Squared = .588 (Adjusted R Squared = .451)

Tests of Between-Subjects Effects

Dependent Variable:Stress Reduction

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	23.738[a]	4	5.935	8.256	.000
Intercept	98.312	1	98.312	136.761	.000
lev	5.410	1	5.410	7.526	.011
epi	9.605	1	9.605	13.362	.001
adult	1.275	1	1.275	1.773	.195
lev * epi	6.240	1	6.240	8.681	.007
Error	17.253	24	.719		
Total	133.830	29			
Corrected Total	40.991	28			

a. R Squared = .579 (Adjusted R Squared = .509)

Output for Question 3: Part II

Tests of Between-Subjects Effects

Dependent Variable:Stress Reduction

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	22.464[a]	3	7.488	10.104	.000
Intercept	97.642	1	97.642	131.753	.000
lev	5.605	1	5.605	7.564	.011
epi	9.370	1	9.370	12.644	.002
lev * epi	6.049	1	6.049	8.162	.008
Error	18.527	25	.741		
Total	133.830	29			
Corrected Total	40.991	28			

a. R Squared = .548 (Adjusted R Squared = .494)

Tests of Between-Subjects Effects

Dependent Variable:Stress Reduction

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	16.415[a]	2	8.207	8.683	.001
Intercept	99.902	1	99.902	105.688	.000
lev	6.992	1	6.992	7.397	.011
epi	10.040	1	10.040	10.622	.003
Error	24.577	26	.945		
Total	133.830	29			
Corrected Total	40.991	28			

a. R Squared = .400 (Adjusted R Squared = .354)

Tests of Between-Subjects Effects

Dependent Variable:Stress Reduction

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.374[a]	1	6.374	4.972	.034
Intercept	94.412	1	94.412	73.639	.000
lev	6.374	1	6.374	4.972	.034
Error	34.617	27	1.282		
Total	133.830	29			
Corrected Total	40.991	28			

a. R Squared = .156 (Adjusted R Squared = .124)

Output for Question 4: Part I

Between-Subjects Factors

		Value Label	N
Insulation	0	Before	26
	1	After	30

Tests of Between-Subjects Effects

Dependent Variable:Gas Consumption (1000 cu. ft.)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	69.589[a]	3	23.196	222.333	.000
Intercept	431.193	1	431.193	4132.905	.000
Insul	14.594	1	14.594	139.882	.000
Temp	45.577	1	45.577	436.850	.000
Insul * Temp	1.345	1	1.345	12.893	.001
Error	5.425	52	.104		
Total	1003.300	56			
Corrected Total	75.014	55			

a. R Squared = .928 (Adjusted R Squared = .924)

Parameter Estimates

Dependent Variable:Gas Consumption (1000 cu. ft.)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.724	.118	40.000	.000	4.487	4.961
[Insul=0]	2.130	.180	11.827	.000	1.769	2.491
[Insul=1]	0[a]	.	.	.	.	.
Temp	-.278	.023	-12.124	.000	-.324	-.232
[Insul=0] * Temp	-.115	.032	-3.591	.001	-.180	-.051
[Insul=1] * Temp	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Output for Question 4: Part II

Tests of Between-Subjects Effects

Dependent Variable:Gas Consumption (1000 cu. ft.)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	68.244[a]	2	34.122	267.114	.000
Intercept	429.865	1	429.865	3365.077	.000
Insul	33.224	1	33.224	260.088	.000
Temp	45.896	1	45.896	359.286	.000
Error	6.770	53	.128		
Total	1003.300	56			
Corrected Total	75.014	55			

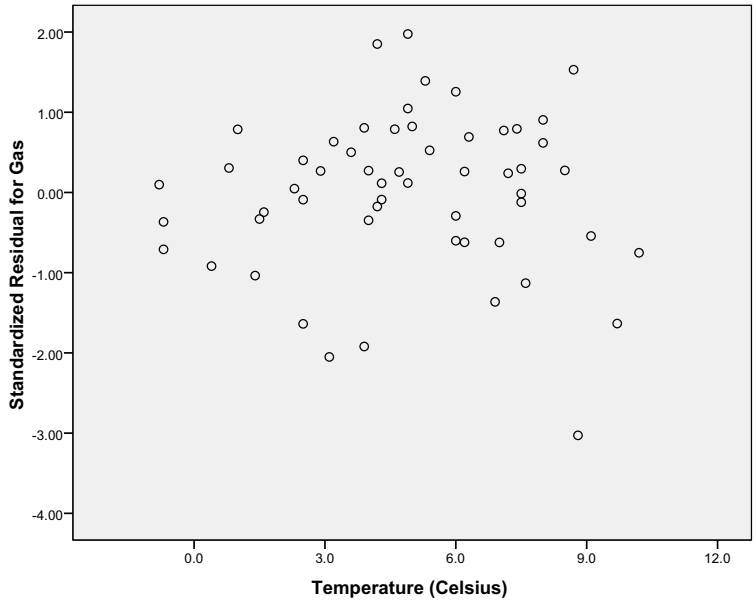
a. R Squared = .910 (Adjusted R Squared = .906)

Parameter Estimates

Dependent Variable:Gas Consumption (1000 cu. ft.)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.986	.103	48.558	.000	4.780	5.192
[Insul=0]	1.565	.097	16.127	.000	1.371	1.760
[Insul=1]	0[a]	.	.	.	.	.
Temp	-.337	.018	-18.955	.000	-.372	-.301

a. This parameter is set to zero because it is redundant.



Output for Question 6: Part I

Analysis 1: Mann-Whitney Test

Ranks

	Time	N	Mean Rank	Sum of Ranks
Sv02	Time 0	22	21.61	475.50
	Time 1	22	23.39	514.50
	Total	44		

Test Statistics[a]

	Sv02 at admission
Mann-Whitney U	222.500
Wilcoxon W	475.500
Z	-.458
Asymp. Sig. (2-tailed)	.647

a. Grouping Variable: Time

Analysis 2: Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
ICUTIME1 - ICUTIME0	Negative Ranks	2[a]	2.50	5.00
	Positive Ranks	8[b]	6.25	50.00
	Ties	0[c]		
	Total	10		

a. ICUTIME1 < ICUTIME0

b. ICUTIME1 > ICUTIME0

c. ICUTIME1 = ICUTIME0

Test Statistics[b]

	ICUTIME1 - ICUTIME0
Z	-2.293[a]
Asymp. Sig. (2-tailed)	.022

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Analysis 3 : Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
HEATIME1 - HEATIME0	Negative Ranks	8[a]	6.75	54.00
	Positive Ranks	4[b]	6.00	24.00
	Ties	0[c]		
	Total	12		

a. HEATIME1 < HEATIME0

b. HEATIME1 > HEATIME0

c. HEATIME1 = HEATIME0

Test Statistics[b]

	HEATIME1 - HEATIME0
Z	-1.177[a]
Asymp. Sig. (2-tailed)	.239

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Output for Question 6: Part II

Analysis 4 : Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
Sv02 after 6 hours - Sv02 at admission	Negative Ranks	10[a]	8.90	89.00
	Positive Ranks	12[b]	13.67	164.00
	Ties	0[c]		
	Total	22		

a. Sv02 after 6 hours < Sv02 at admission

b. Sv02 after 6 hours > Sv02 at admission

c. Sv02 after 6 hours = Sv02 at admission

Test Statistics[b]

	Sv02 after 6 hours - Sv02 at admission
Z	-1.218[a]
Asymp. Sig. (2-tailed)	.223

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Analysis 5 : Mann-Whitney Test

Ranks

	Patie…	N	Mean Rank	Sum of Ranks
Difference	ICU	10	15.15	151.50
	Healthy	12	8.46	101.50
	Total	22		

Test Statistics[b]

	Difference
Mann-Whitney U	23.500
Wilcoxon W	101.500
Z	-2.407
Asymp. Sig. (2-tailed)	.016
Exact Sig. [2*(1-tailed Sig.)]	.014[a]

a. Not corrected for ties.

b. Grouping Variable: Patient Group

Table of the Fisher-F distribution

Entries in table are the $\alpha = 0.05$ tail quantile of Fisher-F(ν_1, ν_2) distribution
ν_1 given in columns, ν_2 given in rows.

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91	244.69	245.36	245.95	246.46	246.92	247.32	247.69	248.01	248.31	248.58	248.83	249.05	249.26
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43	19.43	19.44	19.44	19.44	19.45	19.45	19.45	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67	8.67	8.66	8.65	8.65	8.64	8.64	8.63
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82	5.81	5.80	5.79	5.79	5.78	5.77	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58	4.57	4.56	4.55	4.54	4.53	4.53	4.52
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90	3.88	3.87	3.86	3.86	3.85	3.84	3.83
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47	3.46	3.44	3.43	3.43	3.42	3.41	3.40
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17	3.16	3.15	3.14	3.13	3.12	3.12	3.11
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94	2.93	2.92	2.91	2.90	2.89
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79	2.77	2.76	2.75	2.75	2.74	2.73
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66	2.65	2.64	2.63	2.62	2.61	2.60
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56	2.54	2.53	2.52	2.51	2.51	2.50
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47	2.46	2.45	2.44	2.43	2.42	2.41
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40	2.39	2.38	2.37	2.36	2.35	2.34
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34	2.33	2.32	2.31	2.30	2.29	2.28
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29	2.28	2.26	2.25	2.24	2.24	2.23
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24	2.23	2.22	2.21	2.20	2.19	2.18
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20	2.19	2.18	2.17	2.16	2.15	2.14
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17	2.16	2.14	2.13	2.12	2.11	2.11
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15	2.14	2.12	2.11	2.10	2.09	2.08	2.07
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.11	2.10	2.08	2.07	2.06	2.05	2.05
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10	2.08	2.07	2.06	2.05	2.04	2.03	2.02
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15	2.13	2.11	2.09	2.08	2.06	2.05	2.04	2.02	2.01	2.01	2.00
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11	2.09	2.07	2.05	2.04	2.03	2.01	2.00	1.99	1.98	1.97
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04	2.02	2.01	2.00	1.98	1.97	1.96	1.96
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02	2.00	1.99	1.98	1.97	1.96	1.95	1.94
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.10	2.08	2.06	2.04	2.02	2.00	1.99	1.97	1.96	1.95	1.94	1.93	1.92
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99	1.97	1.96	1.95	1.93	1.92	1.91	1.91
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08	2.05	2.03	2.01	1.99	1.97	1.96	1.94	1.93	1.92	1.91	1.90	1.89
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96	1.95	1.93	1.92	1.91	1.90	1.89	1.88
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	2.11	2.08	2.05	2.03	2.00	1.98	1.96	1.95	1.93	1.92	1.91	1.90	1.88	1.88	1.87
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07	2.04	2.01	1.99	1.97	1.95	1.94	1.92	1.91	1.90	1.88	1.87	1.86	1.85

Table of the Chisquared(ν) distribution

Entries in table are the $\alpha = 0.05$ and $\alpha = 0.01$ tail quantiles

ν	1	2	3	4	5	6	7	8	9	10
$\alpha = 0.05$	3.841	5.991	7.815	9.488	11.070	12.592	14.067	15.507	16.919	18.307
$\alpha = 0.01$	6.635	9.210	11.345	13.277	15.086	16.812	18.475	20.090	21.666	23.209

ν	11	12	13	14	15	16	17	18	19	20
$\alpha = 0.05$	19.675	21.026	22.362	23.685	24.996	26.296	27.587	28.869	30.144	31.410
$\alpha = 0.01$	24.725	26.217	27.688	29.141	30.578	32.000	33.409	34.805	36.191	37.566

McGill University
Faculty of Science

Department of Mathematics and Statistics

April 2008

MATH 204

PRINCIPLES OF STATISTICS II

SOLUTIONS

1. (a) This is a completely randomized design with a single factor, Group, that takes five levels. It is a balanced and complete design, as all treatments (factor levels) have the same non-zero number (five) of observations.

6 MARKS

(b) The completed table is

SOURCE	DF	SS	MS	F
GROUP	4	19.613	4.903	2.136
ERROR	20	45.903	2.295	
TOTAL	24	65.516		

10 MARKS

(c) The ANOVA F-test tests the hypotheses

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1 &: \mu_i \neq \mu_j, \text{ for some } i, j \end{aligned}$$

The null distribution is Fisher$(4, 20)$; comparing the F-statistic (2.136) with the 95 % quantile of the Fisher$(4, 20)$ distribution, we see from the table given that the critical value is 2.87, so we do not reject H_0 - there is no evidence that there is any difference in HDRS reduction.

5 MARKS

(d) The p-value in this table reveals that Levene's test for homogeneity of variances does not reject the null hypothesis of equal variances at the $\alpha = 0.05$ significance level, although the result is almost significant. Equality of variances is essential for the ANOVA F-test to give a valid result.

Given the study design, with independent observations being used, the only assumption that needs to be checked is the Normality assumption, that could be checked using boxplots (or residual plots). Also, to verify the results without considering the normality assumptions, could use permutation or randomization tests.

4 MARKS

2. (a) Sample kr experimental units from each of the b subpopulations identified by the levels of the blocking factor, and then allocate r experimental units at random to each of the treatments.

4 MARKS

(b) The four models are

$$1 \qquad \text{TMT} \qquad \text{B} \qquad \text{TMT} + \text{B}$$

which have 1, k, b and $k + b - 1$ parameters in total. No interaction models can be fitted if there are no replicates.

4 MARKS

(c) (i) The ANOVA table indicates that both main effects and interaction are significant at $\alpha = 0.05$. The blocking factor is confirmed as leading to a significantly different response. The interaction term is not significant at $\alpha = 0.01$, indicating that this term is not as important as the other terms in the model.

6 MARKS

(ii) The significant interaction means that the effect of changing treatment level is **different** in the **different** age groups, that is, in a plot demonstrating the group means, lines plotting mean response against age group are **not piecewise parallel**.

6 MARKS

(d) The Age Group 60+ at dose level 4 subgroup corresponds to the baseline group, thus we can get the difference in expected response by adding the main effect for Age Group 1, the main effect for dose level 1, and the interaction for those two groups. This gives

$$(-3.240) + (-4.602) + 5.785 = -2.057$$

5 MARKS

3. (a) This is a factorial design that is complete but not balanced.

3 MARKS

(b) The five models are

Model 1	L + E + A + L.E + L.A + A.E + L.E.A
Model 2	L + E + A + L.E
Model 3	L + E + L.E
Model 4	L + E
Model 5	L

6 MARKS

(c) Extracting the analysis results we have

Comparison	SSE_C	SSE_R	k	g	F	F_α	Significant at $\alpha = 0.05$
1 vs 2	7	4	16.892	17.253	0.150	3.07	No
2 vs 3	4	3	17.253	18.527	1.772	4.26	No
3 vs 4	3	2	18.527	24.577	8.164	4.24	Yes
4 vs 5	2	1	24.577	34.617	10.621	4.23	Yes

Thus it seems that model 3 is the most appropriate model.

The overall fit is adequate with an R^2 statistic of 0.548.

10 MARKS

(d) Because the study is **not balanced**.

2 MARKS

(e) The simplest way to carry out the procedure would be to form a vector of the response variables, and a list of their corresponding subgroup allocations, and then to cycle through the permutations of the data indices, computing the Fisher-F statistics for the whole table for any hypothesized model. The p-values are computed by enumerating how many times the Fisher-F statistics are more extreme than the observed one.

This may be computationally challenging as the sample size is large, so perhaps sampling permutations would be more amenable.

4 MARKS

4. (a) The five sketches are: (0) Null Model: One horizontal straight line, (1) **Insul**: Two horizontal straight lines, (2) **Temp**: One straight line, (3) **Insul+Temp**: Two parallel straight lines, (4) **Insul*Temp**: Two non-parallel straight lines.

5 MARKS

(b) The full factorial model is the most appropriate; all terms appear to be significant in the ANOVA and in the parameter estimates table. The R^2 value for this model is very high (> 0.9), indicating a very good fit.

6 MARKS

(c) The standardized residual plot against temperature should look like a horizontal band within the bounds ± 2; here, it does, so overall the plot gives no cause for concern - possibly there is an outlier, but this is not certain. The plot allows a check of the necessary assumption that the variance of the residual errors does not depend on the values of the predictors. To check the other assumptions, specifically normality of the residuals, a histogram and P-P plot would be useful.

6 MARKS

(d) We have

$$s^2 = \mathsf{MSE} = \frac{\mathsf{SSE}}{\mathsf{rdf}} = 0.104$$

4 MARKS

(e) The best prediction is

$$((4.724) + (2.130)) + ((-0.278) + (-0.115)) \times 5 = 4.889$$

The answer changes by an amount equal to the coefficient for temperature for uninsulated houses, that is

$$(-0.278) + (-0.115) = -0.393$$

indicating a decrease of 393 cubic feet.

4 MARKS

5. (a) In X^2, the n_{ij} and $\widehat{n}_{ij}$ are the observed and expected frequencies in the cells of the table, where

$$\widehat{n}_{ij} = \frac{n_{i.}n_{.j}}{n} = \frac{i\text{th Row Total} \times j\text{th Column Total}}{\text{Total Sample Size}}$$

If the sample size is large enough, so that the expected counts are at least five, then $X^2 \dot{\sim} \chi^2_1$, so we compare the computed test statistic with the 0.95 quantile of this distribution 3.841.

8 MARKS

(b) We have fitted values

	Hay Fever		
Excema	Yes	No	Total
Yes	39.94	521.06	561.00
No	1029.06	13423.94	14453.00
Total	1069.00	13945.00	15014.00

and hence

$$X^2 = 286.00 \qquad \Longrightarrow \qquad \text{Reject } H_0$$

6 MARKS

(c) We carry out a one-way Chi-squared test of the hypothesis

$$H_0 \; : \; p_1 = 0.4, p_2 = 0.3, p_3 = 0.1, p_4 = 0.1, p_5 = 0.1$$

	Ethnic Subpopulation				
	1	2	3	4	5
Expected Proportion (%)	40	30	10	10	10
Expected Count	800	600	200	200	200
Observed Count	822	638	210	157	173
$(O_i - E_i)^2/E_i$	0.60	2.41	0.50	9.24	3.65

and hence

$$X^2 = 16.402 > 13.277 = \text{Chisq}_{0.01}(4) \qquad \Longrightarrow \qquad \text{Reject } H_0$$

so it seems that the remit has not been met.

6 MARKS

(d) We would need a sample size of 50, so that the smallest expected count is 5.

5 MARKS

6. (a) Paired data have dependence between the two samples, either by matching of the experimental units, or because of repeated sampling of the same individuals. Unpaired data are completely independent.

4 MARKS

(b) (i) Yes: this comes from Analysis 2, where the two-tailed test rejects the null hypothesis of equal medians at $\alpha = 0.05$, and the sum of positive ranks is greater than the sum of positive ranks, so the left-tailed test also rejects the null at $\alpha = 0.05$.

5 MARKS

(ii) Yes: this comes from Analysis 5, where the two-tailed p-value is 0.016.

5 MARKS

(iii) No: this comes from Analysis 3, where the two-tailed p-value is 0.239.

5 MARKS

(c) This would be a Completely Randomized Design, where the null hypothesis could be tested using one-way ANOVA (parametric) or the Kruskal-Wallis test (non-parametric).

6 MARKS