TRENT UNIVERSITY
Faculty of Arts and Science
Final Examinations – 2000/2001

MATHEMATICS 150

PART B: Time: 2 hours and 30 minutes
Books, notes, calculators and 'laptop' computers (with battery power supply) may be used.
Solutions to Part A must be submitted before Part B is commenced.
Each question is worth 16 marks
The **four** questions that you answer best will be counted.

1. a) An office issuing permits has experienced quarterly variation in numbers of permits issued over the past years that indicated a fairly stable pattern with seasonal factors for the four quarters of the year as 0.82 1.04, 1.21. 0.93.

i) For the most recent year, the numbers of permits issued in quarters 1 through 4, respectively were: 762, 989, 1136, 904. What were the seasonally adjusted values?

ii) Numbers of permits issued are assumed to be following a trend line $y = 800 + 9t$ where t is time in quarters and the most recent year was year 5. Predict numbers of permits to be issued for each quarter of the next year (i.e. year 6.)

b) For the most recent ten working days, the numbers of permits (per day) issued (from smallest to largest) were:

9 10 12 13 14 14 16 17 19 24

Comparable figures for another office were:

7 8 10 12 14 15 16 16 17 20

Compare the permits issued by these offices in the ten days with a box-and-whisker plot.

c) In year 1, the average number of working hours from receipt of an application to issuing of a permit was estimated to be 48 hours. In year 5 it was estimated to be 44 hours.

i) What was the total percentage decrease in the estimated number of hours?

ii) What was the average annual year-over-year percentage decrease in the estimated number of hours?

2. A study on lead concentration in wood involved an investigation of sample concentrations reported at five year intervals as listed in the following data.

Concentration of Lead in Wood Samples (µg/g)

year	concentration	year	concentration
1925	2.3	1960	10.1
1930	4.7	1965	11.3
1935	5.9	1970	13.0
1940	6.5	1975	14.7
1945	5.6	1980	14.9
1950	7.8	1985	14.7
1955	7.4	1990	13.9
		1995	12.8

A Minitab session with these data produced the following three pages of output. The year and lead concentration values were entered into C1 and C2, respectively.

2. (Continued)

```
MTB > let c3 = (c1-1925)/5
MTB > let c4=c3*c3
MTB > let c5=c3*c3*c3
```

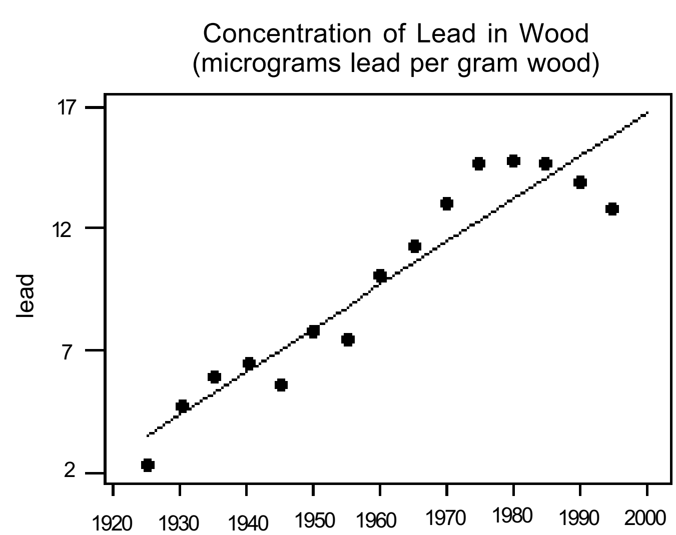
Correlations (Pearson)

```
Correlation of lead and x = 0.941
```

Regression Analysis

```
The regression equation is
lead = 3.51 + 0.884 x

15 cases used 1 cases contain missing values

Predictor        Coef       StDev          T        P
Constant       3.5125      0.7246       4.85    0.000
x             0.88393     0.08809      10.03    0.000

S = 1.474       R-Sq = 88.6%     R-Sq(adj) = 87.7%

Analysis of Variance
Source             DF          SS          MS         F        P
Regression          1      218.77      218.77    100.68    0.000
Residual Error     13       28.25        2.17
Total              14      247.02
```

Plot

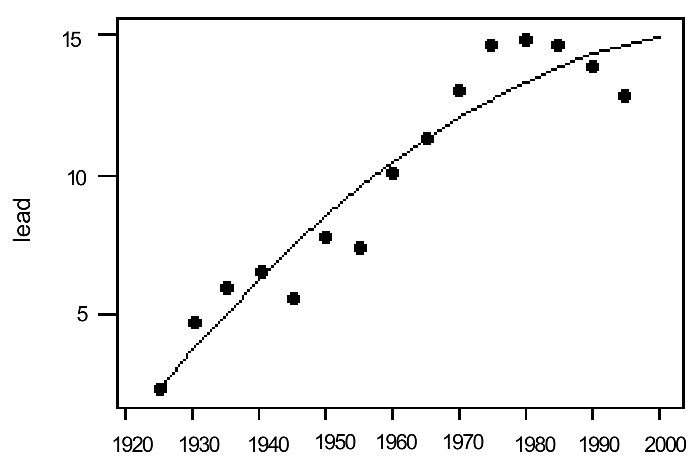


2. (Continued)

Regression Analysis

```
The regression equation is
lead = 2.29 + 1.45 x - 0.0403 x-squared

15 cases used 1 cases contain missing values

Predictor        Coef       StDev          T        P
Constant       2.2912      0.9137       2.51    0.028
x              1.4476      0.3029       4.78    0.000
x-square     -0.04026     0.02087      -1.93    0.078

S = 1.340       R-Sq = 91.3%     R-Sq(adj) = 89.8%

Analysis of Variance
Source            DF          SS          MS          F        P
Regression         2      225.46      112.73      62.74    0.000
Residual Error    12       21.56        1.80
Total             14      247.02
```

Plot



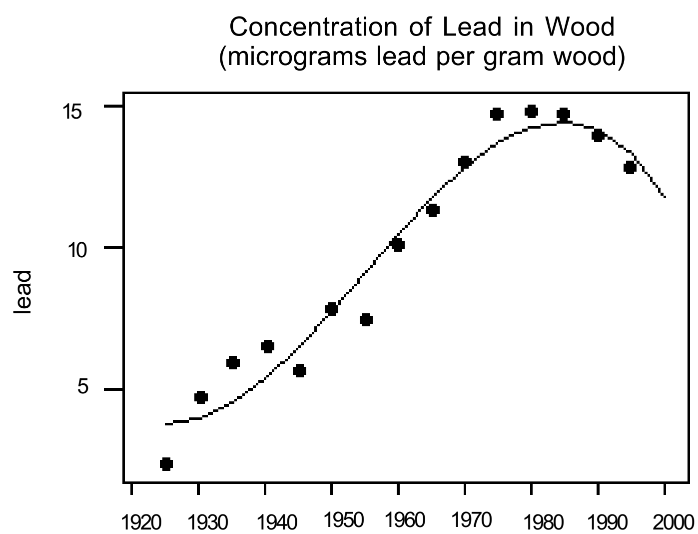
Regression Analysis

```
The regression equation is
lead = 3.73 - 0.050 x + 0.237 x-squared - 0.0132 x-cubed

15 cases used 1 cases contain missing values

Predictor        Coef       StDev          T        P
Constant       3.7312      0.8423       4.43    0.001
x             -0.0504      0.5398      -0.09    0.927
x-square      0.23667     0.09150       2.59    0.025
x-cubed     -0.013187    0.004290      -3.07    0.011
```

2. (Continued)

```
S = 1.027        R-Sq = 95.3%     R-Sq(adj) = 94.0%

Analysis of Variance
Source            DF         SS          MS         F        P
Regression         3    235.422      78.474     74.43    0.000
Residual Error    11     11.598       1.054
Total             14    247.020
```

Plot



Data Display

Row	year	lead	x	x-squared	x-cubed	lin-fit	quad-fit	cube-fit
1	1925	2.3	0	0	0	3.5125	2.2912	3.7312
2	1930	4.7	1	1	1	4.3964	3.6985	3.9042
3	1935	5.9	2	4	8	5.2804	5.0254	4.4715
4	1940	6.5	3	9	27	6.1643	6.2717	5.3538
5	1945	5.6	4	16	64	7.0482	7.4374	6.4721
6	1950	7.8	5	25	125	7.9321	8.5227	7.7473
7	1955	7.4	6	36	216	8.8161	9.5274	9.1001
8	1960	10.1	7	49	343	9.7000	10.4516	10.4516
9	1965	11.3	8	64	512	10.5839	11.2952	11.7225
10	1970	13.0	9	81	729	11.4679	12.0584	12.8338
11	1975	14.7	10	100	1000	12.3518	12.7410	13.7063
12	1980	14.8	11	121	1331	13.2357	13.3431	14.2609
13	1985	14.7	12	144	1728	14.1196	13.8646	14.4185
14	1990	13.9	13	169	2197	15.0036	14.3057	14.1000
15	1995	12.8	14	196	2744	15.8875	14.6662	13.2261
16	2000	*	15	225	3375	16.7714	14.9462	11.7179

2. (Continued)

a) Note the calculations that produced C3, C4, C5. Note that C3 = x; C4 = x^2; C4 = x^3. What is x in terms of year? What is year in terms of x? What is x if year = 2000?
b) What is the correlation between the lead concentration and year? How is this found in the Minitab output? Does this indicate a very strong linear relation between lead and year?
c) Does the data scatter indicate a linear trend of lead over time?
d) If a linear trend is used, what is the prediction equation for lead *vs* year?
e) How much is the model improved
 i) by using a quadratic equation instead of a linear one?
 ii) by using a cubic equation instead of a linear one?
f) What would have been the predicted lead concentration for 2000? What model did you use for this prediction? Why did you use this model?

3. A professional organization has 3785 members. Membership in the association includes a subscription to the newsletter and the main journal.
a) How many of the members of the association must be included in a random sample if it is of interest to estimate the percentage who read the main editorial in the most recent newsletter and if it is desired to have a 95% chance of estimating the percentage to within 7.5 percentage points?
b) If a sample of 165 members included 50 who had read the editorial, determine a 95% confidence limit to indicate at least how many of the 3785 members (i.e. what number) read the editorial. Do these data indicate that over 1000 of the members read the editorial?
c) According to a new proposal, it would be possible to have the membership and newsletter without the journal, or to have the journal without the membership and newsletter (or to continue with the membership, newsletter and journal.) In a survey of 200 long-term members and 80 more recent members, respondents were asked which option they would prefer if the three options were available. The results were as tabulated below. Do the preferences of the long-term and more recent members differ significantly? What is the *P*-value?

Membership/Journal Preferences — Long-Term vs Recent Members
Numbers of Members

Member Type	Preference		
	Membership Only	Membership & Journal	Journal Only
Long-Term	30	122	48
Recent	5	46	29

4 . The compacted Minitab session output on the next page was used to analyze lengths (mm) of clams from two strata — one just below and one just above the mid-tide level.

4. (Continued)

```
Data Display
below
    475    480    525    457    380    424    529    470    419    473
    516    457    358    513    514    488    488    467    520    440
    501    440    474    475
Data Display
above
    520    521    469    485    469    435    341    347    368    434
    190    509    378    284    445    442    319    331
MTB > gstd
MTB > dotplot c1 c2;
SUBC> same.
Dotplot
                                                        :
                              .  .     .. : : ::.: .::..
 -------+---------+---------+---------+---------+---------below

     .            .    .. ..  ..      :..  : .   .:
 -------+---------+---------+---------+---------+---------above
      210       280       350       420       490       560

MTB > stack c1 c2 c5;
SUBC> subs c4.
Homogeneity of Variance
Response    length
Factors     level
ConfLvl     95.0000

F-Test (normal distribution)
Test Statistic: 4.288
P-Value       : 0.001

Two Sample T-Test and Confidence Interval

Two sample T for below vs above

        N      Mean     StDev   SE Mean
below  24     470.1      43.8       8.9
above  18     404.8      90.8        21

90% CI for mu below - mu above: ( 25.5,  105)
T-Test mu below = mu above (vs not =): T = 2.82  P = 0.010  DF = 22

Two Sample T-Test and Confidence Interval

Two sample T for below vs above

        N      Mean     StDev   SE Mean
below  24     470.1      43.8       8.9
above  18     404.8      90.8        21

95% CI for mu below - mu above: ( 22.5,  108)
T-Test mu below = mu above (vs not =): T = 3.08  P = 0.0037  DF = 40
Both use Pooled StDev = 67.9
```

4. (Continued)

a) What differences appear to exist between the strata with regard to the general level and the variability of lengths on the basis of the comparison dot diagram display?
b) Although the dot diagrams may appear to indicate otherwise, assume that the lengths in general are reasonably well represented by normal distributions. Do these data provide evidence at the 5% level that the population variances differ?
c) Use the conclusion from b) to select the appropriate part of the output and hence provide a lower 95% limit for the difference $\mu_{\text{below}} - \mu_{\text{above}}$. Why did you use the part of the output that you selected?
d) Calculate 95% confidence limits for the mean length for clams just below the mid-tide level.

5. a) If 5 test samples of sewage effluent are taken at random from a batch of 20 samples of which 8 came from monitoring station A and 12 from station B, what is the probability that 2 of the 5 will be from station A?
b) If 5 test samples of sewage effluent are taken at random from a batch of 200 samples of which 80 came from monitoring station A and 120 from station B, what is the probability that at most 2 of the 5 will be from station A?
c) If the amount of residue of a common drug in sewage effluent samples is well represented by a normal distribution with a mean of 3500 nanograms per litre and a standard deviation of 250 nanograms per litre, what is the probability that the residue in a given sample will be below 3250 nanograms per litre?
d) If the mean and standard deviation of 16 samples of residue in effluent from a given source were 3640 and 280 nanograms per litre, respectively, and if a normal distribution is assumed to represent well the residue values, would the sample data provide sufficient evidence at the 5% level that the mean residue exceeds 3500 nanograms per litre?

6. a) Prior to development of an appropriate portable power source for field equipment a low-capacity form of power source was used. For 180 field monitors, this type of source functioned adequately 103 times, but failed to function adequately the other 77 times. The power source setup was adjusted in an attempt to improve its performance, with the anticipation that, in some cases, the adjustment might reduce performance. The adjusted units were used with the same 180 field monitors. Of the 103 that had functioned adequately before adjustment, 9 did not function adequately after adjustment. Of the 77 that originally did not function adequately, 27 did function adequately after the adjustment.
i) Is there evidence at the 5% level that the adjustment improves the percentage of cases in which the source will function adequately?
ii) Is there evidence at the 1% level that the adjustment improves the percentage of cases in which the source will function adequately?

6. (Continued)

b) For eight of the units that functioned well before and after adjustment, times of full adequate power were noted before and after adjustment. Sample times in minutes were:

Unit	1	2	3	4	5	6	7	8
Time **after** Adjustment	347	302	306	320	339	379	323	394
Time **before** Adjustment	296	262	245	272	283	322	282	342

Determine 95% confidence limits for the average increase in time with the adjustment.

7. The Minitab session below and on the next page includes output from a session to analyze data used to compare depth measurements for five methods of groove cutting (actually three methods with some 'sub-methods.')

a) Compare depths for the five methods with a mean-and-standard deviation display.

b) What null and alternative hypotheses are implied in using the section labelled "`Analysis of Variance for depth`" in the output?

c) Should the null hypothesis in b) be accepted or rejected? What is the *P*-value?

d) Do the data provide evidence at the 1% level that the methods are not all equivalent with regard to mean depth?

e) What differences between methods, if any, are identified as significant differences? (Use labels such as I-A, II etc. not 1, 2, etc. to label differences.)

```
Descriptive Statistics

Variable          N      Mean     Median     TrMean      StDev    SE Mean
I-A               7    2.3686     2.3700     2.3686     0.0471     0.0178
II                9    2.2478     2.2600     2.2478     0.0497     0.0166
III-A             6    2.1767     2.1750     2.1767     0.0234     0.0095
III-b             8    2.1825     2.1850     2.1825     0.0483     0.0171
I-B               7    2.3557     2.3600     2.3557     0.0424     0.0160

Variable     Minimum    Maximum         Q1         Q3
I-A           2.2900     2.4300     2.3400     2.4100
II            2.1600     2.3100     2.2050     2.2900
III-A         2.1400     2.2100     2.1625     2.1950
III-b         2.1100     2.2500     2.1425     2.2300
I-B           2.2900     2.4100     2.3100     2.3900

MTB > stack c1-c5 c11;
SUBC> subs c10.

One-way Analysis of Variance

Analysis of Variance for depth
Source     DF        SS        MS        F        P
method      4   0.23662   0.05916    30.10    0.000
Error      32   0.06290   0.00197
Total      36   0.29952
```

7. (Continued)

```
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ---+---------+---------+---------+---
1           7    2.3686    0.0471                          (---*---)
2           9    2.2478    0.0497           (---*---)
3           6    2.1767    0.0234  (----*----)
4           8    2.1825    0.0483    (---*---)
5           7    2.3557    0.0424                        (---*----)
                                   ---+---------+---------+---------+---
Pooled StDev =   0.0443             2.160     2.240     2.320     2.400

Fisher's pairwise comparisons

    Family error rate = 0.0701
Individual error rate = 0.0100

Critical value = 2.738

Intervals for (column level mean) - (row level mean)

                  1           2           3           4

       2    0.05962
            0.18197

       3    0.12437     0.00713
            0.25944     0.13509

       4    0.12325     0.00629    -0.07139
            0.24889     0.12426     0.05972

       5   -0.05203    -0.16911    -0.24658    -0.23604
            0.07774    -0.04676    -0.11151    -0.11039
```