TRENT UNIVERSITY

Faculty of Arts and Science

Final Examinations — 1998/99

MATHEMATICS-STATISTICS 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 major-equipment service facility has experienced seasonal variation according to the quarter of the year. A four-year record of the numbers of service calls is shown in the following table with two columns of additional calculations as indicated.

A	B	C	D	E	F
year	quarter	time	number of service calls	centred moving average	D/E
1	1	1	272		
	2	2	144		
	3	3	88	189	0.466
	4	4	244	192	1.271
2	1	5	288	193	1.492
	2	6	152	198	0.768
	3	7	88	209	0.421
	4	8	284	219	1.297
3	1	9	336	227	1.480
	2	10	184	229	0.803
	3	11	120	234	0.513
	4	12	268	243	1.103
4	1	13	392	246	1.593
	2	14	200	254	0.787
	3	15	128		
	4	16	324		

a) Sketch a time series plot of the data.
b) As noted, column F is obtained by dividing the entry in column D by that in column E. What does column F represent?
c) What should be the average value of the empirical seasonal factors for the four quarters?
d) Determine the empirical seasonal factors for these data.
e) Determine the seasonally adjusted values for year 4.
f) The underlying trend has been estimated to be $y = 166 + 6.4\,t$ where y is the number of service calls in the quarter and t is the time value for the quarter.
   i) What are the values of t for the four quarters of year 5?
   ii) Determine the forecast values for the four quarters of year 5.

2 . Five ordinary decks of playing cards (52 cards in each deck — 13 spades [black], 13 hearts [red], 13 diamonds [red], 13 clubs [black]) are indistinguishable. They are designated as decks *A, B, C, D,* and *E*. The diamonds are removed from decks *A, B* and *C* and the clubs are removed from decks *D* and *E*.

a) If one deck is selected at random and one card is selected at random from that deck, what is the probability that the card will be red?

b) If one deck is selected at random and three cards are selected at random from that deck, one at a time *with replacement*, what is the probability that all three cards will be red?

c) One deck was selected at random and three cards were selected at random from that deck, one at a time *with replacement*. All three cards were red. What is the probability that the deck was one with the clubs removed?

d) One deck was selected at random and three cards were selected at random from that deck, one at a time *with replacement*. All three cards were red. The deck of cards now is to be inspected. If it is a deck with the clubs removed, you will win $10.00 from an opponent. If it is a deck with the diamonds removed, the opponent will win *X* dollars from you. What is *X* (nearest cent) if this is a "fair" game?

3 . Studies on operating life for some system components have indicated that the components seem to lose the same percentage of strength each year, thus strength suffers exponential decay. Average strength was determined for several study components for years 3 through 15 and the data were analyzed in MINITAB as shown below and on the following **two** pages.

a) The data first were analyzed with ordinary regression to predict strength *s* on the basis of time *t*.

i) What is the corresponding prediction equation? Is it a good model in terms of correlation?

ii) Predict *s* for $t = 12$ on the basis of the model in i).

iii) What extrapolation problem, if any, does this model have?

b) The model was revised to $s = a_0 \times (1 + c)^t$ with a further re-expression as $y = b_0 + b_1 x$ where $y = ln(s)$ and $x = t$.

i) What is the prediction equation to predict *y* from *x* ? Is it a good model in terms of correlation?

ii) How is *s* found from *y* ? Predict *y* for $x = 12$ and, hence, predict *s* for $t = 12$.

iii) What is the equation to predict s from *t* in the revised model?

iv) Does this model appear reasonable visually?

v) How is *c* found from b_1 ? What was the annual percentage decrease of the underlying trend?

c) i) What was the actual percentage decrease from year 3 to year 15?

ii) How would the actual average percentage decrease from year 3 to year 15 be calculated?

```
MTB > ##    Analysis of Time Decay    ##
MTB > Read 'decay.DAT' c1 c2.
     13 ROWS READ

 ROW    C1     C2

   1     3   16.8
   2     4   14.6
   3     5   11.9
   4     6   11.7
  .  .  .
MTB > name c1 't' c2 's' c3 'y' c4 'yfit' c5 'sfit'
MTB >
MTB > ##  determine y as ln(s)  ##
MTB > let c3 = loge(c2)
```

3 . (Continued)

```
MTB > ## correlations  ##
MTB > corr c1-c3

              t        s
s        -0.972
y        -0.995    0.982

MTB > regress c2 1 c1

The regression equation is
s = 18.0 - 1.03 t

Predictor       Coef       Stdev    t-ratio        p
Constant     18.0500      0.7333      24.61    0.000
t           -1.03462     0.07524     -13.75    0.000

s = 1.015       R-sq = 94.5%     R-sq(adj) = 94.0%

Analysis of Variance

SOURCE       DF          SS          MS         F        p
Regression    1      194.82      194.82    189.10    0.000
Error        11       11.33        1.03
Total        12      206.15

Unusual Observations
Obs.       t         s       Fit Stdev.Fit  Residual   St.Resid
  1      3.0    16.800    14.946     0.532     1.854      2.14R

R denotes an obs. with a large st. resid.

MTB > ##  transformed regression  ##
MTB > regress c3 1 c1 c20 c4

The regression equation is
y = 3.17 - 0.123 t

Predictor       Coef       Stdev    t-ratio        p
Constant     3.16577     0.03634      87.12    0.000
t          -0.122543    0.003728     -32.87    0.000

s = 0.05030     R-sq = 99.0%     R-sq(adj) = 98.9%

Analysis of Variance

SOURCE       DF          SS          MS         F        p
Regression    1      2.7331      2.7331   1080.37    0.000
Error        11      0.0278      0.0025
Total        12      2.7609

Unusual Observations
Obs.       t         y       Fit Stdev.Fit  Residual   St.Resid
  5      7.0    2.4069    2.3080    0.0158    0.0990      2.07R

R denotes an obs. with a large st. resid.

MTB > ##  transform fitted values  ##
MTB > let c5 = exp(c4)
MTB >
```

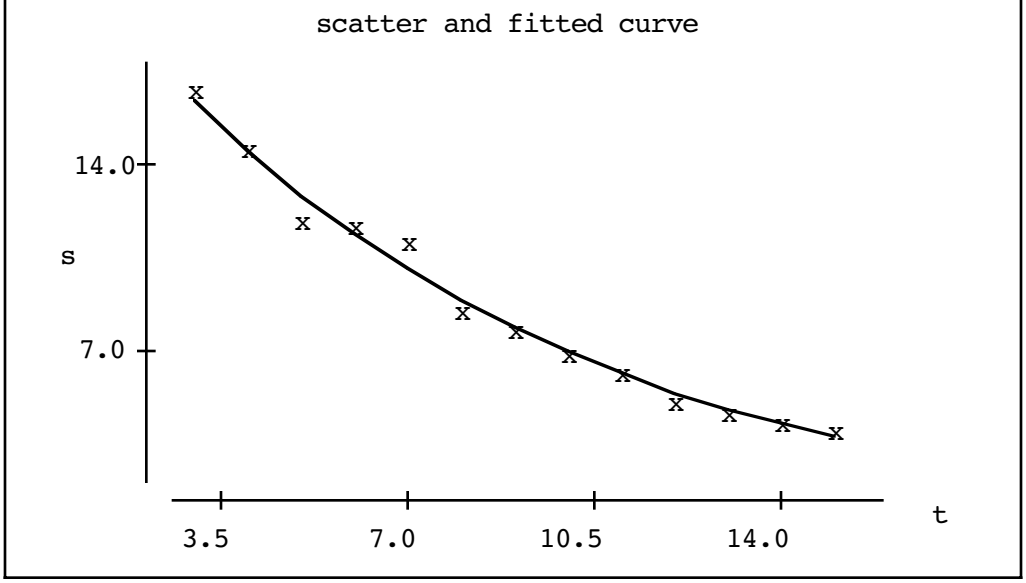
3 . (Continued)

```
MTB > print c1 - c5

 ROW     t      s         y      yfit      sfit

   1     3   16.8   2.82138   2.79814   16.4141
   2     4   14.6   2.68102   2.67560   14.5210
   3     5   11.9   2.47654   2.55305   12.8463
   4     6   11.7   2.45959   2.43051   11.3647
   5     7   11.1   2.40695   2.30797   10.0540
   6     8    8.5   2.14007   2.18542    8.8944
   7     9    7.8   2.05412   2.06288    7.8686
   8    10    6.9   1.93152   1.94034    6.9611
   9    11    6.2   1.82455   1.81779    6.1583
  10    12    5.1   1.62924   1.69525    5.4480
  11    13    4.7   1.54756   1.57271    4.8197
  12    14    4.3   1.45862   1.45017    4.2638
  13    15    4.0   1.38629   1.32762    3.7721

MTB >
MTB > gplot c2 c1;
SUBC> line c5 c1;
SUBC> title 'scatter and fitted curve'.
```



4 . a) It is of interest to estimate the percentage of a company's 3470 employee's who own shares in the company. There is a current belief that the percentage is less than 25%. A survey is to be conducted in such a way as to have a 90% chance of estimating the percentage to within 4 percentage points. How many employees should be in the survey?

b) Calculate a 95% confidence limit to determine at most what percentage own shares in the company if a survey of 300 employees included 57 who owned shares. Do these data provide evidence at the 5% level that the percentage is less than 25%?

c) If 694 of the 3470 employee's own shares and if 300 are surveyed, what is the probability that over 75 of those surveyed will own shares?

d) If one office has a staff of 26 of whom 6 own shares, and if 5 of the 26 are selected at random, what is the probability that 2 of them will own shares?

5 . Consider a study to investigate the effects of different structures for online courses to learn specific computer software applications with the following setup and results. Each of 15 test subjects followed two courses. The courses were for two different applications judged to be of equal difficulty. Approximately half the subjects followed the first course then, after a delay to reduce learning carry-over, followed the second; the remaining subjects followed the second course then, after a delay to reduce learning carry-over, the first. Within each of these arrangements approximately half the presentations of the first course were with a current online course structure and the other half were with a proposed new structure. The presentations were arranged so that each subject used each structure once. Each subject followed the course with online feedback until each step of the course had been completed successfully. An internal log kept a record of the times required to complete the course successfully. With the balancing of first and second courses and the use of each structure for approximately half of each arrangement, it was assumed that any differences in times could be attributed to differences in the course structures. A technician unsure about the appropriate method to compare the means entered the data into Minitab and analyzed the data with several methods that the technician thought could be appropriate. The Minitab session is shown below and on the next page.

```
MTB >  #  ANALYSIS OF ONLINE COURSE TIMES
MTB >
MTB > name c1 'subject' c2 'current' c3 'new'
MTB > set c1
DATA> 1:15
DATA> end
MTB > set c2
DATA> 49.3 56.5 42.8 67.1 35.7 44.0 38.6 30.7 59.8 45.6 33.6 78.3 46.1 55.9 59.8
DATA> end
MTB > set c3
DATA> 26.3 31.6 26.3 34.2 19.7 24.2 27.4 22.0 33.1 31.8 23.1 38.3 24.7 36.9 33.9
DATA> end
MTB > let c4 = c2-c3
MTB > name c4 'timediff'
MTB > print c1-c4  # DATA

 ROW  subject  current    new  timediff

   1        1     49.3   26.3      23.0
   2        2     56.5   31.6      24.9
   3        3     42.8   26.3      16.5
   4        4     67.1   34.2      32.9
   5        5     35.7   19.7      16.0
   6        6     44.0   24.2      19.8
   7        7     38.6   27.4      11.2
   8        8     30.7   22.0       8.7
   9        9     59.8   33.1      26.7
  10       10     45.6   31.8      13.8
  11       11     33.6   23.1      10.5
  12       12     78.3   38.3      40.0
  13       13     46.1   24.7      21.4
  14       14     55.9   36.9      19.0
  15       15     59.8   33.9      25.9

MTB > # COMPARE MEANS - OPTION 1a - INDEPENDENT SAMPLES
MTB > twosample 90 c2 c3;
SUBC> pooled.

TWOSAMPLE T FOR current VS new
          N      MEAN     STDEV   SE MEAN
current  15      49.6      13.3       3.4
new      15     28.90      5.73       1.5
```

5 . (Continued)

```
90 PCT CI FOR MU current - MU new: (14.3, 27.0)

TTEST MU current = MU new (VS NE): T= 5.55  P=0.0000  DF=  28

POOLED STDEV =       10.2

MTB > std c2 k1
   ST.DEV. =      13.252
MTB > std c3 k2
   ST.DEV. =      5.7332
MTB >
MTB > # VARIANCE RATIO
MTB > let k3 = k1*k1/(k2*k2)
MTB > print k3
K3       5.34286

MTB >
MTB > # COMPARE MEANS - OPTION 1b - INDEPENDENT SAMPLES
MTB > twosample 90 c2 c3

TWOSAMPLE T FOR current VS new
          N      MEAN     STDEV   SE MEAN
current  15      49.6      13.3       3.4
new      15     28.90      5.73       1.5

90 PCT CI FOR MU current - MU new: (14.2, 27.1)

TTEST MU current = MU new (VS NE): T= 5.55  P=0.0000  DF=  19

MTB > # COMPARE MEANS - OPTION 2a - PAIRED SAMPLES
MTB > tinterval 90 c4

             N      MEAN    STDEV  SE MEAN   90.0 PERCENT C.I.
timediff    15     20.69     8.61     2.22  (   16.77,   24.60)

MTB > # COMPARE MEANS - OPTION 2b - PAIRED SAMPLES
MTB > ttest 15 c4;
SUBC> alternative 1.

TEST OF MU = 15.00 VS MU G.T. 15.00

             N      MEAN    STDEV   SE MEAN        T    P VALUE
timediff    15     20.69     8.61      2.22     2.56      0.011

MTB >
```

a) As the statistician for the study, you are to use the appropriate Minitab analysis or otherwise to determine whether the new structure produces a 'substantial' improvement. Choose the appropriate Minitab analysis (stating why you have made that choice) or work with the 'raw' data to determine whether the data provide evidence at the 5% level of significance that the new structure reduces the learning time by over 15 minutes "on the average". Does the analysis provide a *P*-value? If so, what is it?

b) Compare the times for the current and new structures with a dot diagram.

c) What are the mean and standard deviation of the times for the new structure? Determine 95% confidence limits for the mean time for the new structure.

6 . a) Because of changing technological needs, an employer is concerned that employee ability to meet customer needs is deteriorating. At the start of a pilot survey, 73 employees who were still on staff after the survey included 54 who were able to meet customer needs. At the end of the survey, only 39 still were able to do so, and the other 15 were not. Of the 19 who originally were unable to meet customer needs, 7 were able to do so by the end of the survey and the other 12 still were not. Do these data indicate at the 5% level of significance that, in general, the proportion of employees able to meet customer needs decreased over the time period?

b) To counteract any potential decline in ability, training programs were introduced. Of 120 sample employees who participated in such a program, 84 showed a substantial improvement. Determine 95% confidence limits for the proportion of all employees who would show substantial improvement with such a program.

c) In another study employees had a choice to participate in an assisted training program or a self-study program. Some employees did not participate at all. Employees subsequently were assessed to determine whether performance improved, showed no change or deteriorated. The results for 125 employees surveyed at random were as tabulated below and analyzed with Minitab as shown. Do these data indicate at the 1% level of significance that assessment is related to program?

program	assessment		
	improved	no change	deteriorated
assisted	18	27	2
self-study	10	28	6
no participation	4	18	12

```
MTB > ##  Training Program Analysis  ##
MTB > Read c1-c3
DATA>  18  27   2
DATA>  10  28   6
DATA>   4  18  12
DATA> end
      3 ROWS READ
MTB >
MTB > Chisquare c1 - c3

Expected counts are printed below observed counts

            C1       C2       C3    Total
    1       18       27        2       47
         12.03    27.45     7.52

    2       10       28        6       44
         11.26    25.70     7.04

    3        4       18       12       34
          8.70    19.86     5.44

Total       32       73       20      125

ChiSq =  2.960 +  0.007 +  4.052 +
         0.142 +  0.207 +  0.154 +
         2.542 +  0.173 +  7.911 = 18.148
df = 4

MTB >
```

7. In an analysis of field equipment battery lifetimes under one set of environmental operating conditions, six batteries of each of four types (one type from group A and three types from group B) were monitored for useful life. The resulting lifetimes were stored in a data file and then analyzed with MINITAB as shown below.

```
MTB > name c1 'A' c2 'B1' c3 'B2' c4 'B3'
MTB > print c1-c4   #     DATA

 ROW      A     B1     B2     B3
   1    443    565    528    562
   2    459    556    542    580
   3    477    579    557    597
   4    482    600    566    617
   5    515    680    615    632
   6    542    718    623    695

MTB > stack c1-c4 c11;   #  STACK DATA FOR ANOVA
SUBC> subs c12.
MTB > name c11 'time' c12 'type'
MTB > let k1 = 0.05/6  #  ADJUSTED ALPHA FOR FISHER(BONFERRONI EQUIVALENT)
MTB > oneway c11 c12;  # ANOVA AND FISHER(AS BONFERRONI)
SUBC> fisher k1.

ANALYSIS OF VARIANCE ON time
SOURCE     DF        SS        MS        F        p
type        3     66326     22109     9.30    0.000
ERROR      20     47566      2378
TOTAL      23    113892
                                  INDIVIDUAL 95 PCT CI'S FOR MEAN
                                  BASED ON POOLED STDEV
 LEVEL      N      MEAN     STDEV  ------+---------+---------+---------+
     1      6    486.33     36.47  (------*------)
     2      6    616.33     66.82                       (------*------)
     3      6    571.83     38.85               (------*------)
     4      6    613.83     47.00                     (------*------)
                                  ------+---------+---------+---------+
POOLED STDEV =    48.77              480       540       600       660

Fisher's pairwise comparisons
    Family error rate = 0.0384
Individual error rate = 0.00833
Critical value = 2.927

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

              1         2         3

     2    -212.4
           -47.6

     3    -167.9     -37.9
            -3.1     126.9

     4    -209.9     -79.9    -124.4
           -45.1      84.9      40.4
```

a) Combine the data for all group B data and display these data with a stem-and-leaf display with stem labels that increase in steps of 50.

a) Do these results indicate that the mean battery lifetimes are not all the same for the four types of batteries? What is the *P*-value?

c) Which battery types differ with regard to mean battery life? What overall α was used?

d) What are the mean and standard deviation for the type A batteries?

e) Using the type A data only, determine 95% confidence limits for the mean lifetime.

f) What distribution assumption is made in all of the above analyses?