THE UNIVERSITY OF MINNESOTA

Statistics 5401/8401 October 18, 2005

Sample Midterm Examination

There were three questions, each with more than one part. The number of points is next to each part.

On separate sheets * were "exhibits" presenting MacAnova output needed to answer the questions. They also included short tables useful with Bonferronized F**.

Methods of analysis in different parts of a question must be consistent.

If not otherwise stated, use α = .05 on all hypothesis tests. Use "small sample" tests where possible. Explicitly give the critical value that should be used, even when it ought to be obvious that the observed value is significant or non-significant, as a z = 5. If a value is not available in the tables in the book, describe what is needed as specifically as possible , for example, upper 5% point of F on 17 and 121 degrees of freedom).re.

* Exhibits are interspersed with questions, and space for answers has been removed.
**Bonferronized F values are at end.

1. In **Exhibit 1** are analyses on 5 measurements on the teeth of Golden retriever dogs from breeders in 4 countries, England, France, United States and Canada. **Tables 1a**, **1b**, and **1c** on p. 5, 6 and 7 of the exhibit booklet are tables of Bonferronized F probability points, that is $F_{f_1,f_2}(\alpha/k)$, for Bonferronizing factor k = 1, 2, 3, 4, 5, 6, denominator degrees of freedom f_2 = 26, 27, ..., 35, and numerator degrees of freedom f_1 =3, 4 and 5, respectively.

Exhibit 1 (for problem 1)

```
Cmd> dogteeth <- read("","dogteeth") # read from file
Dogteeth         35     6 labels
) Tooth measurements on Golden retrievers (dogs) from kennels in
) England, France, United States, and Canada
) Col.  1: Country, 1=England, 2 = France, 3 = USA, 4=Canada
) Col.  2: X1=length of mandible
) Col.  3: X2=breadth of mandible below 1st molar
) Col.  4: X3=breadth of articular condyle
) Col.  5: X4=height of mandlible below 1st molar
) Col.  6: X5=length of 1st molar
Read from file "TP1:Stat5401:Exams:DogData.txt"

Cmd> country <- factor(dogteeth[,1]); list(country)
country         REAL   35    1     FACTOR with 4 levels (labels)

Cmd> y <- dogteeth[,-1] # 35 by 5 data matrix without factor

Cmd> means <- tabs(y,country,means:T); means # rows are group means
(1,1)       127.2      10.138      21.087       21.85      20.225
(2,1)      120.65        9.57       18.34       21.15        19.3
(3,1)      126.68      9.9444      19.956      21.533      19.933
(4,1)      131.86      10.612      22.387      22.675      20.838
              X1          X2          X3          X4          X5

Cmd> n <- tabs(,country); n # sample sizes from the 4 countries
(1)          8          10           9           8
```

```
Cmd> manova("y=country")
Model used is y=country
WARNING: summaries are sequential
NOTE: SS/SP matrices suppressed because of size; use 'manova(,sssp:T)'
                          SS and SP Matrices
                DF
CONSTANT         1
                    Type 'SS[1,,]' to see SS/SP matrix
country          3
                    Type 'SS[2,,]' to see SS/SP matrix
ERROR1          31
                    Type 'SS[3,,]' to see SS/SP matrix

Cmd> h <- matrix(SS[2,,])

Cmd> fh <- DF[2]

Cmd> e <- matrix(SS[3,,])

Cmd> fe <- DF[3]

Cmd> p <- ncols(y)
Cmd> h # hypothesis matrix
               MandLnth   MandBrdth CndyleBrdth      MandHt   MolarLnth
MandLnth         574.46        52.4      208.14      75.056      78.248
MandBrdth          52.4      4.9881      19.685      7.3223      7.3637
CndyleBrdth      208.14      19.685      79.304      28.507      29.323
MandHt           75.056      7.3223      28.507      10.946      10.701
MolarLnth        78.248      7.3637      29.323      10.701      10.933

Cmd> e # error matrix
               MandLnth   MandBrdth CndyleBrdth      MandHt   MolarLnth
MandLnth         1686.5      108.29      438.69      234.21      70.768
MandBrdth        108.29      13.031      36.175      17.569      4.0754
CndyleBrdth      438.69      36.175      202.94      77.266       11.66
MandHt           234.21      17.569      77.266        61.3      1.0675
MolarLnth        70.768      4.0754       11.66      1.0675      20.994

Cmd> s <- e/fe ; s # pooled variance matrix
               MandLnth   MandBrdth CndyleBrdth      MandHt   MolarLnth
MandLnth         54.404      3.4931      14.151      7.5553      2.2828
MandBrdth        3.4931     0.42035      1.1669     0.56675     0.13147
CndyleBrdth      14.151      1.1669      6.5466      2.4924     0.37612
MandHt           7.5553     0.56675      2.4924      1.9774    0.034435
MolarLnth        2.2828     0.13147     0.37612    0.034435     0.67722

Cmd> # Compute T^2 for all pairs of means

Cmd> names <- vector("England","France","USA","Canada")

Cmd> tsq <- matrix(dmat(4,0),labels:structure(names,names)) #EMPTY

Cmd> for(i,run(4)){for(j,run(4)){ # Fill matrix tsq
        dij <- vector(means[i,]-means[j,]) # difference of mean vectors
        vhatij <- (1/n[i]+1/n[j])*s
        tsq[i,j] <- dij' %*% solve(vhatij) %*% dij};;}

Cmd> tsq
            England      France         USA      Canada
England           0      9.4974      2.7166      4.0353
France       9.4974           0      4.9819      24.137
USA          2.7166      4.9819           0      11.069
Canada       4.0353      24.137      11.069           0

Cmd> releigenvals(h,e) # relative eigenvalues
(1)     0.82745    0.069417    0.023853  2.5498e-16  -3.432e-16
```

(a) (15) Use Bonferronized F-tests to test the null hypothesis that the expected tooth measurements are the same in all four Countries. State the null and alternative hypotheses using μ_1, μ_2, μ_3, and μ_4 as notation for the four 5 dimensional mean vectors. Hypothesis and error matrices on on p. 2 of the exhibit booklet.

(b) (15) Use Bonferronized Hotelling's T^2 to test the same hypothesis as in (a). Values of T^2 are computed on p. 2 of the Exhibit booklet.

(c) (15) Use a test involving the eigenvalues of **H** relative to **E** to test the same null hypothesis as in (a). Relative eigenvalues are on p. 2 of the exhibit booklet.

2. $\boldsymbol{X}$ is a 50 by 4 data matrix whose rows are a random sample from a population with mean $\boldsymbol{\mu}=[\mu_1,\mu_2,\mu_3,\mu_4]'$ and variance matrix $\boldsymbol{\Sigma} = [\sigma_{ij}]$. The sample variance matrix is $\boldsymbol{S} = [s_{ij}]$ and sample mean is $\bar{\mathbf{x}}=[\bar{x}_1,\bar{x}_2,\bar{x}_3,\bar{x}_4]'$.

Give the dimensions of each of the following matrices and describe what they represent statistically (for example, "sample regression coefficients in the regression of the last column of $\boldsymbol{X}$ on the first 3 columns" or "variance of $\bar{x}_3$"). No justification is necessary.

(a) (10) $$(1/50)\mathbf{1}_{50}'\boldsymbol{X}$$

(b) (10) $$\frac{1}{50}\begin{bmatrix}1 & 1 & -1 & -1\end{bmatrix}\mathbf{S}\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}$$

(c) (10) $$\begin{bmatrix}1/\sqrt{\sigma_{11}} & 0 & 0 & 0\\ 0 & 1/\sqrt{\sigma_{22}} & 0 & 0\\ 0 & 0 & 1/\sqrt{\sigma_{33}} & 0\\ 0 & 0 & 0 & 1/\sqrt{\sigma_{44}}\end{bmatrix}\Sigma\begin{bmatrix}1/\sqrt{\sigma_{11}} & 0 & 0 & 0\\ 0 & 1/\sqrt{\sigma_{22}} & 0 & 0\\ 0 & 0 & 1/\sqrt{\sigma_{33}} & 0\\ 0 & 0 & 0 & 1/\sqrt{\sigma_{44}}\end{bmatrix}$$

3. This problem deals with the analysis of data in Table 4.3 in Johnson and Wichern. The stiffness of 30 boards was measured in four different ways, two dynamic testing methods (sending a shock wave down the board and vibrating a board), and using two static methods. **Exhibit 3** contains MacAnova output related to analysis of these data.

Exhibit 3 (for problem 3)

```
Cmd> data <- read("","t04_03")
T04_03    30    4 format
) Data from Table 4.3 p. 187 in
) Applied Mulivariate Statistical Analysis, 5th Edition
) by Richard A. Johnson and Dean W. Wichern, Prentice Hall, 2002
) These data were edited from file T4-3.DAT on disk from book
) Omitted was the last column (d^2) as this can be computed directly
) using distcomp(T04_03)
) Four measurements of stiffness
) Col. 1: x1 (from shock wave down board)
) Col. 2: x2 (from vibrating board)
) Col. 3: x3 (from static test)
) Col. 4: x4 (from static test)
Read from file "TP1:Stat5401:Data:JWData5.txt"

Cmd> n <- nrows(data) # sample size
```

```
Cmd> stats <- tabs(data,covar:T,mean:T);stats
component: mean
(1)      1906.1      1749.5      1509.1        1725
component: covar
(1,1)  1.0562e+05       94614       87290       94231
(2,1)       94614  1.0151e+05       76137       81064
(3,1)       87290       76137       91917       90352
(4,1)       94231       81064       90352  1.0423e+05

Cmd> xbar <- stats$mean; s <- stats$covar

Cmd> cor(data) #correlation matrix
(1,1)           1     0.91376     0.88593     0.89812
(2,1)     0.91376           1     0.78821      0.7881
(3,1)     0.88593     0.78821           1      0.9231
(4,1)     0.89812      0.7881      0.9231           1

Cmd> eigen(s)
component: values
(1)  3.6322e+05       26814      7688.5      5550.9
component: vectors
(1,1)     0.52638    -0.19881    -0.23971     0.79116
(2,1)     0.48659    -0.72687     0.13627    -0.46511
(3,1)     0.47569     0.44462     0.75856    0.025065
(4,1)     0.50977     0.48421    -0.59039    -0.39637

Cmd> c1 # previously entered 3 by 4 matrix
(1,1)           1           1          -1          -1
(2,1)           1          -1           0           0
(3,1)           0           0           1          -1

Cmd> c1xbar <- c1 %*% xbar; c1xbar'
(1,1)      421.53      156.57     -215.83

Cmd> c1sc1 <- c1 %*% s %*% c1'; c1sc1
(1,1)       95759      -20213       442.6
(2,1)      -20213       17899     -2013.8
(3,1)       442.6     -2013.8       15440

Cmd> c1xbar' %*% solve(c1sc1/n) %*% c1xbar
(1,1)      254.72

Cmd> sqrt(diag(c1sc1/n))
(1)      56.498      24.426      22.686

Cmd> c2 # previously entered matrix
(1,1)           1          -1           0           0
(2,1)           1           0          -1           0
(3,1)           1           0           0          -1
(4,1)           0           1          -1           0
(5,1)           0           1           0          -1
(6,1)           0           0           1          -1

Cmd> c2xbar <- vector(c2 %*% xbar)

Cmd> c2sc2 <- c2 %*% s %*% c2'

Cmd> vhat2 <- c2sc2/n
```

```
Cmd> print(format:"10.4f",c2xbar:c2xbar',format:"10.3f",c2sc2,vhat2)
c2xbar:
(1,1)   156.5667   396.9667   181.1333   240.4000    24.5667  -215.8333
c2sc2:
(1,1)  17899.357   -149.843  -2163.595 -18049.200 -20062.953  -2013.753
(2,1)   -149.843  22953.964  14448.246  23103.807  14598.089  -8505.718
(3,1)  -2163.595  14448.246  21382.809  16611.841  23546.405   6934.563
(4,1) -18049.200  23103.807  16611.841  41153.007  34661.041  -6491.966
(5,1) -20062.953  14598.089  23546.405  34661.041  43609.357   8948.316
(6,1)  -2013.753  -8505.718   6934.563  -6491.966   8948.316  15440.282
vhat2:
(1,1)    596.645     -4.995    -72.120   -601.640   -668.765    -67.125
(2,1)     -4.995    765.132    481.608    770.127    486.603   -283.524
(3,1)    -72.120    481.608    712.760    553.728    784.880    231.152
(4,1)   -601.640    770.127    553.728   1371.767   1155.368   -216.399
(5,1)   -668.765    486.603    784.880   1155.368   1453.645    298.277
(6,1)    -67.125   -283.524    231.152   -216.399    298.277    514.676
```

There is interest in comparing static with dynamic testing methods, comparing the dynamic methods and comparing the static methods.

(a) (15) Describe in words what the 3 elements of variable `c1xbar` on p. 3 of the exhibits are and why they might be relevent things to compute. Is there statistical evidence that expectations of the elements of `c1xbar` are non-zero?

(b) (15) Find a 95% confidence interval for the difference between the mean measurement using the shock-wave measurement (variable 1) and the mean measurement using the first static test (variable 3). (Hint: Look at row 3 of matrix `c2` on p. 3 of exhibits.) Do it using a method that would be appropriate for simultaneous confidence intervals of all comparisons of two measurement methods. Use the shortest limits that would be appropriate.

Table 1a

Table of 5% F-probability points, all for $f_1 = 3$ numerator degrees of freedom, Bonferronized for k tests

Denom	Number of tests k					
D.F. f_2	1	2	3	4	5	6
26	2.9752	3.6697	4.0903	4.3956	4.6366	4.8362
27	2.9604	3.6472	4.0624	4.3635	4.6009	4.7975
28	2.9467	3.6264	4.0367	4.3339	4.5681	4.7619
29	2.9340	3.6072	4.0129	4.3066	4.5378	4.7290
30	2.9223	3.5894	3.9909	4.2812	4.5097	4.6986
31	2.9113	3.5728	3.9704	4.2577	**4.4837**	4.6704
32	2.9011	3.5573	3.9513	4.2358	4.4594	4.6441
33	2.8916	3.5429	3.9335	4.2153	4.4368	4.6196
34	2.8826	3.5293	3.9168	4.1962	4.4156	4.5967
35	2.8742	3.5166	3.9011	4.1782	4.3957	4.5752
36	2.8663	3.5047	3.8864	4.1614	4.3771	4.5550
37	2.8588	3.4934	3.8726	4.1455	4.3595	4.5360
38	2.8517	3.4828	3.8595	4.1305	4.3430	4.5181
39	2.8451	3.4728	3.8472	4.1164	4.3274	4.5012
40	2.8387	3.4633	3.8355	4.1030	4.3126	4.4852
41	2.8327	3.4542	3.8244	4.0903	4.2986	4.4701
42	2.8270	3.4457	3.8139	4.0783	4.2853	4.4557
43	2.8216	3.4376	3.8039	4.0668	4.2726	4.4420
44	2.8165	3.4298	3.7944	4.0560	4.2606	4.4291
45	2.8115	3.4224	3.7853	4.0456	4.2492	4.4167

Table 1b
Table of 5% F-probability points, all for f_1 = 4 numerator degrees of freedom, Bonferronized for k tests

Denom	Number of tests k					
D.F. f_2	1	2	3	4	5	6
26	2.7426	3.3289	3.6823	3.9383	4.1400	4.3069
27	2.7278	3.3067	3.6551	3.9072	4.1056	4.2698
28	2.7141	3.2863	3.6301	3.8786	4.0740	4.2356
29	2.7014	3.2674	3.6070	3.8521	4.0449	4.2041
30	2.6896	3.2499	3.5855	3.8276	4.0179	4.1750
31	2.6787	3.2336	3.5656	3.8049	3.9928	4.1479
32	2.6684	3.2185	3.5471	3.7837	3.9695	4.1227
33	2.6589	3.2043	3.5297	3.7640	3.9477	4.0992
34	2.6499	3.1910	3.5135	3.7455	3.9273	4.0772
35	2.6415	3.1785	3.4983	3.7281	3.9082	4.0566
36	2.6335	3.1668	3.4840	3.7118	3.8903	4.0373
37	2.6261	3.1557	3.4705	3.6965	3.8734	4.0191
38	2.6190	3.1453	3.4578	3.6820	3.8575	4.0020
39	2.6123	3.1354	3.4458	3.6684	3.8425	3.9858
40	2.6060	3.1261	3.4345	3.6555	3.8283	3.9705
41	2.6000	3.1173	3.4237	3.6432	3.8148	3.9560
42	2.5943	3.1089	3.4135	3.6316	3.8021	3.9422
43	2.5888	3.1009	3.4038	3.6206	3.7899	3.9292
44	2.5837	3.0933	3.3945	3.6101	3.7784	3.9168
45	2.5787	3.0860	3.3857	3.6001	3.7674	3.9049

Table 1c
Table of 5% F-probability points, all for f_1 = 5 numerator degrees of freedom,
Bonferronized for k tests

Denom	Number of tests k					
D.F. f_2	1	2	3	4	5	6
26	2.5868	3.1048	3.4160	3.6411	3.8183	3.9649
27	2.5719	3.0828	3.3893	3.6106	3.7848	3.9287
28	2.5581	3.0626	3.3646	3.5826	3.7539	3.8954
29	2.5454	3.0438	3.3418	3.5567	3.7254	3.8647
30	2.5336	3.0265	3.3207	3.5327	3.6990	3.8363
31	2.5225	3.0103	3.3011	3.5104	3.6745	3.8099
32	2.5123	2.9953	3.2829	3.4896	3.6517	3.7853
33	2.5026	2.9812	3.2658	3.4703	3.6305	3.7625
34	2.4936	2.9680	3.2498	3.4521	3.6106	3.7410
35	2.4851	2.9557	3.2348	3.4351	3.5919	3.7210
36	2.4772	2.9440	3.2208	3.4192	3.5744	3.7021
37	2.4696	2.9331	3.2075	3.4041	3.5579	3.6844
38	2.4625	2.9227	3.1950	3.3900	3.5424	3.6677
39	2.4558	2.9130	3.1832	3.3766	3.5277	3.6520
40	2.4495	2.9037	3.1720	3.3639	3.5138	3.6370
41	2.4434	2.8950	3.1614	3.3519	3.5007	3.6229
42	2.4377	2.8866	3.1514	3.3406	3.4882	3.6095
43	2.4322	2.8787	3.1418	3.3298	3.4764	3.5968
44	2.4270	2.8712	3.1327	3.3195	3.4651	3.5847
45	2.4221	2.8640	3.1241	3.3097	3.4544	3.5732