30

HUDM 6122 Name: KEY

Test 1

6p

1. Use the matrices:

$$A = \begin{pmatrix} 4 & 8 & 8 \\ 3 & 6 & -9 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & 8 & 6 \\ 4 & -3 & 5 \end{pmatrix}.$$

1p

(a) Calculate $\boldsymbol{A} - \boldsymbol{B}$.

$$A - B = \begin{pmatrix} 1 & 0 & 2 \\ -1 & 9 & -14 \end{pmatrix}$$

1p

(b) Find $\boldsymbol{A}'\boldsymbol{A}$.

$$A'A = \begin{pmatrix} 4 & 3 \\ 8 & 6 \\ 8 & -9 \end{pmatrix} \begin{pmatrix} 4 & 8 & 8 \\ 3 & 6 & -9 \end{pmatrix} = \begin{pmatrix} 25 & 50 & 5 \\ 50 & 100 & 10 \\ 5 & 10 & 145 \end{pmatrix}$$

−0.5 for arithmetic mistakes

1p

(c) Is $\boldsymbol{A}$ a square matrix?

No, # rows = 2 ≠ 3 = # columns

1p

(d) Let $\boldsymbol{x} = (1 \ -1 \ 0)'$. Find $\boldsymbol{Bx}$.

$$Bx = \begin{pmatrix} 3 & 8 & 6 \\ 4 & -3 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -5 \\ 7 \end{pmatrix}$$

2p

(e) Find the determinant $|\boldsymbol{BB}'|$.

$$BB' = \begin{pmatrix} 3 & 8 & 6 \\ 4 & -3 & 5 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 8 & -3 \\ 6 & 5 \end{pmatrix} = \begin{pmatrix} 109 & 18 \\ 18 & 50 \end{pmatrix}$$

$$\det(BB') = 109(50) - 18^2 = \boxed{5126}$$

7p

2. A management consultant was engaged by a company to analyze the cost effectiveness of its travel expenses. The consultant selected 10 managers from each of the "Sales" and "Research and development" divisions. He collected data on the dollar costs of domestic and international travel made by the managers during the past month. The data along with other R code is presented below (first 10 rows represent the "Sales" division and second 10 rows the "R&D" division).

```
> x <- read.table("e:/Work/HUDM6122/datasets/E1.txt", header = TRUE)
> x
   Domestic International
1       666           705
2       920          1040
3       495           502
4       602           803
5      1499          1526
6       960          1982
7       796           824
8       343           428
9       894           901
10      813           925
11      391           251
12      450           351
13      609           729
14      910           820
15      705           620
16      472           301
17      645           692
18      496           301
19      763           729
20     1309          1822
> Y1 <- as.matrix(x[1:10,])
> Y2 <- as.matrix(x[11:20,])
> y1.bar <- apply(Y1, 2, mean)
> y1.bar
   Domestic International
      798.8         963.6
> y2.bar <- apply(Y2, 2, mean)
> y2.bar
   Domestic International
      675.0         661.6
> S1 <- cov(Y1); S2 <- cov(Y2)
> n1 <- nrow(Y1); n2 <- nrow(Y2) # This calculates the number of rows
> Sp <- (1/(n1 + n2 -2))*((n1 - 1)*S1 + (n2 - 1)*S2)
>  T.sq.obs  <-  (n1*n2/(n1+n2))*t(y1.bar-y2.bar)%*%solve(Sp)%*%(y1.bar-
y2.bar)
> T.sq.obs
         [,1]
[1,] 2.449405
> pf(((n1+n2-2-1)/((n1+n2-2)*2))*T.sq.obs, 2, n1+n2-2-1, lower.tail =
FALSE)
          [,1]
[1,] 0.3380886
> pf(((n1+n2-2+1)/((n1+n2-2)*2))*T.sq.obs, 2, n1+n2-2+1, lower.tail =
FALSE)
          [,1]
[1,] 0.2975927
```

1p

a) Suppose that the consultant wants to test the claim that the average travel costs are different for the two divisions. State the two hypotheses and explain any notation.

$H_0: \mu_{Sales} = \mu_{R\&D}$ vs. $H_1: \mu_{Sales} \neq \mu_{R\&D}$

where $\mu = \begin{pmatrix} \mu_{domestic} \\ \mu_{international} \end{pmatrix}$

1p

b) Report the value of the test statistic.

$T^2 = 2.45$

1p

c) What is the distribution of the test statistic under H_0?

$T^2 \sim T^2_{2, 18}$ 0.5 points

1p

d) Write down the rejection rule for the test in part a). Use $\alpha = 0.05$ and table value.

Reject H_0 if $T^2 > 7.606$

1p

e) State the conclusion using the specific context of the problem.

Since $2.45 \not> 7.606$ we don't reject H_0
Average travel expenses are not different between "sales" & "R&D" division

2p

f) TRUE or FALSE. If we use the same α level and perform two univariate t-tests (one for domestic and one for international travel) to compare the two divisions, then the conclusions are guaranteed to be the same as these from the multivariate test. Explain why.

Univariate tests do not take into account the correlation between μ_{Sales} & $\mu_{R\&D}$

11 8

3. A study was performed on the readability levels of magazine advertisements. Magazines were classified as "High education", "Medium Education" and "Low Education", depending on the education level of the majority of their readers. (For example, "Scientific American" was classified as "High Education", "Sports Illustrated" as "Medium Education", and "National Enquirer" as "Low Education".) A random sample of 18 advertisements was selected from each of the three types of magazines and the number of words per ad, words per sentence, and number of words with at least 3 syllables was recorded for each ad. Partial data and R code are provided below.

```
> x <- read.table("e:/Work/Columbia/HUDM6122/datasets/Readability.txt", header = TRUE); x
               group words.ad words.sentence words.w.3.syl
1     high.education      205              9            34
...
18    high.education       88             12             6
19  medium.education      191             25            13
...
36  medium.education       67              7             8
37     low.education      162             14            16
...
54     low.education      208             20            15
> y.. <- mean(x[, 2:4])
> y1. <- mean(x[x[,1]=="high.education",  2:4])
> y2. <- mean(x[x[,1]=="medium.education", 2:4])
> y3. <- mean(x[x[,1]=="low.education", 2:4])
> H <- 18*((y1.-y..)%*%t(y1.-y..)+(y2.-y..)%*%t(y2.-y..)+(y3.-y..)%*%t(y3.-y..))
> e1 <- e2 <- e3 <- matrix(rep(0,9), 3, 3)
> index <- 1:nrow(x)
> for (j in index[x[,1]=="high.education"])
  e1 <- e1+t(as.matrix((x[j, 2:4]-y1.)))%*%as.matrix(x[j, 2:4]-y1.)
> for (j in index[x[,1]=="medium.education"])
  e2 <- e2+t(as.matrix((x[j, 2:4]-y2.)))%*%as.matrix(x[j, 2:4]-y2.)
> for (j in index[x[,1]=="low.education"])
  e3 <- e3+t(as.matrix((x[j, 2:4]-y3.)))%*%as.matrix(x[j, 2:4]-y3.)
> E <- e1 + e2 + e3
> round(eigen(solve(E)%*%H)$value,3)
[1] 0.166 0.073 0.000
```

1p

(a) Is there enough evidence that reading difficulty of magazine advertisements varied across the three groups? Specify the appropriate hypotheses to answer the above question. Define any parameter(s) used in the hypotheses.

$H_0: \mu_{HE} = \mu_{ME} = \mu_{LE}$ vs. H_1: At least two means are different

$$\mu = \begin{pmatrix} \mu_{w/A} \\ \mu_{w/S} \\ \mu_{w3} \end{pmatrix}$$

2 p

(b) What are the values of p, k, n, ν_E, ν_H, and s?

4 p $p=3$, $k=3$, $n=18$, $\nu_E = 51$, $\nu_H = 2$, $s=2$

(c) Calculate all four test statistics. (Show your work)

$$\Lambda = \left(\frac{1}{1+0.166}\right)\left(\frac{1}{1+0.073}\right) = 0.799$$

$$\theta = \frac{0.166}{1+0.166} = 0.142$$

$$V^{(s)} = 0.142 + \frac{0.073}{0.073} = 0.21$$

$$U^{(s)} = 0.166 + 0.073 = 0.239$$

1 p

(d) Use Wilks' test, the appropriate table value and $\alpha = 0.05$ to draw a conclusion in the context of the problem.

$\Lambda_{3,2,51} \approx \Lambda_{3,2,60} = 0.808$ (or 0.724 or 0.766)

Since $0.799 < 0.808$ we reject H_0

There are differences between the 3 groups of magazines' ads

1 p

(e) Calculate A_Λ measure of association.

$$A_\Lambda = 1-\sqrt{\Lambda} = 1-\sqrt{0.799} = 0.106$$

3 p

(f) Specify the assumptions of the MANOVA model for the above problem.

1. Each group has multivariate normal distr.
2. Cov. matrices are the same
3. Groups are indep.

6p

4. Let $y = (y_1, y_2)'$ be a random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, given by

$$\mu = \begin{pmatrix} 36 \\ 26 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 65 & 34 \\ 34 & 46 \end{pmatrix}.$$

2p

(a) Find the correlation coefficient between y_1 and y_2.

$$\text{cor}(y_1, y_2) = \frac{34}{\sqrt{65}\sqrt{46}} = 0.62$$

1p

(b) Find the total variance of $\boldsymbol{y}$.

$$\text{Total var. } \cancel{of}\ y = \text{tr}(\Sigma) = 65+46 = 111$$

2p

(c) Let $\boldsymbol{a} = (1\ {-1})'$. Find the mean and the covariance of $\boldsymbol{a}'\boldsymbol{y}$

$$E(a'y) = a'\mu = (1\ \ {-1}) \begin{pmatrix} 36 \\ 26 \end{pmatrix} = \boxed{10}$$

$$\text{cov}(a'y) = a'\Sigma a = (1\ {-1}) \begin{pmatrix} 65 & 34 \\ 34 & 46 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$= (31\ \ {-12}) \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \boxed{43}$$

1p

(d) In addition now assume $\boldsymbol{y} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Specify the distributions of y_1 and y_2

$$y_1 \sim N(36, 65)$$

$$y_2 \sim N(26, 46)$$

Table A.7 Upper Percentage Points of Hotelling's T^2 Distribution

Degrees of Freedom, ν	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
					$\alpha = .05$					
2	18.513									
3	10.128	57.000								
4	7.709	25.472	114.986							
5	6.608	17.361	46.383	192.468						
6	5.987	13.887	29.661	72.937	289.446					
7	5.591	12.001	22.720	44.718	105.157	405.920				
8	5.318	10.828	19.028	33.230	62.561	143.050	541.890			
9	5.117	10.033	16.766	27.202	45.453	83.202	186.622	697.356		
10	4.965	9.459	15.248	23.545	36.561	59.403	106.649	235.873	872.317	
11	4.844	9.026	14.163	21.108	31.205	47.123	75.088	132.903	290.806	1066.774
12	4.747	8.689	13.350	19.376	27.656	39.764	58.893	92.512	161.967	351.421
13	4.667	8.418	12.719	18.086	25.145	34.911	49.232	71.878	111.676	193.842
14	4.600	8.197	12.216	17.089	23.281	31.488	42.881	59.612	86.079	132.582
15	4.543	8.012	11.806	16.296	21.845	28.955	38.415	51.572	70.907	101.499
16	4.494	7.856	11.465	15.651	20.706	27.008	35.117	45.932	60.986	83.121
17	4.451	7.722	11.177	15.117	19.782	25.467	32.588	41.775	54.041	71.127
18	4.414	7.606	10.931	14.667	19.017	24.219	30.590	38.592	48.930	62.746
19	4.381	7.504	10.719	14.283	18.375	23.189	28.975	36.082	45.023	56.587
20	4.351	7.415	10.533	13.952	17.828	22.324	27.642	34.054	41.946	51.884
21	4.325	7.335	10.370	13.663	17.356	21.588	26.525	32.384	39.463	48.184
22	4.301	7.264	10.225	13.409	16.945	20.954	25.576	30.985	37.419	45.202
23	4.279	7.200	10.095	13.184	16.585	20.403	24.759	29.798	35.709	42.750
24	4.260	7.142	9.979	12.983	16.265	19.920	24.049	28.777	34.258	40.699
25	4.242	7.089	9.874	12.803	15.981	19.492	23.427	27.891	33.013	38.961
26	4.225	7.041	9.779	12.641	15.726	19.112	22.878	27.114	31.932	37.469

(continued)

Table A.9 (*Continued*)

ν_E	ν_H 1	2	3	4	5	6	7	8	9	10	11	12
						$p = 3$						
1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
2	.000	.000	.000	.000	.000	.001^a	.002^a	.004^a	.005^a	.008^a	.010^a	.013^a
3	1.70^a	.354^a	.179^a	.127^a	.105^a	.095^a	.091^a	.090^a	.091^a	.092^a	.095^a	.098^a
4	.034	.010	.004	.002	.001	.001	.809^a	.659^a	.562^a	.496^a	.449^a	.416^a
5	.097	.036	.018	.010	6.36^a	4.37^a	3.20^a	2.46^a	1.97^a	1.64^a	1.40^a	1.22^a
6	.168	.074	.040	.024	.016	.011	.008	.006	.004	3.94^a	3.28^a	2.79^a
7	.236	.116	.068	.043	.029	.021	.016	.012	9.49^a	7.67^a	6.35^a	5.35^a
8	.296	.160	.099	.066	.046	.034	.026	.020	.016	.013	.011	9.00^a
9	.349	.203	.131	.091	.066	.049	.038	.030	.024	.020	.016	.014
10	.396	.243	.164	.117	.086	.066	.052	.041	.034	.028	.023	.020
11	.437	.281	.196	.143	.108	.084	.067	.054	.044	.037	.031	.026
12	.473	.316	.226	.169	.130	.103	.083	.067	.056	.047	.040	.034
13	.505	.348	.255	.194	.152	.122	.099	.082	.068	.058	.049	.042
14	.534	.378	.283	.219	.174	.141	.116	.096	.081	.069	.059	.051
15	.560	.405	.309	.243	.195	.160	.133	.111	.095	.081	.070	.061
16	.583	.431	.334	.266	.216	.179	.149	.127	.108	.093	.081	.071
17	.603	.454	.357	.288	.236	.197	.166	.142	.122	.106	.092	.081
18	.622	.476	.379	.309	.256	.215	.183	.157	.136	.118	.104	.092
19	.639	.496	.399	.329	.275	.233	.199	.172	.149	.131	.115	.102
20	.655	.515	.419	.348	.293	.250	.215	.187	.163	.144	.127	.113
21	.669	.532	.437	.366	.310	.266	.230	.201	.177	.156	.139	.124
22	.683	.548	.454	.383	.327	.282	.246	.215	.190	.169	.150	.135
23	.695	.564	.470	.399	.343	.298	.260	.229	.203	.181	.162	.146
24	.706	.578	.486	.415	.359	.313	.275	.243	.216	.193	.173	.156
25	.717	.591	.500	.430	.374	.327	.289	.256	.229	.205	.185	.167
26	.727	.604	.514	.444	.388	.341	.302	.269	.241	.217	.196	.178
27	.736	.616	.527	.458	.401	.355	.315	.282	.253	.229	.207	.188
28	.744	.627	.540	.471	.415	.368	.328	.294	.265	.240	.218	.199
29	.752	.638	.552	.483	.427	.380	.340	.306	.277	.251	.229	.209
30	.760	.648	.563	.495	.439	.392	.352	.318	.288	.262	.239	.219
40	.816	.724	.651	.591	.539	.494	.454	.419	.387	.359	.334	.311
60	.875	.808	.752	.704	.661	.623	.587	.555	.526	.498	.473	.449
80	.905	.853	.808	.769	.733	.700	.670	.641	.615	.590	.566	.544
100	.924	.881	.844	.810	.780	.751	.725	.700	.676	.654	.632	.612
120	.936	.900	.868	.839	.813	.788	.764	.742	.721	.700	.681	.663
140	.945	.913	.886	.861	.837	.815	.794	.774	.755	.736	.719	.702
170	.955	.928	.905	.884	.864	.845	.827	.809	.792	.776	.761	.746
200	.961	.939	.919	.900	.883	.866	.850	.835	.820	.806	.792	.779
240	.968	.949	.932	.916	.901	.887	.873	.860	.848	.835	.823	.811
320	.976	.961	.948	.936	.925	.914	.903	.893	.883	.873	.864	.854
440	.982	.972	.962	.953	.945	.937	.929	.921	.913	.906	.899	.891
600	.987	.979	.972	.966	.959	.953	.947	.941	.936	.930	.924	.919
800	.990	.984	.979	.974	.969	.965	.960	.956	.951	.947	.943	.939
1000	.992	.987	.983	.979	.975	.972	.968	.964	.961	.957	.954	.950

^{a}Multiply entry by 10^{-3}.

(continued)