TUTORIAL 5
STA437 WINTER 2015

AL NOSEDAL

Contents

1. Two-sample profile analysis

Example. Four psychological tests were given to 32 men and 32 women. The data are recorded in Table 5.1. The variables are:

y_1 = pictorial inconsistencies,
y_2 = paper from board,
y_3 = tool recognition,
y_4 = vocabulary.

The mean vectors are

$$\bar{\mathbf{y}}_1 = \begin{pmatrix} 15.97 \\ 15.91 \\ 27.19 \\ 22.75 \end{pmatrix}$$

$$\bar{\mathbf{y}}_2 = \begin{pmatrix} 12.34 \\ 13.91 \\ 16.66 \\ 21.94 \end{pmatrix}$$

The covariance matrices of the two samples are

$$\mathbf{S_1} = \begin{pmatrix} 5.192 & 4.545 & 6.522 & 5.250 \\ 4.545 & 13.18 & 6.760 & 6.266 \\ 6.522 & 6.760 & 28.67 & 14.47 \\ 5.250 & 6.266 & 14.47 & 16.65 \end{pmatrix}$$

$$\mathbf{S_2} = \begin{pmatrix} 9.136 & 7.549 & 4.864 & 4.151 \\ 7.549 & 18.60 & 10.22 & 5.446 \\ 4.864 & 10.22 & 30.04 & 13.49 \\ 4.151 & 5.446 & 13.49 & 28.00 \end{pmatrix}$$

The pooled covariance matrix is

$$\mathbf{S_p} = \frac{1}{32+32-2}[(32-1)\mathbf{S}_1 + (32-1)\mathbf{S}_2] = \begin{pmatrix} 7.164 & 6.047 & 5.693 & 4.701 \\ 6.047 & 15.89 & 8.492 & 5.856 \\ 5.693 & 8.492 & 29.36 & 13.98 \\ 4.701 & 5.856 & 13.98 & 22.32 \end{pmatrix}$$

We use the psychological data in PSYCH.DAT to illustrate two-sample profile analysis.

1.1. **Parallelism Hypothesis.** To test for parallelism, $H_{01} : \mathbf{C}\mu_\mathbf{1} = \mathbf{C}\mu_\mathbf{2}$ at level α, we use the matrix

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

and

$$T^2 = (\mathbf{C}\bar{\mathbf{y}}_1 - \mathbf{C}\bar{\mathbf{y}}_2)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{C}\mathbf{S_p}\mathbf{C}'\right]^{-1} (\mathbf{C}\bar{\mathbf{y}}_1 - \mathbf{C}\bar{\mathbf{y}}_2).$$

Reject H_{01} when $T^2 > c^2$, where $c^2 = \frac{(n_1+n_2-2)(p-1)}{n_1+n_2-p} F_{p-1,n_1+n_2-p}(\alpha)$.

R code

```
## Reading data

data<-read.table(file="PSYCH.DAT")

## breaking down data
males<-data[1:32 ,-1]
females<-data[33:64 ,-1]
```

```
## sample sizes
n1<-dim(males)[1]
n2<-dim(females)[1]

## mean vectors
y.bar.1<-apply(males,2,FUN=mean)
y.bar.2<-apply(females,2,FUN=mean)

## covariance matrices
S.1<-cov(males)
S.2<-cov(females)
Sp<-(n1+n2-2)^(-1)*((n1-1)*S.1 + (n2-1)*S.2)


## Test for parallelism H01: Cmu1 = Cmu2

## C matrix

C<-matrix(c(-1,0,0,1,-1,0,0,1,-1,0,0,1),nrow=3,ncol=4)

## Hotelling's T^2

K<-(n1*n2/(n1+n2))

T.2<-K*t( C%*%(y.bar.1-y.bar.2) )%*%solve( C%*%Sp%*%t(C) )%*%C%*%(y.bar.1-y.bar.2)

T.2

## Critical value

p<-dim(C)[1]

crit.val<-((n1+n2-2)*p/(n1+n2-p-1))*qf(0.99,p,n1+n2-p-1)

crit.val
```

Another way, using MANOVA.
R code

```
## Test for parallelism H01: Cmu1 = Cmu2

new.data<-matrix(unlist(data[ ,-1]),nrow=64,ncol=4)%*%t(C)
```

```
groups<-factor(data[ ,1])

groups

Y<-cbind(new.data[ ,1],new.data[ ,2],new.data[ ,3])

Y

fit<-manova(Y~groups)

## showing MANOVA table

summary(fit,test="Roy")

sum.roy<-summary(fit,test="Roy")

## largest eigenvalue

lambda.1<-sum.roy$Eigen[1]

## N = total number of individuals

N<-n1 + n2

## g = number of groups or samples

g<-2

## largest univariate F

F.a<-(N-g)*lambda.1/(g-1)

F.a
```

Upon comparison of this value (our $T.2$ or $F.a$) with $T_{0.01,3,62} = 12.79$ (obtained from our F approximation or by interpolation in Table A.7), we reject the hypothesis of parallelism.

1.2. **Levels Hypothesis.** To test for equal levels, $H_{02} : \mathbf{j}'\mu_1 = \mathbf{j}'\mu_2$ at level α, we use

$$T^2 = (\mathbf{j}'\bar{\mathbf{y}}_1 - \mathbf{j}'\bar{\mathbf{y}}_2)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{j}'\mathbf{S_p}\mathbf{j} \right]^{-1} (\mathbf{j}'\bar{\mathbf{y}}_1 - \mathbf{j}'\bar{\mathbf{y}}_2).$$

Reject H_{02} when $T^2 > c^2$, where $c^2 = F_{1,n_1+n_2-2}(\alpha)$.

```
## Test for equal levels H02: 1' mu1 = 1' mu2

C2<-matrix(c(1,1,1,1),ncol=4)

## Hotelling's T^2

K<-(n1*n2/(n1+n2))

T.2<-K*t(C2%*%(y.bar.1-y.bar.2))%*%solve(C2%*%Sp%*%t(C2))%*%C2%*%(y.bar.1-y.bar.2)

## Critical value (ours), using F distribution

our.crit.val<-qf(1-0.01,1,n1+n2-2)

our.crit.val

## On your textbook, test statistic and critical value.

t2<-sqrt(T.2)

t2

## Critical value

crit.val<-qt(1-0.005,n1+n2-2)

crit.val
```

Another way, using ANOVA.

R code

```
## Test for equal levels H02: 1' mu1 = 1' mu2

C2<-matrix(c(1,1,1,1),ncol=4)
```

```
new.data<-matrix(unlist(data[ ,-1]),nrow=64,ncol=4)%*%t(C2)

groups<-factor(data[ ,1])

groups

Y<-cbind(new.data[ ,1])

Y

fit<-aov(Y~groups)

## showing ANOVA table

anova(fit)

aov.table<-anova(fit)

## T.2 = F value

F.val<-aov.table$F[1]

F.val
```

Comparing our test statistic with $t_{0.005,62} = 2.658$ (or using T^2 and our critical value from an F distribution), we reject the hypothesis of equal levels.

1.3. **Flatness Hypothesis.** To test the flatness hypothesis, $H_{03} : \frac{1}{2}\mathbf{C}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = \mathbf{0}$ at level α, we use

$$T^2 = (n_1 + n_2)(\mathbf{C\bar{y}})'(\mathbf{CS_pC}')^{-1}\mathbf{C\bar{y}}.$$

Reject H_{03} when $T^2 > c^2$, where $c^2 = \frac{(n_1+n_2-1)(p-1)}{(n_1+n_2-p+1)} F_{1,n_1+n_2-2}(\alpha)$.

We first calculate $\bar{\mathbf{y}} = \frac{32\bar{\mathbf{y}}_1+32\bar{\mathbf{y}}_2}{32+32} = \begin{pmatrix} 14.16 \\ 14.91 \\ 21.92 \\ 22.34 \end{pmatrix}$

Using

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

we obtain $T^2 = 254.004$.

R code

```
## Test for flatness H03: (1/2)C(mu1+mu2)=0
##                    H03: C mu1 = -C mu2

y.bar<-(0.5)*(y.bar.1 + y.bar.2)

## C matrix

C3<-matrix(c(1,0,0,-1,1,0,0,-1,1,0,0,-1),nrow=3,ncol=4)

## Hotelling's T^2

T.2<-(n1+n2)*t( C3%*%(y.bar) )%*%solve( C3%*%Sp%*%t(C3) )%*%C3%*%(y.bar)


## Critical value

p<-dim(C3)[2]

crit.val<-((n1+n2-1)*(p-1)/(n1+n2-p+1))*qf(1-0.01,p-1,n1+n2-p+1)

crit.val
```

Another way, using MANOVA.

```
## Test for flatness H03: (1/2)C(mu1+mu2)=0
##                    H03: C mu1 = -C mu2

C3<-matrix(c(1,1,1,-1,0,0,0,-1,0,0,0,-1),nrow=3,ncol=4)

new.data.1<-matrix(unlist(data[1:32 ,-1]),nrow=32,ncol=4)%*%t(C3)

new.data.2<-(-1)*matrix(unlist(data[33:64 ,-1]),nrow=32,ncol=4)%*%t(C3)

new.data<-rbind(new.data.1,new.data.2)
```

```
groups<-factor(data[ ,1])

groups

Y<-cbind(new.data[ ,1],new.data[ ,2],new.data[ ,3])

Y

fit<-manova(Y~groups)

## showing MANOVA table

summary(fit,test="Roy")

sum.roy<-summary(fit,test="Roy")

## largest eigenvalue

lambda.1<-sum.roy$Eigen[1]

## N = total number of individuals

N<-n1 + n2

## g = number of groups or samples
g<-2

## largest univariate F

F.a<-(N-g)*lambda.1/(g-1)

F.a
```

which exceeds $T^2_{0.01,3,62} = 12.796$, so we reject the hypothesis of flatness.

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

$\alpha = .05$

27	4.210	6.997	9.692	12.493	15.496	18.770	22.388	26.428	30.985	36.176
28	4.196	6.957	9.612	12.359	15.287	18.463	21.950	25.818	30.149	35.043
29	4.183	6.919	9.539	12.236	15.097	18.184	21.555	25.272	29.407	34.044
30	4.171	6.885	9.471	12.123	14.924	17.931	21.198	24.781	28.742	33.156
35	4.121	6.744	9.200	11.674	14.240	16.944	19.823	22.913	26.252	29.881
40	4.085	6.642	9.005	11.356	13.762	16.264	18.890	21.668	24.624	27.783
45	4.057	6.564	8.859	11.118	13.409	15.767	18.217	20.781	23.477	26.326
50	4.034	6.503	8.744	10.934	13.138	15.388	17.709	20.117	22.627	25.256
55	4.016	6.454	8.652	10.787	12.923	15.090	17.311	19.600	21.972	24.437
60	4.001	6.413	8.577	10.668	12.748	14.850	16.992	19.188	21.451	23.790
70	3.978	6.350	8.460	10.484	12.482	14.485	16.510	18.571	20.676	22.834
80	3.960	6.303	8.375	10.350	12.289	14.222	16.165	18.130	20.127	22.162
90	3.947	6.267	8.309	10.248	12.142	14.022	15.905	17.801	19.718	21.663
100	3.936	6.239	8.257	10.167	12.027	13.867	15.702	17.544	19.401	21.279
110	3.927	6.216	8.215	10.102	11.934	13.741	15.540	17.340	19.149	20.973
120	3.920	6.196	8.181	10.048	11.858	13.639	15.407	17.172	18.943	20.725
150	3.904	6.155	8.105	9.931	11.693	13.417	15.121	16.814	18.504	20.196
200	3.888	6.113	8.031	9.817	11.531	13.202	14.845	16.469	18.083	19.692
400	3.865	6.052	7.922	9.650	11.297	12.890	14.447	15.975	17.484	18.976
1000	3.851	6.015	7.857	9.552	11.160	12.710	14.217	15.692	17.141	18.570
∞	3.841	5.991	7.815	9.488	11.070	12.592	14.067	15.507	16.919	18.307

(continued)

Table A.7. (*Continued*)

Degrees of Freedom, ν	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
					$\alpha = .01$					
2	98.503									
3	34.116	297.000								
4	21.198	82.177	594.997							
5	16.258	45.000	147.283	992.494						
6	13.745	31.857	75.125	229.679	1489.489					
7	12.246	25.491	50.652	111.839	329.433	2085.984				
8	11.259	21.821	39.118	72.908	155.219	446.571	2781.978			
9	10.561	19.460	32.598	54.890	98.703	205.293	581.106	3577.472		
10	10.044	17.826	28.466	44.838	72.882	128.067	262.076	733.045	4472.464	
11	9.646	16.631	25.637	38.533	58.618	93.127	161.015	325.576	902.392	5466.956
12	9.330	15.722	23.588	34.251	49.739	73.969	115.640	197.555	395.797	1089.149
13	9.074	15.008	22.041	31.171	43.745	62.114	90.907	140.429	237.692	472.742
14	8.862	14.433	20.834	28.857	39.454	54.150	75.676	109.441	167.499	281.428
15	8.683	13.960	19.867	27.060	36.246	48.472	65.483	90.433	129.576	196.853
16	8.531	13.566	19.076	25.626	33.672	44.240	58.241	77.755	106.391	151.316
17	8.400	13.231	18.418	24.458	31.788	40.975	52.858	68.771	90.969	123.554
18	8.285	12.943	17.861	23.487	30.182	38.385	48.715	62.109	80.067	105.131
19	8.185	12.694	17.385	22.670	28.852	36.283	45.435	56.992	71.999	92.134
20	8.096	12.476	16.973	21.972	27.734	34.546	42.779	52.948	65.813	82.532
21	8.017	12.283	16.613	21.369	26.781	33.088	40.587	49.679	60.932	75.181
22	7.945	12.111	16.296	20.843	25.959	31.847	38.750	46.986	56.991	69.389
23	7.881	11.958	16.015	20.381	25.244	30.779	37.188	44.730	53.748	64.719
24	7.823	11.820	15.763	19.972	24.616	29.850	35.846	42.816	51.036	60.879
25	7.770	11.695	15.538	19.606	24.060	29.036	34.680	41.171	48.736	57.671
26	7.721	11.581	15.334	19.279	23.565	28.316	33.659	39.745	46.762	54.953
27	7.677	11.478	15.149	18.983	23.121	27.675	32.756	38.496	45.051	52.622

$\alpha = .01$

28	7.636	11.383	14.980	18.715	22.721	27.101	31.954	37.393	43.554	50.604
29	7.598	11.295	14.825	18.471	22.359	26.584	31.236	36.414	42.234	48.839
30	7.562	11.215	14.683	18.247	22.029	26.116	30.589	35.538	41.062	47.283
35	7.419	10.890	14.117	17.366	20.743	24.314	28.135	32.259	36.743	41.651
40	7.314	10.655	13.715	16.750	19.858	23.094	26.502	30.120	33.984	38.135
45	7.234	10.478	13.414	16.295	19.211	22.214	25.340	28.617	32.073	35.737
50	7.171	10.340	13.181	15.945	18.718	21.550	24.470	27.504	30.673	33.998
55	7.119	10.228	12.995	15.667	18.331	21.030	23.795	26.647	29.603	32.682
60	7.077	10.137	12.843	15.442	18.018	20.613	23.257	25.967	28.760	31.650
70	7.011	9.996	12.611	15.098	17.543	19.986	22.451	24.957	27.515	30.139
80	6.963	9.892	12.440	14.849	17.201	19.536	21.877	24.242	26.642	29.085
90	6.925	9.813	12.310	14.660	16.942	19.197	21.448	23.710	25.995	28.310
100	6.895	9.750	12.208	14.511	16.740	18.934	21.115	23.299	25.496	27.714
110	6.871	9.699	12.125	14.391	16.577	18.722	20.849	22.972	25.101	27.243
120	6.851	9.657	12.057	14.292	16.444	18.549	20.632	22.705	24.779	26.862
150	6.807	9.565	11.909	14.079	16.156	18.178	20.167	22.137	24.096	26.054
200	6.763	9.474	11.764	13.871	15.877	17.819	19.720	21.592	23.446	25.287
400	6.699	9.341	11.551	13.569	15.473	17.303	19.080	20.818	22.525	24.209
1000	6.660	9.262	11.426	13.392	15.239	17.006	18.743	20.376	22.003	23.600
∞	6.635	9.210	11.345	13.277	15.086	16.812	18.475	20.090	21.666	23.209

Note: p = number of variables.