MVA HW3

Due in class on March 1

1. Prove the facts about the multivariate normal distribution.

   (a) Conditional distribution: If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \underset{\approx}{\Sigma})$ and

$$\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix}, \underset{\sim}{\mu}_X = \begin{pmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{pmatrix}, \underset{\approx}{\Sigma}_X = \begin{pmatrix} \underset{\approx}{\Sigma}_{11} & \underset{\approx}{\Sigma}_{12} \\ \underset{\approx}{\Sigma}_{21} & \underset{\approx}{\Sigma}_{22} \end{pmatrix},$$
$$\underset{\sim}{X}_1 \in \mathbb{R}^{p_1},\ \underset{\sim}{X}_2 \in \mathbb{R}^{p_2},\ \underset{\sim}{X} \in \mathbb{R}^{p}, p = p_1 + p_2$$

   then, $\underset{\sim}{X}_1 | \underset{\sim}{X}_2 = \underset{\sim}{x}_2 \sim N_{p_1}(\underset{\sim}{\mu}_{1\cdot 2}(\underset{\sim}{x}_2), \underset{\approx}{\Sigma}_{11\cdot 2})$, where $\underset{\sim}{\mu}_{1\cdot 2}(\underset{\sim}{x}_2) = \underset{\sim}{\beta}_0 + \underset{\approx}{B}_1 \underset{\sim}{x}_2$, where $\underset{\approx}{B}_1 = \underset{\approx}{\Sigma}_{12} \underset{\approx}{\Sigma}_{22}^{-1}$, $\underset{\sim}{\beta}_0 = \underset{\sim}{\mu}_1 - \underset{\approx}{B}_1 \underset{\sim}{\mu}_2$ and $\underset{\approx}{\Sigma}_{11\cdot 2} = \underset{\approx}{\Sigma}_{11} - \underset{\approx}{\Sigma}_{12} \underset{\approx}{\Sigma}_{22}^{-1} \underset{\approx}{\Sigma}_{21}$.

(b) MLE: Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n$ be iid random samples from $N_p(\underset{\sim}{\mu}, \underset{\approx}{\Sigma})$. Then the MLEs are

$$\begin{aligned}\hat{\underset{\sim}{\mu}} &= \bar{\underset{\sim}{x}} \\ \hat{\underset{\approx}{\Sigma}} &= \frac{1}{n}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})',\end{aligned}$$

and the maximized value of the likelihood is

$$L(\hat{\underset{\sim}{\mu}}, \hat{\underset{\approx}{\Sigma}}) = (2\pi)^{-np/2} e^{-np/2} \frac{1}{|\hat{\underset{\approx}{\Sigma}}|^{n/2}},$$

which depends on the sample only through the estimated generalized variance $|\hat{\underset{\approx}{\Sigma}}|$.

5.1. (a) Evaluate T^2, for testing H_0: $\boldsymbol{\mu}' = [7, \quad 11]$, using the data

$$\mathbf{X} = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$$

(b) Specify the distribution of T^2 for the situation in (a).

(c) Using (a) and (b), test H_0 at the $\alpha = .05$ level. What conclusion do you reach?

5.3. (a) Use expression (5-15) to evaluate T^2 for the data in Exercise 5.1.

(b) Use the data in Exercise 5.1 to evaluate Λ in (5-13). Also, evaluate Wilks' lambda.

$$T^2 = \frac{(n-1)|\hat{\boldsymbol{\Sigma}}_0|}{|\hat{\boldsymbol{\Sigma}}|} - (n-1)$$

$$= \frac{(n-1)\left|\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right|}{\left|\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right|} - (n-1) \qquad (5\text{-}15)$$

$$\Lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{n/2} = \left(\frac{\left|\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right|}{\left|\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right|}\right)^{n/2} < c_\alpha \qquad (5\text{-}13)$$

$\Lambda^{2/n} = |\hat{\boldsymbol{\Sigma}}|/|\hat{\boldsymbol{\Sigma}}_0|$ is called *Wilks' lambda.*

5.4. Use the sweat data in Table 5.1. (See Example 5.2.)

(a) Determine the axes of the 90% confidence ellipsoid for $\boldsymbol{\mu}$. Determine the lengths of these axes.

Table 5.1 Sweat Data

Individual	X_1 (Sweat rate)	X_2 (Sodium)	X_3 (Potassium)
1	3.7	48.5	9.3
2	5.7	65.1	8.0
3	3.8	47.2	10.9
4	3.2	53.2	12.0
5	3.1	55.5	9.7
6	4.6	36.1	7.9
7	2.4	24.8	14.0
8	7.2	33.1	7.6
9	6.7	47.4	8.5
10	5.4	54.1	11.3
11	3.9	36.9	12.7
12	4.5	58.8	12.3
13	3.5	27.8	9.8
14	4.5	40.2	8.4
15	1.5	13.5	10.1
16	8.5	56.4	7.1
17	4.5	71.6	8.2
18	6.5	52.8	10.9
19	4.1	44.1	11.2
20	5.5	40.9	9.4

Source: Courtesy of Dr. Gerald Bargman.

5.7. Use the sweat data in Table 5.1 (See Example 5.2.) Find simultaneous 95% T^2 confidence intervals for μ_1, μ_2, and μ_3 using Result 5.3. Construct the 95% Bonferroni intervals using (5-29). Compare the two sets of intervals.

5.13. Determine the approximate distribution of $-n \ln(|\hat{\mathbf{\Sigma}}|/|\hat{\mathbf{\Sigma}}_0|)$ for the sweat data in Table 5.1. (See Result 5.2.)

5.18. Use the college test data in Table 5.2. (See Example 5.5.)

(a) Test the null hypothesis H_0: $\boldsymbol{\mu}' = [500, 50, 30]$ versus H_1: $\boldsymbol{\mu}' \neq [500, 50, 30]$ at the $\alpha = .05$ level of significance. Suppose $[500, 50, 30]'$ represent average scores for thousands of college students over the last 10 years. Is there reason to believe that the group of students represented by the scores in Table 5.2 is scoring differently? Explain.

(b) Determine the lengths and directions for the axes of the 95% confidence ellipsoid for $\boldsymbol{\mu}$.

Table 5.2 College Test Data

Individual	X_1 (Social science and history)	X_2 (Verbal)	X_3 (Science)	Individual	X_1 (Social science and history)	X_2 (Verbal)	X_3 (Science)
1	468	41	26	45	494	41	24
2	428	39	26	46	541	47	25
3	514	53	21	47	362	36	17
4	547	67	33	48	408	28	17
5	614	61	27	49	594	68	23
6	501	67	29	50	501	25	26
7	421	46	22	51	687	75	33
8	527	50	23	52	633	52	31
9	527	55	19	53	647	67	29
10	620	72	32	54	647	65	34
11	587	63	31	55	614	59	25
12	541	59	19	56	633	65	28
13	561	53	26	57	448	55	24
14	468	62	20	58	408	51	19
15	614	65	28	59	441	35	22
16	527	48	21	60	435	60	20
17	507	32	27	61	501	54	21
18	580	64	21	62	507	42	24
19	507	59	21	63	620	71	36
20	521	54	23	64	415	52	20
21	574	52	25	65	554	69	30
22	587	64	31	66	348	28	18
23	488	51	27	67	468	49	25
24	488	62	18	68	507	54	26
25	587	56	26	69	527	47	31
26	421	38	16	70	527	47	26
27	481	52	26	71	435	50	28
28	428	40	19	72	660	70	25
29	640	65	25	73	733	73	33
30	574	61	28	74	507	45	28
31	547	64	27	75	527	62	29
32	580	64	28	76	428	37	19
33	494	53	26	77	481	48	23
34	554	51	21	78	507	61	19
35	647	58	23	79	527	66	23
36	507	65	23	80	488	41	28
37	454	52	28	81	607	69	28
38	427	57	21	82	561	59	34
39	521	66	26	83	614	70	23
40	468	57	14	84	527	49	30
41	587	55	30	85	474	41	16
42	507	61	31	86	441	47	26
43	574	54	31	87	607	67	32
44	507	53	23				

Source: Data courtesy of Richard W. Johnson.

5.21. Using the data on bone mineral content in Table 1.8, construct the 95% Bonferroni intervals for the individual means. Also, find the 95% simultaneous T^2-intervals. Compare the two sets of intervals.

Table 1.8 Mineral Content in Bones

Subject number	Dominant radius	Radius	Dominant humerus	Humerus	Dominant ulna	Ulna
1	1.103	1.052	2.139	2.238	.873	.872
2	.842	.859	1.873	1.741	.590	.744
3	.925	.873	1.887	1.809	.767	.713
4	.857	.744	1.739	1.547	.706	.674
5	.795	.809	1.734	1.715	.549	.654
6	.787	.779	1.509	1.474	.782	.571
7	.933	.880	1.695	1.656	.737	.803
8	.799	.851	1.740	1.777	.618	.682
9	.945	.876	1.811	1.759	.853	.777
10	.921	.906	1.954	2.009	.823	.765
11	.792	.825	1.624	1.657	.686	.668
12	.815	.751	2.204	1.846	.678	.546
13	.755	.724	1.508	1.458	.662	.595
14	.880	.866	1.786	1.811	.810	.819
15	.900	.838	1.902	1.606	.723	.677
16	.764	.757	1.743	1.794	.586	.541
17	.733	.748	1.863	1.869	.672	.752
18	.932	.898	2.028	2.032	.836	.805
19	.856	.786	1.390	1.324	.578	.610
20	.890	.950	2.187	2.087	.758	.718
21	.688	.532	1.650	1.378	.533	.482
22	.940	.850	2.334	2.225	.757	.731
23	.493	.616	1.037	1.268	.546	.615
24	.835	.752	1.509	1.422	.618	.664
25	.915	.936	1.971	1.869	.869	.868

Source: Data courtesy of Everett Smith.