KEY

Assignment 1 (Due at 1 p.m. on Friday, February 17, 2006)

This is an open-book, open-note assignment. Collaboration is not allowed. Late assignments will not be accepted. Attach only those printouts that you used in answering the questions. Indicate on these printouts where you got your answer to each question. If you do any hand calculations, please show all your work, as no credit will be given for unsupported answers. Failure to follow these instructions will result in a loss of points from your score. Good luck!

Consider Exercise 5.18, p. 267, in our text. (Refer also to Example 5.5, pp. 226-229 and Table 5.2, p. 228). These data are available on the www.BIOS6244.com website.

Testing for Multivariate Normality

(1) Univariate normality of each variable (marginal distributions).

(a) Stem and Leaf Plots.

If the stem and leaf plot suggests skewness, indicate the direction (right or left). If it indicates kurtosis, indicate the type of tails (heavy or light).

4/

Variable	Suggested Shape
X_1	symmetrical
X_2	slight negative skewness
X_3	symmetrical

(b) Shapiro-Wilk Tests.

6/

Variable	Test Statistic	p-value	Interpretation	Conclusion
X_1	.989	.686	no evidence	normal
X_2	.970	.039	moderate evidence	non-normal
X_3	.987	.536	no evidence	normal

(c) Normal Probability Plots.

Variable	Conclusion
X_1	normal
X_2	normal
X_3	normal

3/

(d) Correlation Coefficient Tests.

Variable	Test Statistic	p-value	Interpretation	Conclusion
X_1	.996	.750	no evidence	normal
X_2	.987	.081	weak evidence	probably normal
X_3	.9979	.990	no evidence	normal

8/

(See work on p-7)

(e) Consensus of Univariate Tests on Marginal Distributions.

Variable	Conclusion	Reason(s)
X_1	normal	4/4 tests indicate normality
X_2	non-normal	2/4 tests indicate non-normality
X_3	normal	4/4 indicate normality

6/

(2) Univariate normality of each principal component.

(a) Stem and Leaf Plots.

If the stem and leaf plot suggests skewness, indicate the direction (right or left). If it indicates kurtosis, indicate the type of tails (heavy or light).

P.C.	Suggested Shape
1	slightly skewed right
2	slightly skewed left
3	slightly skewed left

4/

6/

(b) Shapiro-Wilk Tests.

P.C.	Test Statistic	p-value	Interpretation	Conclusion
1	.990	.736	no evidence	normal
2	.976	.107	no evidence	normal
3	.985	.412	no evidence	normal

3/

(c) Normal Probability Plots.

P.C.	Conclusion
1	normal
2	non-normal
3	normal

8/

(d) Correlation Coefficient Tests.

P.C.	Test Statistic	p-value	Interpretation	Conclusion
1	.995	.625	no evidence	normal
2	.987	.081	weak evidence	probably normal
3	.995	.625	no evidence	normal

(See work on p. 7.)

6/

(f) Consensus of Tests on PC's.

P.C.	Conclusion	Reason(s)
1	normal	4/4 tests indicate normality
2	probably normal	3/4 tests indicate normality
3	normal	4/4 tests indicate normality

3/

(3) Beta Plot of Squared Radii

What is your conclusion from this plot? Why?

It appears to be fairly linear and to support the assumption of MVN. There appears to be at least one outlier in terms of the squared radii, however.

4/

(4) Srivastava-Hui Tests of Multivariate Normality

Test	Test Statistic	p-value	Interpretation	Conclusion
M_1	4.74	.577	no evidence	fail to reject MVN
M_2	0.976	.407	no evidence	fail to reject MVN

(Note: Use the estimated values of γ, δ, and ε provided in class.)

3/

(5) Summary.

For each of the following techniques, give your conclusion and a brief justification.

Technique	Conclusion	Justification
Marginal Tests	Reject MVN	X_2 appears to be non-normal
Tests on PC's	Fail to reject MVN	All PC's appear to be UVN
Direct Tests of MVN	Fail to reject MVN	Neither test indicates non-MVN

4/

(6) What is your overall assessment of MVN? Why? If you reject MVN, what remedy(ies) would you use?

The assumption of MVN appears to be reasonable. The significant result for the S-W test for X_2 may be a false positive.

Tests on Mean Vector

Suppose that we wish to compare these 87 students with students typically admitted by L.S.U. The mean vector for L.S.U. is given by $\underset{\sim}{\mu} = [525\ \ 56\ \ 26]$.

(1) Give the results for Hotelling's T^2.

3 ✓

T^2_{cal}	F_{cal}	d.f.	p-value	Intrepretation
8.02	2.61	3,84	.057	Fail to reject H_0

$df_1 = p = 3,\ df_2 = n-p = 87-3 = 84$

Based on these results, is there sufficient reason to believe that the group of students represented by the scores in Table 5.2 is scoring any differently from those admitted by L.S.U.? Why or why not?

3 ✓

No, there is insufficient evidence to reject H_0 since $p > .05$, Therefore, we conclude that there is no difference between the population mean vector for these students and that for the students typically admitted by LSU.

(2) Give the results for the "robust" version of Hotelling's T^2.

2 ✓

χ^2_{cal}	d.f.	p-value	Intrepretation
8.02	3	.046	Reject H_0

What do the results for this test indicate about the hypothesized mean vector?

3 ✓

This is moderate evidence to reject H_0 since $p = .046 < .05$. Therefore, conclude that the population mean for these students does differ significantly from that for students typically admitted to LSU.

Simultaneous Tests on Means

(1) Consider the simultaneous confidence intervals given in Example 5.4, pp. 227-229. Note that these are the T^2-based intervals. What differences from the hypothesized values, if any, are indicated by these intervals? Give a reason in each case.

5 ✓

Mean	Interval	Conclusion	Justification
μ_1	(503.30, 549.88)	$\mu_{10} = 525$ is reasonable	since $525 \in CI(\mu_1)$
μ_2	(51.22, 58.16)	$\mu_{20} = 56$ is reasonable	since $56 \in CI(\mu_2)$
μ_3	(23.65, 26.61)	$\mu_{30} = 26$ is reasonable	since $26 \in CI(\mu_3)$

5

(2) Now construct the same intervals using the "robust" approach. What differences from the hypothesized values, if any, are indicated by these intervals? Give a reason in each case. (See work below.)

Mean	Interval	Conclusion	Justification
μ_1	(503.99, 549.19)	Same as in Question (1) above.	
μ_2	(51.33, 58.05)		
μ_3	(23.69, 26.57)		

7

(3) Compare the lengths of your intervals in Questions (1) and (2) above. Comment.

Mean	T^2 Length	Robust Length	Difference
μ_1	46.58	45.20	1.38
μ_2	6.94	6.72	0.22
μ_3	2.96	2.88	0.08

The robust interval is shorter in every case, but the difference is never more than 1.5 points, a negligible amount in this case.

Choice of Technique

4

Which technique is more appropriate in this case, Hotelling's T^2 or the "robust" approach? Why? Since the evidence against MVN is not conclusive, the results based on Hotelling's T^2 are more appropriate.

Work for (2) above:

$$\chi^2_p(\alpha) = \chi^2_3(.05) = 7.81$$

$$95\%\ CI(\mu_1):\ 526.59 \pm \sqrt{7.81}\sqrt{\frac{5691.34}{87}} = 526.59 \pm 22.60 = (503.99, 549.19)$$

$$95\%\ CI(\mu_2):\ 54.69 \pm \sqrt{7.81}\sqrt{\frac{126.05}{87}} = 54.69 \pm 3.36 = (51.33, 58.05)$$

$$95\%\ CI(\mu_3):\ 25.13 \pm \sqrt{7.81}\sqrt{\frac{23.11}{87}} = 25.13 \pm 1.44 = (23.69, 26.57)$$

#

Work for Correlation Coefficient P-Values

You must first interpolate to find the appropriate critical values for $n = 87$. From Table 2 of Looney & Gulledge (1985) handed out in class:

	.050	.100	.500	.750	.990
$n = 85$	.985	.988	.994	.996	.9979
$n = 87$	$\boxed{.9854}$	$\boxed{.988}$	$\boxed{.994}$	$\boxed{.996}$	$\boxed{.9979}$
$n = 90$	.986	.988	.994	.996	.9979

So, for $n=87$, the .05 percentage point is calculated as follows:

$$\frac{87-90}{85-90} = \frac{x - .986}{.985 - .986}$$

$$\Rightarrow x = .9854$$

Now, for a given r-value for $n=87$, interpolate the p-value:

$r = .996 \Rightarrow p = .750$

$r = .987$:

$$\frac{.987 - .9854}{.988 - .9854} = \frac{p - .05}{.10 - .05}$$

$$\Rightarrow p = .081$$

$r = .9979 \Rightarrow p = .990$

$r = .995 \Rightarrow p = \frac{.5 + .75}{2} = .625$ #