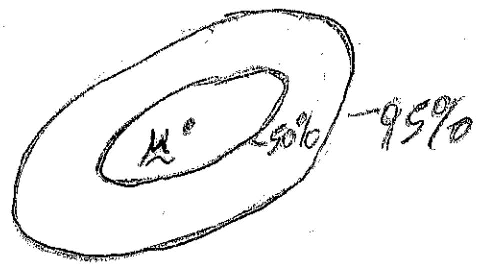
6) p114 The minimum volume $100(1-\alpha)\%$ covering region if $\underline{x}_1, \ldots, \underline{x}_n, \underline{x}_f$ are iid $EC(\underline{\mu}, \Sigma, g)$ with continuous decreasing g, is

$$\{\underline{z} \mid D^2_{\underline{z}}(\underline{\mu}, \Sigma) \le u_{1-\alpha}\}$$ where

$U = D^2_{\underline{x}}(\underline{\mu}, \Sigma)$ has pdf given by (3.10)

and $P(U \le u_{1-\alpha}) = 1-\alpha$.



7) To estimate a $100(1-\alpha)\%$ percentile, use the $\lceil n(1-\alpha) \rceil$th order statistic

eg $\lceil \frac{n}{2} \rceil$ for the median and $\lceil \frac{9n}{10} \rceil$ for the 90th percentile where $\lceil 7.7 \rceil = 8 = \lceil 8 \rceil$.

8) Let $$q_n = \begin{cases} \min\left(1-\alpha+0.05,\ 1-\alpha+\frac{p}{n}\right), & \alpha \ge 0.1 \\ \min\left(1-\frac{\alpha}{2},\ 1-\alpha+\frac{10\alpha p}{n}\right), & \alpha \le 0.1. \end{cases}$$

If $q_n < 1-\alpha+0.001$, use $q_n = 1-\alpha$.

$n = 20p$

Idea: 50%–90% prediction regions tend to have up to 5% undercoverage. 90–99.9% regions also have undercoverage for small n, so for small n, increase nominal 80% PR coverage to 85% and nominal 95% PR coverage to 97.5%. (18.5

9) P114-5 Let $D_{(U_P)}$ be the q_nth percentile of the $D_i(T, C)$. The $100(1-\alpha)\%$ large sample nonparametric PR is $\{\underline{z} \mid D_{\underline{z}}(\underline{\bar{x}}, S) \le D_{(U_P)}(\underline{\bar{x}}, S)\}$,

the semiparametric PR is (works for EC)

$$\{\underline{z} \mid D_{\underline{z}}(T_{RMVN}, C_{RMVN}) \le D_{(U_P)}(T_{RMVN}, C_{RMVN})\},$$

the parametric MVN PR is

$$\{\underline{z} \mid D_{\underline{z}}(T_{RMVN}, C_{RMVN}) \le \chi^2_{p, q_n}\}.$$

10) The 3 regions are asymptotically optimal for $N_p(\mu, \Sigma)$ data with nonsingular Σ and the 1st 2 regions are asymptotically optimal for a large class of EC distributions. The nonparametric PR is a large sample PR if $\Sigma_{\underline{x}}$ is nonsingular, and is very robust to nonnormality, but not to outliers. The semiparametric

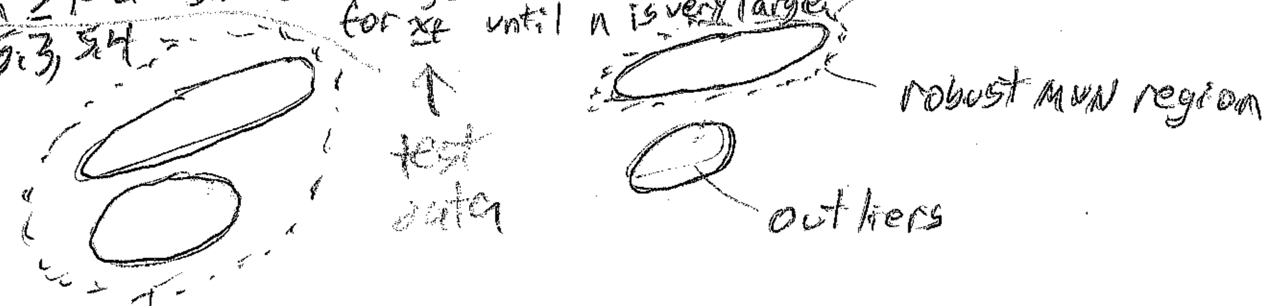
PR can tolerate about $100(1-\delta_n)\%$ outliers, $\delta_n > .5$. MU 19

The MVN region is not robust to nonnormality: it can have severe undercoverage if the bulk of the data is not MVN.

The semi and non parametric regions cover $100\,\delta_n$ % of $\underline{x}_1, \ldots, \underline{x}_n$ if S and C are nonsingular, even if iid or EC assumption does not hold. So these regions are robust to model misspecifications for coverage, although volume could be huge. Use $\delta_n \geq 1-\alpha$ since coverage on training data $\underline{x}_1, \ldots, \underline{x}_n$ is better than that for $\underline{x}_f$ until n is very large.

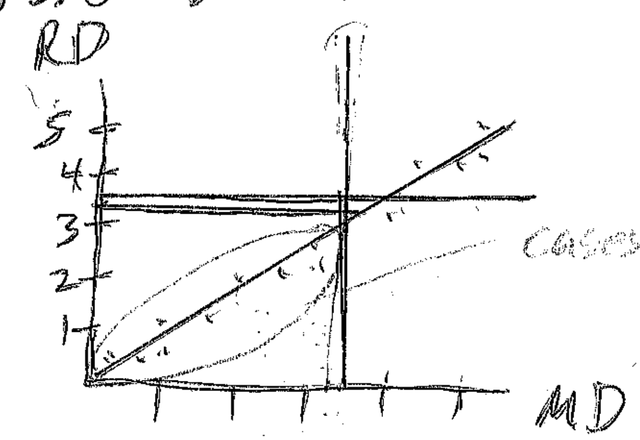
ex Fig 5.3, 5.4

$p = 2$



ex] Fig 5.8 Buxton data without outliers.

See HW4 6 for 80% PR.



cases in nonparametric region $MD_i = D_i(\bar{\underline{x}}, S) \leq D_{(.95n)} \approx 3.09$

semiparametric region $RD_i \leq 3.44$

MVN region $RD_i \leq 3.33$

$p = 5 \quad n = 82 \quad$ 90% PR uses $\delta_n = 0.95 = 1 - .1 + .05$

since $1 - .1 + \frac{5}{82} = .957$

Ch 6 (JW ch 8) Principal Component Analysis (PCA)

1) p124 Let $\Sigma = (\sigma_{ij}) > 0$ be $p \times p$. A generalized correlation matrix $\rho = (\rho_{ij})$ where

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\ \sigma_{jj}}}. \qquad \text{If } \Sigma = c\,\Sigma_x = c\,\text{Cov}(\underline{x}),\ c > 0$$

then the generalized correlation matrix ρ is the correlation matrix $\rho_{\underline{x}} = \rho$.

2) PCA extracts the eigenvalue eigenvector pairs $(\hat{\lambda}_1, \hat{\underline{e}}_1), \ldots, (\hat{\lambda}_p, \hat{\underline{e}}_p)$ of $C = \hat{\Sigma}$ or $C = \hat{\rho}$.

3) $\frac{\text{Know}}{\text{p128}}$ A <u>scree plot</u> is a plot of i vs $\hat{\lambda}_i$ for $i = 1, \ldots, p$.

4) Want to keep $(\hat{\lambda}_1, \hat{\underline{e}}_1), \ldots, (\hat{\lambda}_k, \hat{\underline{e}}_k)$.

want trace explained or variance explained

$$= \frac{\sum_{i=1}^{k} \hat{\lambda}_i}{\sum_{i=1}^{p} \hat{\lambda}_i} \geq 0.95 \text{ or } 0.9 \text{ or } 0.8 \text{ or } 0.7,$$

and there is often a sharp bend in the scree plot when the components are no longer important. For the classical or robust correlation matrix R or R_U, $\sum_{i=1}^{p} \hat{\lambda}_i = p$. For S, $\sum_{i=1}^{p} \hat{\lambda}_i = \sum_{i=1}^{p} \hat{\sigma}_i^2 = tr(S)$.

5) R output shows
"standard deviations" $= \sqrt{\text{eigenvalues}}$

$\sqrt{\hat{\lambda}_1}, \ldots, \sqrt{\hat{\lambda}_p}$

and PC1 ... PCP

$x_1, \ldots, x_p$ $\quad \hat{\underline{e}}_1 \quad \ldots \quad \hat{\underline{e}}_p$

6) Interpretation of $\hat{\underline{e}}_i$: find coefficients that are large in magnitude and ignore the remaining ones. See Exam2 review 57)–59), and ex done in class. This works if R or Ru is used or if the s_i^2 are approx equal.

ex) HW4 output know for Quiz 4

Buxton data with corr matrix R

(HW4 uses rprcomp(z) output not prcomp(z, scale=T) output
← use R instead of S)

a) Make scree plot. What proportion of trace is explained by the 1st 4 principal components?

b) Which principal components correspond to

Soln b)
i) bigonal — PC3
ii) nasal + buxy — PC2
iii) length + cephalic — PC5
iv) length − cephalic — PC1
v) nasal − buxy — PC4

} soln

can multiple eigenvector by ±1

20.5

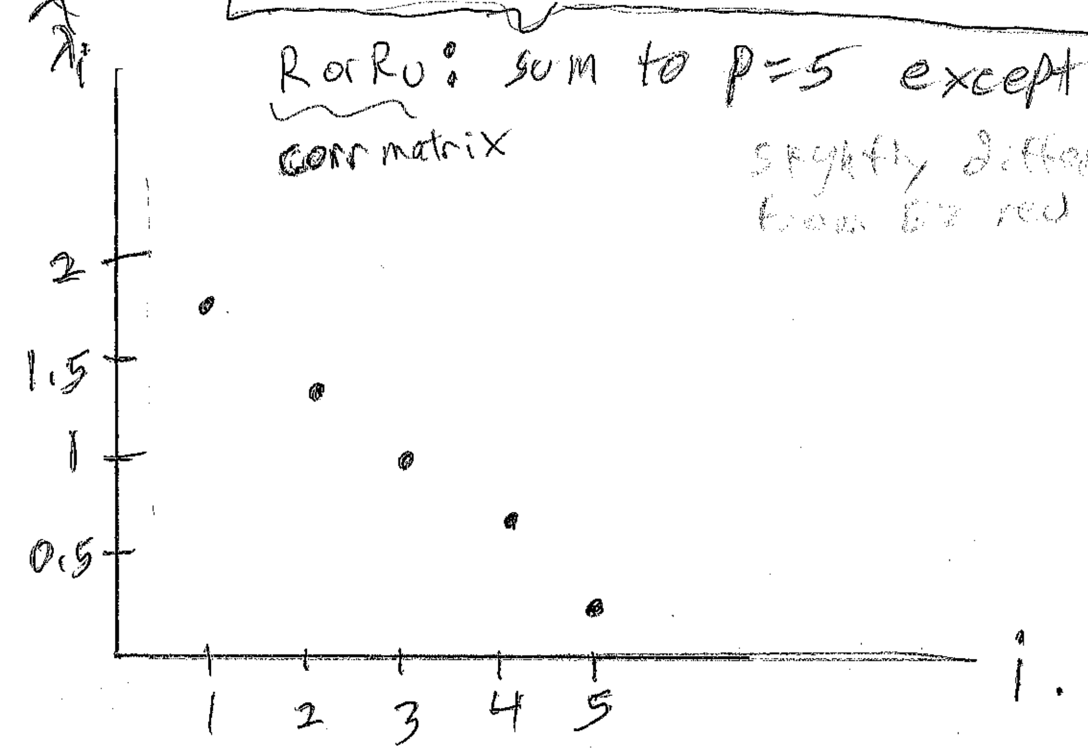
Soln: a) $\sqrt{\hat\lambda_i}$ 1.3184 1.1724 1.0155 0.7867 0.4868

$(\sqrt{\hat\lambda_i})^2 = \hat\lambda_i$ $\underbrace{1.7383 \quad 1.3745 \quad 1.0313 \quad 0.6190 \quad 0.2370}$

R or R_U: sum to $p=5$ except for rounding

corr matrix

(slightly different from Ex red #9)



a) $$\frac{\sum_{i=1}^{4}\hat\lambda_i}{\sum_{i=1}^{p}\hat\lambda_i} = \frac{5-\hat\lambda_5}{5} = \frac{5-.237}{5} = \frac{4.763}{5} = \boxed{0.9526}$$

p for corr matrix

since $\sum_{i=1}^{5}\hat\lambda_i = 5$, not true in general

7) SAS output proc princomp

	eigenvalue	difference	proportion	cumulative
			$\hat\lambda_i / \sum_1^p \hat\lambda_i$	$\hat\lambda_1 / \sum_{i=1}^p \hat\lambda_i$
prin 1	$\hat\lambda_1$			
prin k	$\hat\lambda_K$		$\hat\lambda_K / \sum_1^p \hat\lambda_i$	$\sum_{i=1}^{k}\hat\lambda_i / \sum_{i=1}^{p}\hat\lambda_i$
prin p	$\hat\lambda_p$		$\hat\lambda_p / \sum_1^p \hat\lambda_i$	1

eigenvectors

	prin1	...	prin p
x_1 ⋮ x_p	$\hat e_1$		$\hat e_p$

} like R