56:295 – 001
Multivariate Statistics and Advanced Quality Control
Fall 05

HW5 Due: October 26 (Wednesday), 6:15pm

Solution

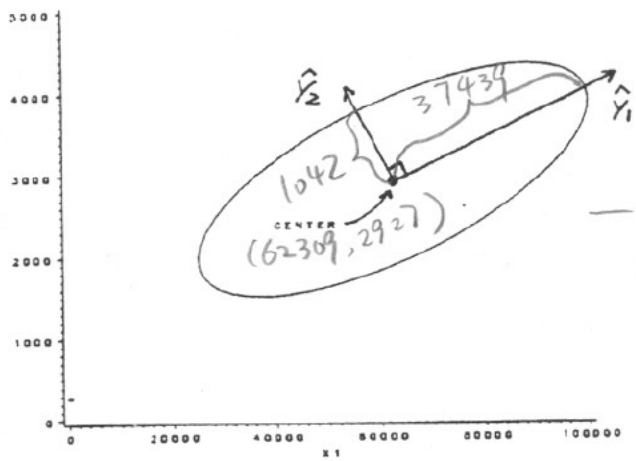
Solve the following problems from the textbook (J&W)

8.6(a)(b)(c) (you can use a software such as Matlab to calculate the eigenvalues and eigenvectors; but do NOT use the PCA function of software)

(a) $\hat{y}_1 = 0.999673x_1 + 0.025574x_2$. Sample variance of $\hat{y}_1 = \hat{\lambda}_1 = 1.0012 \times 10^9$.
$\hat{y}_2 = -0.025574x_1 + 0.999673x_2$. Sample variance of $\hat{y}_2 = \hat{\lambda}_2 = 775,734.3$.

(b) Proportion of total sample variance explained by $\hat{y}_1$ is $\hat{\lambda}_1/(\hat{\lambda}_1 + \hat{\lambda}_2) = 0.99923$.

(c)



— flat. X_1-axis & X_2-axis use different scales here.

8.7(a)(b)(d)

(a) $\hat{y}_1 = 0.707107z_1 + 0.707107z_2$. Sample variance of $\hat{y}_1 = \hat{\lambda}_1 = 1.67615$.
$\hat{y}_2 = -0.707107z_1 + 0.707107z_2$. Sample variance of $\hat{y}_2 = \hat{\lambda}_2 = 0.32385$.

Correlation Matrix

	SALES	PROFITS
SALES	1.0000	0.6762
PROFITS	0.6762	1.0000

(b) Proportion of total sample variance explained by $\hat{y}_1$ is $\hat{\lambda}_1/(\hat{\lambda}_1 + \hat{\lambda}_2) = 0.83808$.

(d) The nature of the principal component is heavily influenced by the relative sizes of the variances of the variables. The correlation coefficients between the components and the variables give some indication of the importance of the variables taking account differences in variances. Standardizing the variables makes the variable units comparable and puts the variables on similar scales.

8.18(a)(b)(c) (use Minitab or another statistical package for PCA; for part (b), do not need to show the correlations of the standardized variables with the components; the data file for Table 1.9 is posted on the homework page).

(a) Correlation Matrix

	X1	X2	X3	X4	X5	X6	X7
X1	1.0000	0.9528	0.8347	0.7277	0.7284	0.7417	0.6863
X2	0.9528	1.0000	0.8570	0.7241	0.6984	0.7099	0.6856
X3	0.8347	0.8570	1.0000	0.8984	0.7878	0.7776	0.7054
X4	0.7277	0.7241	0.8984	1.0000	0.9016	0.8636	0.7793
X5	0.7284	0.6984	0.7878	0.9016	1.0000	0.9692	0.8779
X6	0.7417	0.7099	0.7776	0.8636	0.9692	1.0000	0.8998
X7	0.6863	0.6856	0.7054	0.7793	0.8779	0.8998	1.0000

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulat:
PRIN1	5.80569	5.15204	0.829384	0.82:
PRIN2	0.65365	0.35376	0.093378	0.92:
PRIN3	0.29988	0.17440	0.042840	0.96:
PRIN4	0.12548	0.07166	0.017925	0.98:
PRIN5	0.05382	0.01477	0.007688	0.99:
PRIN6	0.03905	0.01661	0.005578	0.99(
PRIN7	0.02244		0.003206	1.00(

Eigenvectors

	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5	PRIN6	PRIN7
X1	0.368356	0.490060	0.286012	-.319386	0.231169	0.619825	0.052177
X2	0.365364	0.536580	0.229819	0.083302	0.041455	-.710765	-.109225
X3	0.381610	0.246538	-.515367	0.347377	-.572178	0.190946	0.208497
X4	0.384559	-.155402	-.584526	0.042076	0.620324	-.019089	-.315210
X5	0.389104	-.360409	-.012912	-.429539	0.030261	-.231248	0.692562
X6	0.388866	-.347539	0.152728	-.363120	-.463355	0.009277	-.598359
X7	0.367004	-.369208	0.484370	0.672497	0.130536	0.142281	0.069598

(b) $\hat{y}_1 = 0.3684z_1 + 0.3654z_2 + 0.3816z_3 + 0.3846z_4 + 0.3891z_5 + 0.3889z_6 + 0.3670z_7.$

$\hat{y}_2 = 0.4901z_1 + 0.5366z_2 + 0.2465z_3 - 0.1554z_4 - 0.3604z_5 - 0.3475z_6 + 0.3692z_7.$

The cumulative percentage of total variance explained by the first two components is 92.3%.

(c) The first principal component has strong positive correlations with all of the standardized variables. This component may be identified as a measure of athletic excellence (excellence in all running events). The second component has positive correlations with variables z_1—z_3 (100m, 200m, 400m) and negative correlations with variables z_4—z_7 (800m, 1500m, 300m, marathon). This component contrasts excellence at running short distances (primarily speed) with excellence at running the long distances (primarily endurance).