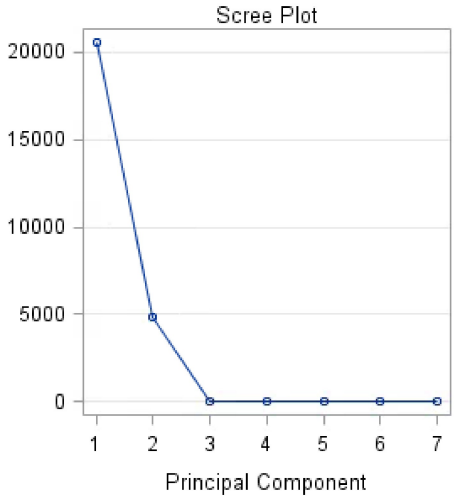
9.32. Perform a factor analysis of the data on bulls given in Table 1.10. Use the seven variables YrHgt, FtFrBody, PrctFFB, Frame, BkFat, SaleHt, and SaleWt. Factor the sample covariance matrix **S** and interpret the factors. Compute factor scores, and check for outliers. Repeat the analysis with the sample correlation matrix **R**. Compare the results obtained from **S** with the results from **R**. Does it make a difference if **R**, rather than **S**, is factored? Explain.

For the factor analysis on matrix **S**:

First, we check the scree plot of matrix S:



Here, we will use m=2, since rest eigenvalues are much smaller than the first two.

The factor analysis using m=2 (PCM) with rotation:

Factor Pattern		
	Factor1	Factor2
YrHgt	0.48777	0.39033
FtFrBody	0.75367	0.65725
PrctFFB	0.37408	0.62342
Frame	0.48170	0.36809
BkFat	0.11083	-0.38394
SaleHt	0.66769	0.29875
SaleWt	0.96506	-0.26204

Final Communality Estimates and Variable Weights		
Total Communality: Weighted = 25454.287 Unweighted = 3.981148		
Variable	Communality	Weight
YrHgt	0.39028268	2.9980
FtFrBody	0.99999925	8594.3439
PrctFFB	0.52858921	10.6917
Frame	0.36752067	0.8589
BkFat	0.15969673	0.0080
SaleHt	0.53505948	4.0180
SaleWt	0.99999999	16850.6618

Here, the first factor might be SaleWt factor. The meaning of second factor is not very clear. Based on the communality result, the two factor model try to cover the FtFrBody and SaleWt, because the variance of these two variables is much larger than the rest variables. Factor analysis on the matrix **S** may not be a good approach.

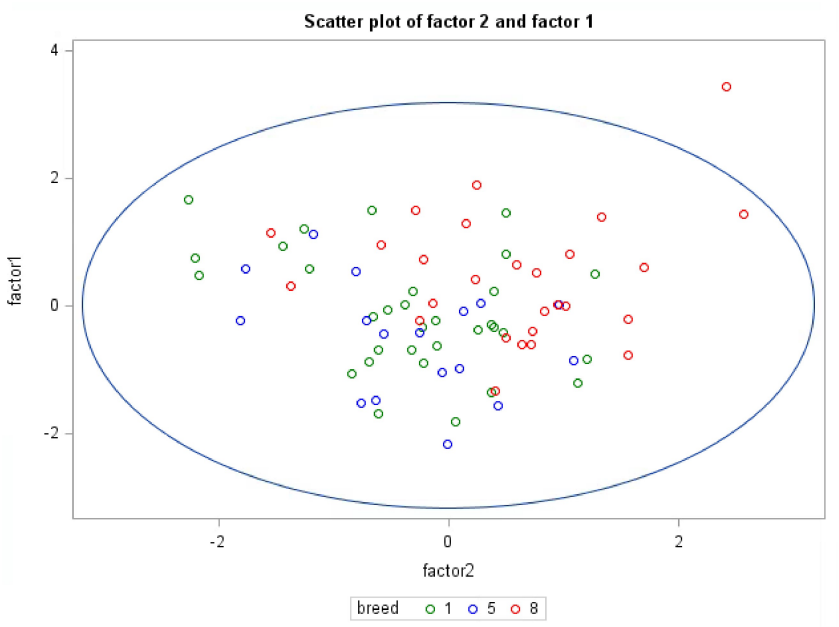
The factor analysis using m=2 (MLE) with rotation:

Rotated Factor Pattern		
	Factor1	Factor2
YrHgt	0.92547	0.37883
FtFrBody	0.28122	0.95964
PrctFFB	0.30694	0.63025
Frame	0.86139	0.37765
BkFat	-0.34109	-0.07551
SaleHt	0.71640	0.51872
SaleWt	0.18331	0.52474

Final Communality Estimates and Variable Weights		
Total Communality: Weighted = 12.811876 Unweighted = 4.589345		
Variable	Communality	Weight
YrHgt	1.00000000	Infty
FtFrBody	1.00000000	Infty
PrctFFB	0.49143386	1.96631258
Frame	0.88460518	8.66589992
BkFat	0.12204697	1.13901310
SaleHt	0.78230426	4.59356717
SaleWt	0.30895454	1.44708281

Here, the first factor might be YrHgt factor. The second factor is FtFrBody factor. Based on the communality result, the two factor model only covers the YrHgt and FtFrBody. Also the factor loading value from MLE method is different to the value from PCM method, this also indicates that the factor analysis (m=2 factors model) on matrix **S** is not good fit.

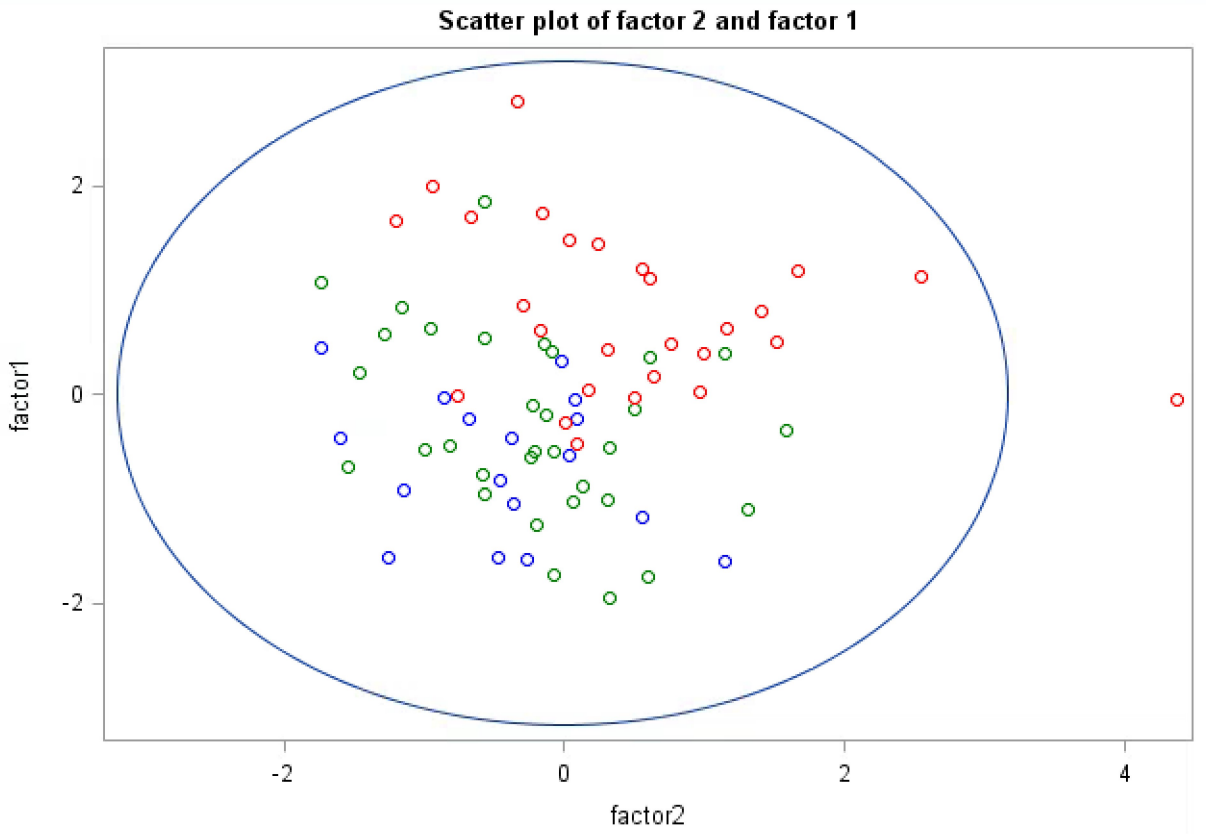
The factor score plot form PCM result:



Apparently, there is an outlier. It is observation no 51,

48	5	1000	47.7	913	68.2	5	0.15	49.4	1345	-1.56232	0.42841
49	5	975	47.2	844	70.6	5	0.15	50.1	1285	-2.16165	-0.01435
50	8	1750	54.0	1252	76.5	8	0.15	56.9	1648	1.43437	2.55687
51	8	1450	53.3	1383	81.4	8	0.20	59.6	1904	3.43792	2.41023
52	8	1200	52.8	1076	74.0	7	0.15	55.5	1615	0.63565	0.58579

The factor score plot form MLE result:

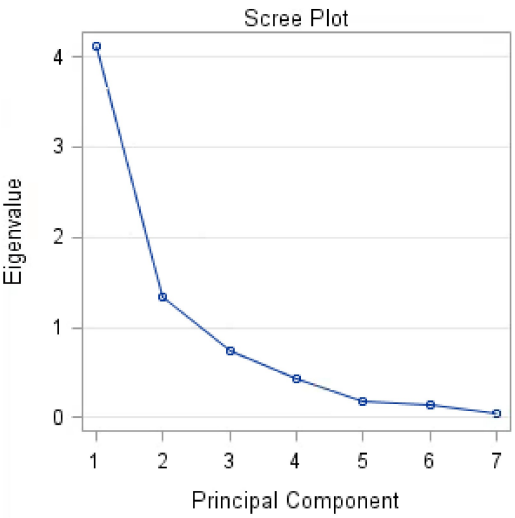


Apparently, there is an outlier. It is observation no 51,

48	5	1000	47.7	913	68.2	5	0.15	49.4	1345	-1.56771	-0.47296
49	5	975	47.2	844	70.6	5	0.15	50.1	1285	-1.56151	-1.25037
50	8	1750	54.0	1252	76.5	8	0.15	56.9	1648	1.12732	2.54779
51	8	1450	53.3	1383	81.4	8	0.20	59.6	1904	-0.05396	4.36645
52	8	1200	52.8	1076	74.0	7	0.15	55.5	1615	1.19656	0.54919
53	8	2000	53.5	1175	74.5	8	0.10	57.4	1686	1.17534	1.66821

For the factor analysis on matrix **R**:

First, we check the scree plot of matrix R:



Based on the scree plot, we will try m=2 and m=3 models.

For the model m=2:

PCM result

Factor Pattern		
	Factor1	Factor2
YrHgt	0.91334	-0.04948
FtFrBody	0.83700	0.15014
PrctFFB	0.72177	-0.36484
Frame	0.88091	0.00894
BkFat	-0.37900	0.82646
SaleHt	0.91927	0.11715
SaleWt	0.54798	0.69440

MLE result

Rotated Factor Pattern		
	Factor1	Factor2
YrHgt	0.92547	0.37883
FtFrBody	0.28122	0.95964
PrctFFB	0.30694	0.63025
Frame	0.86139	0.37765
BkFat	-0.34109	-0.07551
SaleHt	0.71640	0.51872
SaleWt	0.18331	0.52474

AIC value = 41.31

Chi-Square without Bartlett's Correction	57.314320
Akaike's Information Criterion	41.314320
Schwarz's Bayesian Criterion	22.668453
Tucker and Lewis's Reliability Coefficient	0.712184

For the model m=3:

PCM result

Factor Pattern			
	Factor1	Factor2	Factor3
YrHgt	0.91334	-0.04948	-0.35794
FtFrBody	0.83700	0.15014	0.38772
PrctFFB	0.72177	-0.36484	0.48930
Frame	0.88091	0.00894	-0.38949
BkFat	-0.37900	0.82646	-0.03335
SaleHt	0.91927	0.11715	-0.15210
SaleWt	0.54798	0.69440	0.21811

MLE result

Rotated Factor Pattern			
	Factor1	Factor2	Factor3
YrHgt	0.94416	0.28505	0.16526
FtFrBody	0.41180	0.50161	0.55676
PrctFFB	0.22931	0.94889	0.21686
Frame	0.88793	0.25083	0.18397
BkFat	-0.25673	-0.51439	0.27074
SaleHt	0.75319	0.26700	0.43735
SaleWt	0.25284	-0.05300	0.87631

AIC value = 2.659

Chi-Square without Bartlett's Correction	8.6588140
Akaike's Information Criterion	2.6588140
Schwarz's Bayesian Criterion	-4.3333860
Tucker and Lewis's Reliability Coefficient	0.9153059

Based on the AIC value, we choose the m = 3 model.

The interpretation of factors for the PCM result:

Factor Pattern			
	Factor1	Factor2	Factor3
YrHgt	0.91334	-0.04948	-0.35794
FtFrBody	0.83700	0.15014	0.38772
PrctFFB	0.72177	-0.36484	0.48930
Frame	0.88091	0.00894	-0.38949
BkFat	-0.37900	0.82646	-0.03335
SaleHt	0.91927	0.11715	-0.15210
SaleWt	0.54798	0.69440	0.21811

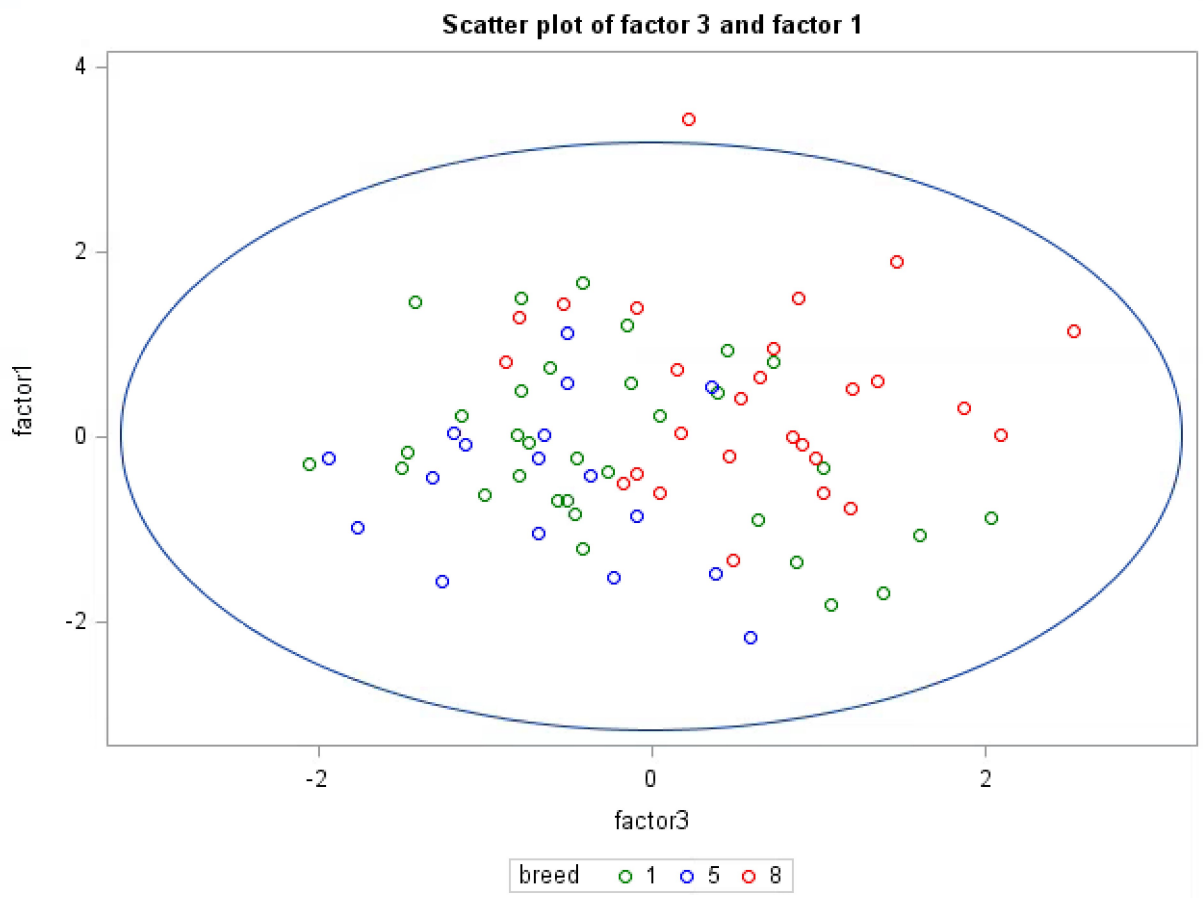
The first factor is the **body size factor**. The loading on YrHgt, FtFrBody, Frame and SaleHt are all high. The second factor is **back fat factor**. The loading on BkFat is very large. The third factor is not very clear, it reflect the difference between YrHgt/Frame/SaleHt and FtFrbody/PrctFFB/SaleWt. It might be viewed as **body configuration factor**,

The interpretation of factors for the PCM result:

Rotated Factor Pattern			
	Factor1	Factor2	Factor3
YrHgt	0.94416	0.28505	0.16526
FtFrBody	0.41180	0.50161	0.55676
PrctFFB	0.22931	0.94889	0.21686
Frame	0.88793	0.25083	0.18397
BkFat	-0.25673	-0.51439	0.27074
SaleHt	0.75319	0.26700	0.43735
SaleWt	0.25284	-0.05300	0.87631

The first factor is the **body height factor**. The loading on YrHgt, Frame and SaleHt are all high. The second factor is **back fat factor**. The loading on PrctFFB is very large, and the BkFat are negative. The third factor is **weight factor**, with large loading on SaleWt and FtFrBody.

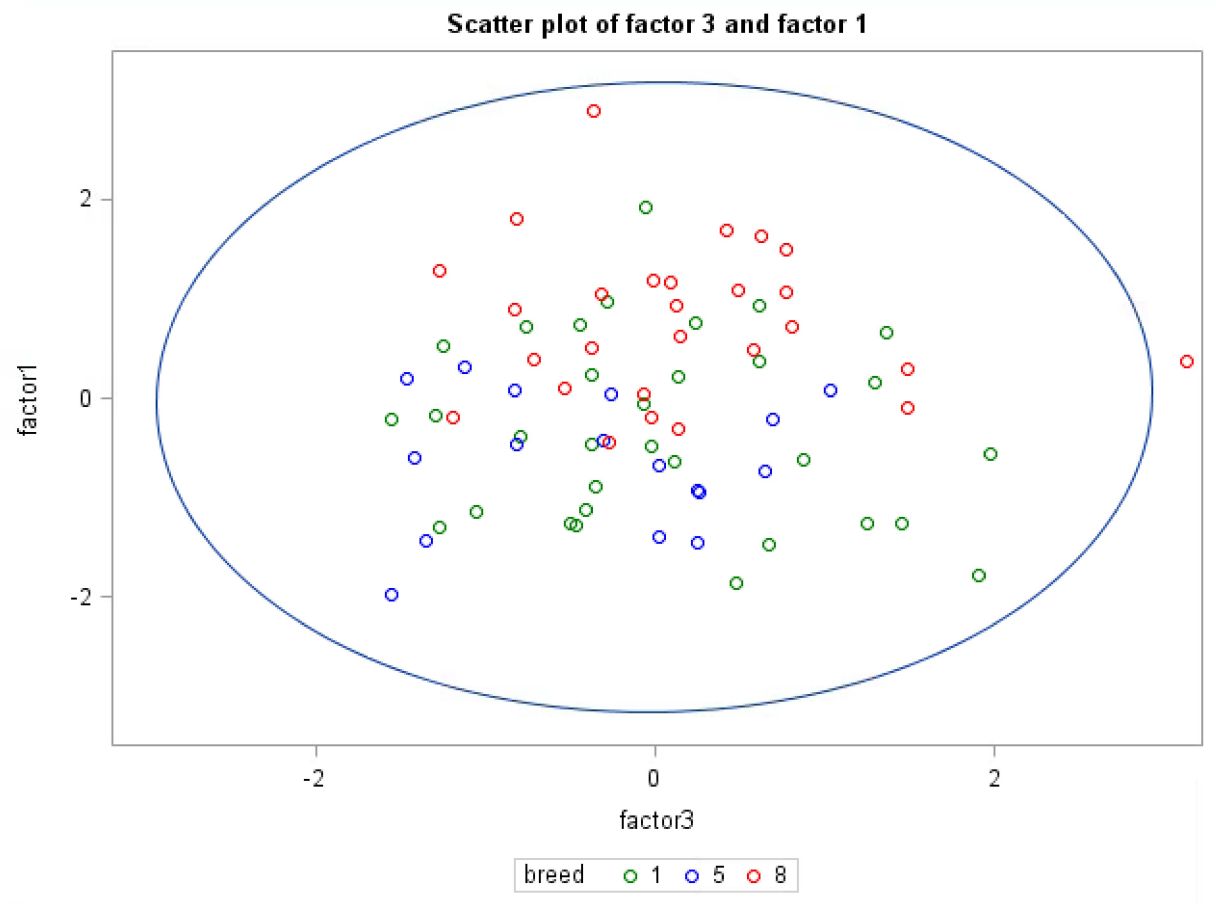
The factor3 and factor1 score plot form PCM result:



The outlier is No. 51.

51	8	1450	53.3	1383	81.4	8	0.20	59.6	1904	3.43792	2.41023	0.21942

The factor3 and factor1 score plot form MLE result:



The outlier is No. 51.

51	8	1450	53.3	1383	81.4	8	0.20	59.6	1904	0.37020	2.58421	3.13479

When we use MLE method, the factor analysis on R and S are same. However, if we use PCM, the factor analysis result on R is more reasonable, especially when the variance of some variables is much larger than the rest in the S matrix.