

多元分析：第十一周作业

蒋翌坤 20307100013

《实用多元统计分析》P364: 8.6

a

$$|S - \lambda| = \begin{vmatrix} 7476.45 - \lambda & 303.62 \\ 306.62 & 26.19 - \lambda \end{vmatrix} = 0 \Rightarrow \begin{matrix} \lambda_1 = 7489 & \lambda_2 = 13.84 \\ e_1 = (-0.999, -0.041) & e_2 = (0.041, -0.999) \end{matrix}$$

因此，主成分及对应方差为：

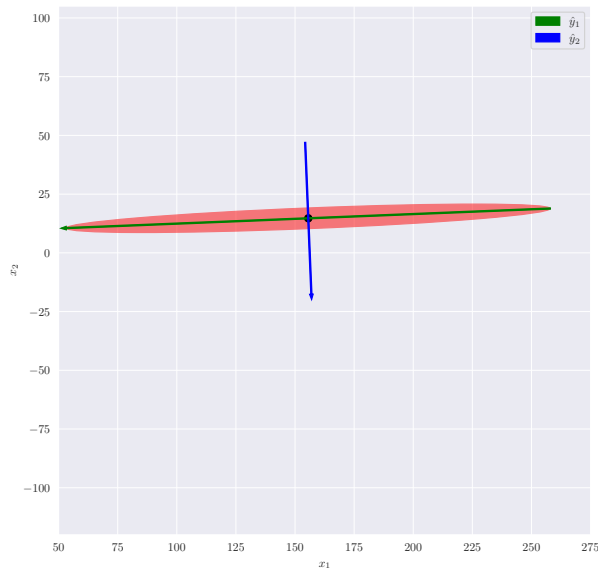
$$\hat{y} = \begin{bmatrix} -0.999x_1 + -0.041x_2 \\ 0.041x_1 + -0.999x_2 \end{bmatrix} \quad S_y = \begin{bmatrix} 7489 & 0 \\ 0 & 13.84 \end{bmatrix}$$

b

样本总方差用 \hat{y}_1 解释的比例为 $\frac{7489}{7489+13.84} = 99.82\%$

c

常数密度椭圆 $(x - \bar{x})' S^{-1} (x - \bar{x}) = 1.4$ 及主成分 \hat{y}_1, \hat{y}_2 的方向如下所示：



d

$$r_{\hat{y}_1, x_1} = \frac{0.999 \times \sqrt{7489}}{\sqrt{7476.45}} = -0.999$$

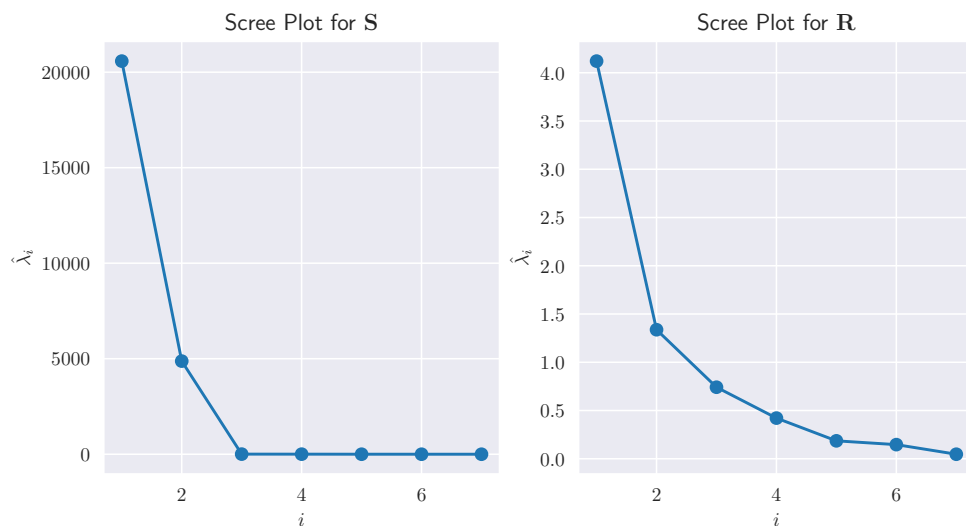
$$r_{\hat{y}_1, x_2} = \frac{-0.041 \times \sqrt{7489}}{\sqrt{26.19}} = -0.687$$

$r_{\hat{y}_1, x_1} \approx -1$ ，说明第一主成分几乎完全由 x_1 决定。

《实用多元统计分析》P370: 8.22

a

对于 S 与 R 作主成分分析，可得崖底碎石图如下所示：



从图中可以发现，两者的恰当成分个数都是 2，到第三个成分就出现拐点，特征值都很小且差不多。

b

样本主成分相当于是将原来 7 个变量对应的坐标点进行平移、旋转，使得特征向量对应的方向上可以获得最大方差

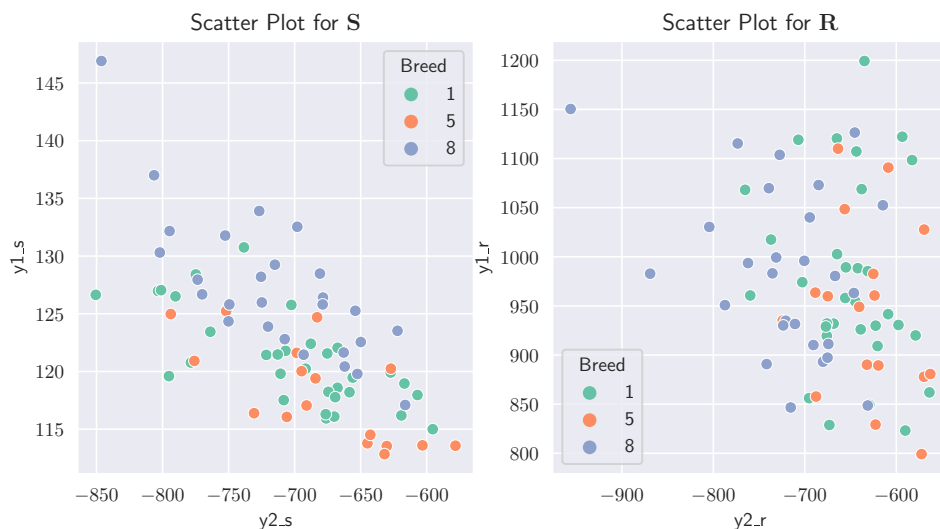
由于样本前 2 个主成分特征值很大，而后 5 个特征值很小，那么这前 2 个主成分就可以近似的表达这 7 个变量的信息。

c

由以上 7 个变量的数据可能是可以得到“身体尺寸”或“身体外形”的指标，因为从样本主成分来看，第一主成分对应的特征值特别大，说明第一主成分涵盖了这 7 个变量中的大多数信息，那么，第一主成分对应的 7 个变量的线性组合，就有可能描述了公牛的一个特征，比如“身体尺寸”或“身体外形”。

d

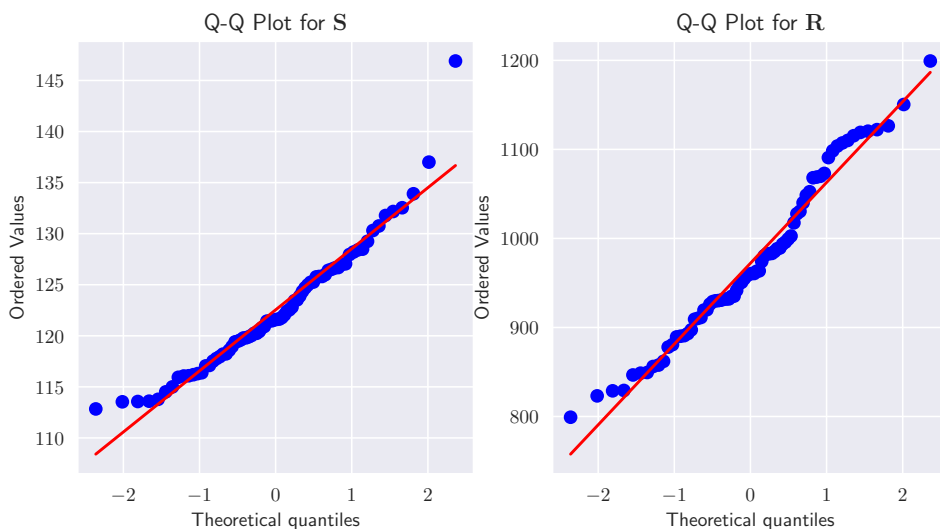
由 \hat{y}_1 和 \hat{y}_2 构成的数据点图如下所示:



从图中可以发现, 两者都不能区分三种牛。在第 8 种公牛中, 有一个离群值距离其他的点都很远。

e

用第一主成分可得如下 Q-Q 图:



从 Q-Q 图中可以看出, 无论是用 S 还是 R 得到的第一主成分, 都不服从正态分布。

附录

解答题目所使用的代码及输出请见:

<https://thisiskunmeng.github.io/multivariate/hw11.html>