多元回归 第七周作业

20307100013 蒋翌坤

《实用多元统计分析》P199: 5.3

解：(a) $T^2=\frac{(n-1)\left|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right|}{\left|\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right|}-(n-1)$

$$=\frac{3\times\left|\sum_{j=1}^{4}\left(x_j-[7\quad 11]'\right)\left(x_j-[7\quad 11]'\right)'\right|}{\left|\sum_{j=1}^{4}\left(x_j-[6\quad 10]'\right)\left(x_j-[6\quad 10]'\right)'\right|}-3$$

$$=\frac{3\times\left|\begin{bmatrix}25&-5\\-5&1\end{bmatrix}+\begin{bmatrix}1&-2\\-2&4\end{bmatrix}+\begin{bmatrix}1&2\\2&4\end{bmatrix}+\begin{bmatrix}1&-1\\-1&1\end{bmatrix}\right|}{\left|\begin{bmatrix}16&-8\\-8&4\end{bmatrix}+\begin{bmatrix}4&-2\\-2&1\end{bmatrix}+\begin{bmatrix}0&0\\0&1\end{bmatrix}+\begin{bmatrix}4&0\\0&0\end{bmatrix}\right|}-3$$

$$=\frac{3\times\left|\begin{bmatrix}28&-6\\-6&10\end{bmatrix}\right|}{\left|\begin{bmatrix}24&-10\\-10&6\end{bmatrix}\right|}-3=\frac{3\times(280-36)}{(144-100)}-3=\frac{150}{11}$$

(b) $\Lambda=\left(1+\frac{T^2}{n-1}\right)^{-\frac{n}{2}}=\left(1+\frac{150/11}{3}\right)^{-2}=0.0325$

《实用多元统计分析》P200: 5.11

解：

(a) 可以得到样本均值$\bar{X}=[5.19\quad 16.07]$，$S=\begin{bmatrix}176.00&287.24\\287.24&527.85\end{bmatrix}$

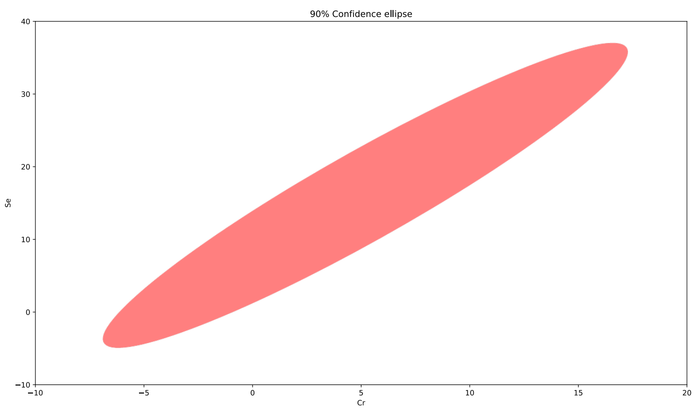
S对应的特征值及特征向量：$\lambda_1,\lambda_2=688.8,15.1$，$e_1,e_2=(-0.49,-0.87),(-0.87,0.49)$

于是可以得到 90%联合置信椭圆：中心位于$\bar{X}$，长、短半轴方向由e_1,e_2给出

长半轴长为$\sqrt{\lambda_1}\times\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(0.9)}=\sqrt{688.8\times\frac{2\times 8}{9\times 7}\times 3.26}=23.87$

短半轴长为$\sqrt{\lambda_2}\times\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(0.9)}=\sqrt{15.1\times\frac{2\times 8}{9\times 7}\times 3.26}=3.53$

椭圆图像如下图所示：



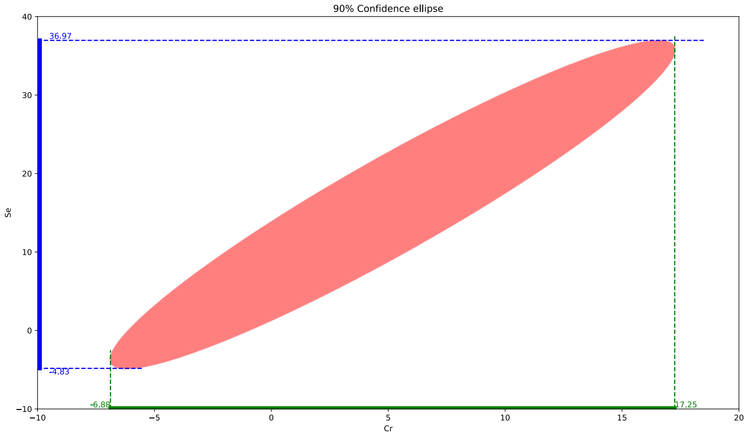
（b）

μ_1的置信区间为$\overline{x_1} \mp \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(0.9)s_{11}} = [-6.88, 17.25]$

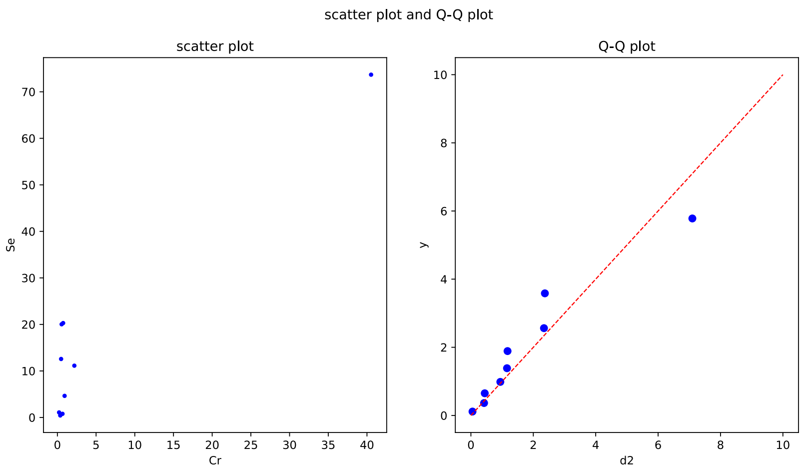
μ_2的置信区间为$\overline{x_2} \mp \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(0.9)s_{22}} = [-4.83, 36.97]$

由于 10 在μ_2的置信区间内，所以有迹象表明这种古秘鲁人头发的平均锶含量为 10

椭圆向坐标轴投影如下图所示：



（c）数据的散点图与 Q-Q 图如下所示：



可以发现 Q-Q 图中点与直线$y = x$距离很大，说明数据不显示出二元正态性。而（a）、（b）问中的结论都依赖数据有二元正态性的假设，这说明（a）、（b）问中得出的结论都不可靠。

（d）能够轻松发现(40.53, 73.68)为一组明显的离群观测值，重新推断可以得出以下结论：

样本均值$\bar{X} = [0.77 \quad 8.87]$，$S = \begin{bmatrix} 0.38 & 1.03 \\ 1.03 & 69.86 \end{bmatrix}$

S对应的特征值及特征向量：$\lambda_1, \lambda_2 = 69.88, 0.36$，$e_1, e_2 = (0.01, -0.99), (-0.99, 0.01)$

于是可以得到 90%联合置信椭圆：中心位于$\bar{X}$，长、短半轴方向由e_1, e_2给出

长半轴长为$\sqrt{\lambda_1} \times \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(0.9)} = \sqrt{69.88 \times \frac{2\times7}{8\times6} \times 3.46} = 8.40$

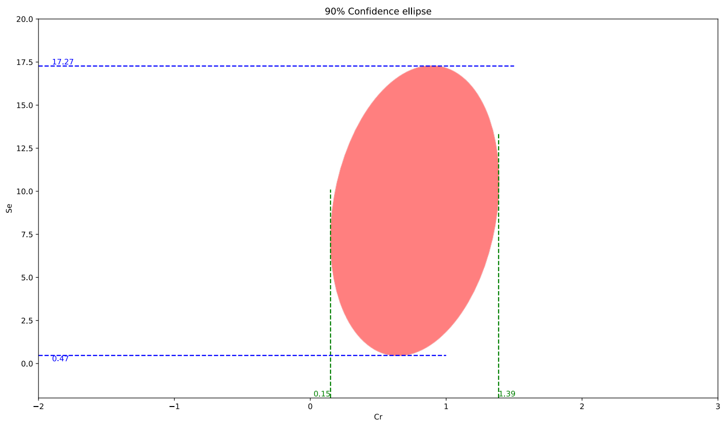
短半轴长为$\sqrt{\lambda_2} \times \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(0.9)} = \sqrt{0.36 \times \frac{2\times7}{8\times6} \times 3.46} = 0.61$

μ_1的置信区间为$\overline{x_1} \mp \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(0.9) s_{11}} = [0.15, 1.39]$

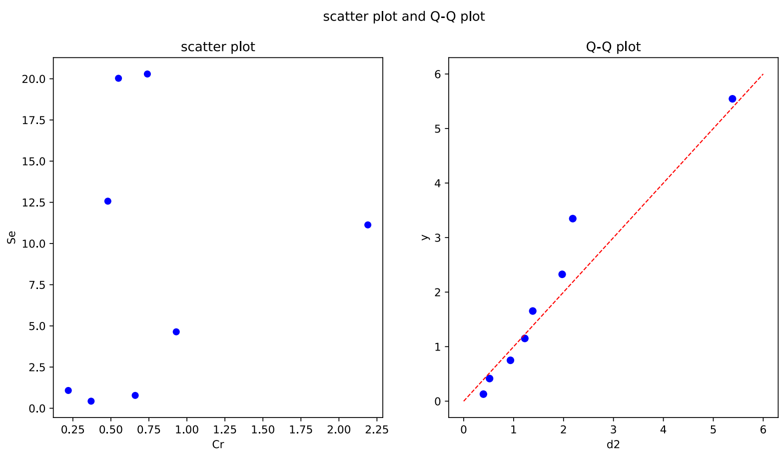
μ_2的置信区间为$\overline{x_2} \mp \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(0.9) s_{22}} = [0.47, 17.27]$

由于 10 在μ_2的置信区间内，所以有迹象表明这种古秘鲁人头发的平均锶含量为 10

椭圆图像以及向坐标轴投影如下图所示：



散点图以及 Q-Q 图如下图所示：



附录：

解答题目所使用的代码及输出请见：

https://thisiskunmeng.github.io/multivariate/hw7.html