作业 **2**

1. 随机向量 $\mathbf{x}=(x_1,...,x_n)^\top$, 假设随机变量 $x_1,...,x_n$ iid $\sim N(0,1)$, 我们称随机向量 $\mathbf{x}$ 服从 n 元标准正态分布，记作 $\mathbf{x}\sim N_n(0,I_n)$。假设 A 是 $n\times n$ 正交矩阵，证明 $\mathbf{y}=A\mathbf{x}\sim N_n(0,I_n)$.

2. 设 $\mathbf{a},\mathbf{b}\in R^n$ 为常数向量，假设 $\|\mathbf{a}\|=\|\mathbf{b}\|=1$, 且 $\mathbf{a}^\top\mathbf{b}=0$。假设 $n\times 1$ 随机向量 $\mathbf{x}\sim N_n(0,I_n)$，试证明

$$\sqrt{n-2}\times\frac{\mathbf{a}^\top\mathbf{x}}{\sqrt{\|\mathbf{x}\|^2-(\mathbf{a}^\top\mathbf{x})^2-(\mathbf{b}^\top\mathbf{x})^2}}\sim t_{n-2}.$$

3. 假设两组样本：

$$y_1,y_2,...,y_{n_0}\ \text{iid}\ \sim N(\mu_0,\sigma^2),$$
$$y_{n_0+1},y_{n_0+2},...,y_{n_0+n_1}\ \text{iid}\ \sim N(\mu_1,\sigma^2).$$

零假设 $H_0:\mu_1=\mu_0$ 的两样本 t-检验统计量为

$$t_1=\frac{\bar{y}_1-\bar{y}_0}{\sqrt{\left(\frac{1}{n_0}+\frac{1}{n_1}\right)s^2}},$$

其中 $\bar{y}_0=\frac{1}{n_0}\sum_{i=1}^{n_0}y_i,\bar{y}_1=\frac{1}{n_1}\sum_{i=n_0+1}^{n_0+n_1}y_i$, $s^2=\frac{1}{n_0+n_1-2}\left(\sum_{i=1}^{n_0}(y_i-\bar{y}_0)^2+\sum_{i=n_0+1}^{n_0+n_1}(y_i-\bar{y}_1)^2\right)$。
另外一方面，我们记 $x_i=0,1\le i\le n_0$ 为第一组的标号, $x_i=1,n_0+1\le i\le n_0+n_1$ 为第二组的标号。记 r 为 $(x_i,y_i),i=1,...,n=n_0+n_1$ 的样本相关系数，x,y 的相关性检验统计量

$$t_2=\sqrt{n-2}\frac{r}{\sqrt{1-r^2}},$$

试验证 $t_1=t_2$。

4. 假设有二元数据 (x_i,y_i), $i=1,...,n$ iid，其中 x_i,y_i 各分别仅取 0,1 值, 其样本相关系数记为 r。另一方面，这种二元属性 (离散) 数据常常以列联表的形式表示:

$$\begin{array}{cc|c|c|c}
 & & \multicolumn{2}{c}{y} & \\
 & & 1 & 0 & \\
\hline
x & 1 & a & b & n_1 \\
\hline
 & 0 & c & d & n_0 \\
\hline
 & & m_1 & m_0 & n
\end{array}$$

其中 a 为 $x_i=y_i=1$ 的样本个数 (即 $a=\sum x_iy_i,c=\sum(1-x_i)y_i$ 等等), $n_1=a+b,n_0=c+d,m_1=a+c,m_0=b+d$。该列联表的独立性检验的 Pearson 卡方统计量为

$$X^2=\frac{n(ad-bc)^2}{n_0n_1m_0m_1}.$$

试验证等于 $X^2=nr^2$。