1. 考虑$y = x^t\beta + \varepsilon, \varepsilon \sim N(0,\sigma^2)$线性模型，x 为 p 维已知列向量，y 为观测值标量，β为 p 维系数列向量，ε为随机误差项。请说明极大似然估计与最小二乘法，在求解该模型β时是等价的。

由于题目没有说明所以此处不考虑异方差、内生性问题等带来的参数估计问题，仅针对一般情况下的 OLS 和极大似然估计进行讨论。

首先使用 OLS（最小二乘法）求解该模型

$$y = x^t\beta + \varepsilon = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

已知残差平方和表述如下：

$$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}\right)^2$$

最小二乘法使残差平方和最小，分别对每个估计项求导得到如下一阶条件

$$\sum_{i=1}^{n} -2 * \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}\right) = 0$$

$$\sum_{i=1}^{n} -2 * x_{i1} * (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0$$

$$\cdots$$

$$\sum_{i=1}^{n} -2 * x_{ik} * (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0$$

求解以上方程得到我们要的系数

结果如下：

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$$

式中的$\hat{r}_{i1}$表示用x_1对剩下的变量做回归得到的残差。即这个过程可以描述为，先拿x_1对剩下的变量做回归得到的残差$\hat{r}_{i1}$，再用 y 对$\hat{r}_{i1}$做回归，得到$\hat{\beta}_1$。对于除了$\hat{\beta}_0$之外的其他变量都有这种表述，而$\hat{\beta}_0$则通过式$\bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2 - \cdots$得到。

其实关于这些变量的具体表述和这一题的证明无关，但是和下一题的证明相关，这

里顺便先解出。

然后使用极大似然估计来解这个模型。

极大似然估计是指在给定样本的情况下估计模型，使得在这个模型下抽取到指定样本的概率是最大的。先得到目标函数 L。根据定义

$$\mathrm{L}=\prod P(Y|\vec{X},\vec{\beta})$$

这样考虑这个模型，对于给定的样本 i 有：

$$\mathrm{y_i}=\mathrm{x_i^t}\beta+\varepsilon=\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik}+\varepsilon_\mathrm{i}$$

所以，对于那个样本的残差（定值）有

$$\varepsilon_\mathrm{i}=\mathrm{y_i}-\beta_0-\beta_1 x_{i1}-\cdots-\beta_k x_{ik}$$

而残差是个服从正态分布的随机变量：$\varepsilon\sim N(0,\sigma^2)$，不影响结论的可以设置 L 为（实际上和标准正态分布差个σ^2但不影响结论）

$$\mathrm{L}=\prod \phi(\mathrm{y_i}-\beta_0-\beta_1 x_{i1}-\cdots-\beta_k x_{ik})$$

根据正态分布函数的概率密度函数，我们有

$$\mathrm{lnL}=\mathrm{C}\sum_{i=1}^{n}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\cdots-\hat{\beta}_k x_{ik}\right)^2$$

C 是一个常数，这里就不具体算了。

根据极大似然条件得到：

$$\frac{\partial \mathrm{lnL}}{\partial \mathrm{\beta_i}}=0$$

即

$$\sum_{i=1}^{n}-2*\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\cdots-\hat{\beta}_k x_{ik}\right)=0$$

$$\sum_{i=1}^{n}-2*x_{i1}*(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\cdots-\hat{\beta}_k x_{ik})=0$$

$$\cdots$$

$$\sum_{i=1}^{n}-2*x_{ik}*(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\cdots-\hat{\beta}_k x_{ik})=0$$

可以得到，这个条件和上面的是相同的，所以在这里极大似然估计和最小二乘法是等价的。

2. 单变量回归：对于 x 向量某一元x_i，我们可以考虑模型$y_i = x_i\beta_i^* + \varepsilon, \varepsilon \sim N(0, \sigma^2)$线性模型。$\beta_i^*$为此模型通过极大似然或者最小二乘法求解时，得到的系数估计。请说明β_i^*与β_i的区别是什么？可以从理论推导或者实验模拟角度。

正如上面推导的

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 - \cdots$$

所以$\hat{\beta}_1$是x_1对其他变量做回归得到残差再对 y 做的回归得到的系数，即，$\hat{\beta}_1$是x_1对 y“纯”的影响。而β_1^*则对应于一般的一元线性回归模型，即直接使用x_1对 y 做回归。所以在 X 向量的各维都正交的情况下（如通过 PCA 得到的特征空间），这两种方法是等价的，否则两者是不相同的。使用多元线性回归得到的是x_1对 y“纯”的影响，而单变量回归的得到的是包含其他与x_1相关的变量对 y 的综合影响。