回归分析 (01714601)
课程主页：http://staff.ustc.edu.cn/~ynyang/lm2020

第十三讲 最小二乘估计（续）

2020.4.1

$$\hat{\boldsymbol{\beta}}_J = (X_J^{\perp\top} X_J^{\perp})^{-1} X_J^{\perp\top} \mathbf{y} \overset{0 \notin J}{=} \hat{\Sigma}_{JJ\bullet -J}^{-1} \hat{\Sigma}_{Jy\bullet -J}$$

中心化

矩阵中心化：数据矩阵$Z_{n\times m}=\begin{pmatrix}\mathbf{x}_1{}^{\top}\\ \vdots \\ \mathbf{x}_n{}^{\top}\end{pmatrix}$，样本均值$\bar{\mathbf{x}}=Z^{\top}\mathbf{1}/n$,

Z的中心化矩阵为

$$Z_{(c)}=\begin{pmatrix}(\mathbf{x}_1-\bar{\mathbf{x}})^{\top}\\ \vdots \\ (\mathbf{x}_n-\bar{\mathbf{x}})^{\top}\end{pmatrix}=Z-\mathbf{1}\bar{\mathbf{x}}^{\top}=(I_n-P_{\mathbf{1}})Z$$

其中$P_{\mathbf{1}}=\mathbf{1}\mathbf{1}^{\top}/n=\mathbf{1}(\mathbf{1}^{\top}\mathbf{1})^{-1}\mathbf{1}^{\top}$，样本协方差矩阵为

$$\hat{\Sigma}=\frac{1}{n-1}Z_{(c)}{}^{\top}Z_{(c)}=\frac{1}{n-1}Z^{\top}(I_n-P_{\mathbf{1}})Z.$$

在带截距的模型$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + \mathbf{x}_{(1)}\beta_1, \ldots, \mathbf{x}_{(p-1)}\beta_{p-1} + \boldsymbol{\varepsilon}$中

我们将截距项独立出来，为此划分

$$X_{n\times p} = (\mathbf{1}, \mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(p-1)}) = (\mathbf{1}, Z_{n\times(p-1)}),\ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\gamma} \end{pmatrix},\ \boldsymbol{\gamma}_{(p-1)\times 1} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

其中Z的列为$p-1$个自变量的n次观察。

模型为 $$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = (\mathbf{1}, Z)\begin{pmatrix} \beta_0 \\ \boldsymbol{\gamma} \end{pmatrix} = \mathbf{1}\beta_0 + Z\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

- $\hat{\Sigma}_{\mathbf{xx}} = Z_{(c)}{}^{\mathsf{T}} Z_{(c)}/(n-1)$：为自变量的样本协方差阵；
- $\hat{\Sigma}_{\mathbf{x}y} = Z_{(c)}{}^{\mathsf{T}} \mathbf{y}/(n-1)$：为自变量与响应变量的样本协方差阵。

回忆：总体参数含义（第五讲命题3）：

记号：$\mathbf{x}=(x_1,\ldots,x_{p\text{-}1})^{\mathsf{T}}$，$\mathbf{x}_{-k}=(x_1,\ldots,x_{k-1,}x_{k+1,}\ldots,x_{p\text{-}1})^{\mathsf{T}}$.

(第五讲)命题3. 假设线性回归模型：

$$y=\alpha+\boldsymbol{\beta}^{\mathsf{T}}\mathbf{x}+\varepsilon,\quad \varepsilon\sim(0,\sigma^2)\text{与}\mathbf{x}\text{独立}$$

其中$\boldsymbol{\beta}=(\beta_1,\ldots,\beta_{p\text{-}1})^{\mathsf{T}}$. 记$\mu_y=E(y)$，$\boldsymbol{\mu}_x=E(\mathbf{x})$，则

(1) $\boldsymbol{\beta}=\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}$，$\alpha=\mu_y-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_x$

从而$\varepsilon=(y-\mu_y)-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)$是$y$关于$\mathbf{x}$的不相关化。

(2) $\beta_k=\Sigma_{x_kx_k\bullet\mathbf{x}_{-k}}^{-1}\Sigma_{x_ky\bullet\mathbf{x}_{-k}}$

(3) $\sigma^2=\Sigma_{yy\bullet\mathbf{x}}=\Sigma_{yy}-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}$

$\Sigma_{\mathbf{uv}}=\mathrm{cov}(\mathbf{u},\mathbf{v})$

$y^{\perp}=y-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$

$\Sigma_{\mathbf{uv}\bullet\mathbf{w}}=$

$\Sigma_{\mathbf{uv}}-\Sigma_{\mathbf{uw}}\Sigma_{\mathbf{ww}}^{-1}\Sigma_{\mathbf{wv}}$

由上节课命题1，我们可知LS估计与总体参数含义一致（推论1：参数表达中把协方差换成样本协方差，即得LS估计）

设$J \subset \{0,\ldots,p-1\}$为任何一个非空下标集合

样本协方差阵为$\hat{\Sigma}$，记

$$\hat{\Sigma}_{JJ} = \hat{\Sigma}_{\mathbf{x}_J \mathbf{x}_J}, \quad \hat{\Sigma}_{J,-J} = \hat{\Sigma}_{\mathbf{x}_J \mathbf{x}_{-J}}, \quad \hat{\Sigma}_{-J,-J} = \hat{\Sigma}_{\mathbf{x}_{-J} \mathbf{x}_{-J}}, \quad \hat{\Sigma}_{Jy} = \hat{\Sigma}_{\mathbf{x}_J y}.$$

$$\hat{\Sigma}_{JJ\bullet -J} = \hat{\Sigma}_{JJ} - \hat{\Sigma}_{J,-J} \left(\hat{\Sigma}_{-J,-J} \right)^{-1} \hat{\Sigma}_{-J,J}^{-1},$$

$$\hat{\Sigma}_{Jy\bullet -J} = \hat{\Sigma}_{Jy} - \hat{\Sigma}_{J,-J} \left(\hat{\Sigma}_{-J,-J} \right)^{-1} \hat{\Sigma}_{-J,y}.$$

推论1. 对于模型 $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + Z\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim (0, \sigma^2)$

记 $\bar{\mathbf{x}} = Z^{\mathsf{T}}\mathbf{1}/n$，$Z$的中心化矩阵$Z_{(c)} = Z - \mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}}$，则

(1) $\hat{\boldsymbol{\gamma}} = (Z_{(c)}{}^{\mathsf{T}} Z_{(c)})^{-1} Z_{(c)}{}^{\mathsf{T}} \mathbf{y} = \hat{\Sigma}_{\mathbf{xx}}^{-1} \hat{\Sigma}_{\mathbf{x}y}, \quad \hat{\beta}_0 = \bar{y} - \bar{\mathbf{x}}^{\mathsf{T}} \hat{\boldsymbol{\gamma}}.$

$$\operatorname{var}(\hat{\boldsymbol{\gamma}} \mid X) = \sigma^2 (Z_{(c)}{}^{\mathsf{T}} Z_{(c)})^{-1} = \frac{\sigma^2}{n-1} \hat{\Sigma}_{\mathbf{xx}}^{-1}.$$

(2) $\hat{\sigma}^2 = \dfrac{1}{n-p} RSS = \dfrac{n-1}{n-p} \hat{\Sigma}_{yy\bullet\mathbf{x}}$

(3) 若$0 \notin J$，则$\hat{\boldsymbol{\beta}}_J = \hat{\Sigma}_{JJ\bullet-J}^{-1} \hat{\Sigma}_{Jy\bullet-J}, \operatorname{var}(\hat{\boldsymbol{\beta}}_J \mid X) = \sigma^2 \hat{\Sigma}_{JJ\bullet-J}^{-1} / (n-1).$

特别地，$k \geq 1$，$\hat{\beta}_k = \hat{\Sigma}_{kk\bullet-k}^{-1} \hat{\Sigma}_{ky\bullet-k}$，$\operatorname{var}(\hat{\beta}_k \mid X) = \dfrac{\sigma^2}{(n-1)\hat{\Sigma}_{kk\bullet-k}}.$

$$\hat{\Sigma}_{JJ\bullet-J} = \hat{\Sigma}_{\mathbf{x}_J\mathbf{x}_J\bullet\mathbf{x}_{-J}} = \hat{\Sigma}_{\mathbf{x}_J\mathbf{x}_J} - \hat{\Sigma}_{\mathbf{x}_J\mathbf{x}_{-J}} \left(\hat{\Sigma}_{\mathbf{x}_{-J}\mathbf{x}_{-J}}\right)^{-1} \hat{\Sigma}_{\mathbf{x}_{-J}\mathbf{x}_J},$$

$$\hat{\Sigma}_{Jy\bullet-J} = \hat{\Sigma}_{\mathbf{x}_J y\bullet\mathbf{x}_{-J}} = \hat{\Sigma}_{\mathbf{x}_J y} - \hat{\Sigma}_{\mathbf{x}_J\mathbf{x}_{-J}} \left(\hat{\Sigma}_{\mathbf{x}_{-J}\mathbf{x}_{-J}}\right)^{-1} \hat{\Sigma}_{\mathbf{x}_{-J},y}.$$

$$\hat{\Sigma}_{kk\bullet-k} = \hat{\Sigma}_{\mathbf{x}_k\mathbf{x}_k\bullet\mathbf{x}_{-k}}, \quad \hat{\Sigma}_{ky\bullet-k} = \hat{\Sigma}_{\mathbf{x}_k y\bullet\mathbf{x}_{-k}}.$$

注1: 截距项的LS估计与样本均值有关。 除了截距项之外的参数LS估计由样本协方差阵$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{\mathbf{xx}} & \hat{\Sigma}_{\mathbf{x}y} \\ \hat{\Sigma}_{y\mathbf{x}} & \hat{\Sigma}_{yy} \end{pmatrix}$完全决定。

事实上，正态假设下,$(\bar{y}, \bar{\mathbf{x}}, \hat{\Sigma})$是充分统计量。

注2: $\hat{\beta}_k = \hat{\Sigma}_{kk\bullet-k}^{-1} \hat{\Sigma}_{ky\bullet-k} \propto \dfrac{\hat{\Sigma}_{ky\bullet-k}}{\hat{\Sigma}_{kk\bullet-k}^{1/2} \hat{\Sigma}_{yy\bullet-k}^{1/2}} = r_{ky\bullet-k}$偏相关系数。

注3. 注意到 $\hat{\mathbf{x}}_{(k)} = \mathrm{P}_{X_{(-k)}}\mathbf{x}_{(k)}$为$\mathbf{x}_{(k)}$在其它自变量$X_{(-\mathrm{k})}$生成空间上的投影（拟合值），所以

$\mathbf{x}_{(k)}^{\perp} = \mathbf{x}_{(k)} - \hat{\mathbf{x}}_{(k)}$为“$\mathbf{x}_{(k)}$对设计阵其它列回归后的残差”，

而$\hat{\beta}_k = ({\mathbf{x}_{(k)}^{\perp}}^{\mathsf{T}}\mathbf{x}_{(k)}^{\perp})^{-1}{\mathbf{x}_{(k)}^{\perp}}^{\mathsf{T}}\mathbf{y}$，因此$\hat{\beta}_k$可由如下两步求得：

(1) $\mathbf{x}_{(k)} \sim \mathbf{x}_{(0)} + \mathbf{x}_{(1)} + \ldots + \mathbf{x}_{(k-1)} + \mathbf{x}_{(k+1)} + \ldots + \mathbf{x}_{(p-1)}$，得到残差$\mathbf{x}_{(k)}^{\perp}$

(2) $\mathbf{y} \sim \mathbf{x}_{(k)}^{\perp}$，得到的斜率即$\hat{\beta}_k$,

证明:(1). 可由命题1立得,或在模型中将$\mathbf{1}, Z$正交化:

$$\mathbf{y} = \mathbf{1}\beta_0 + Z\boldsymbol{\gamma} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + (Z - \mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}} + \mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}})\boldsymbol{\gamma} + \boldsymbol{\varepsilon} =$$

$$= \mathbf{1}(\beta_0 + \bar{\mathbf{x}}^{\mathsf{T}}\boldsymbol{\gamma}) + (Z - \mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}})\boldsymbol{\gamma} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0^{\ *} + Z_{(c)}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

则$\hat{\boldsymbol{\gamma}} = (Z_{(c)}{}^{\mathsf{T}} Z_{(c)})^{-1} Z_{(c)}{}^{\mathsf{T}}\mathbf{y}$, $\hat{\beta}_0^{\ *} = \bar{y} \Rightarrow \hat{\beta}_0 = \bar{y} - \bar{\mathbf{x}}^{\mathsf{T}}\hat{\boldsymbol{\gamma}}$.

另外,$Z_{(c)}{}^{\mathsf{T}} Z_{(c)} = (n-1)\hat{\Sigma}_{\mathbf{xx}}$, $Z_{(c)}{}^{\mathsf{T}}\mathbf{y} = (n-1)\hat{\Sigma}_{\mathbf{x}y}$

$$\Rightarrow \hat{\boldsymbol{\gamma}} = (Z_{(c)}{}^{\mathsf{T}} Z_{(c)})^{-1} Z_{(c)}{}^{\mathsf{T}}\mathbf{y} = \hat{\Sigma}_{\mathbf{xx}}^{-1}\hat{\Sigma}_{\mathbf{x}y}$$

$$(2)\quad RSS = \|\mathbf{y}-\hat{\mathbf{y}}\|^2 = \|\mathbf{y}-\mathbf{1}\bar{y}-Z_{(c)}\hat{\boldsymbol{\gamma}}\|^2 = \|\mathbf{y}-\mathbf{1}\bar{y}\|^2 - \|Z_{(c)}\hat{\boldsymbol{\gamma}}\|^2$$

$$=(n-1)\hat{\Sigma}_{yy}-\hat{\boldsymbol{\gamma}}^{\mathsf{T}}Z_{(c)}{}^{\mathsf{T}}Z_{(c)}\hat{\boldsymbol{\gamma}}$$

$$=(n-1)\hat{\Sigma}_{yy}-(n-1)\hat{\boldsymbol{\gamma}}^{\mathsf{T}}\hat{\Sigma}_{\mathbf{xx}}\hat{\boldsymbol{\gamma}}$$

因为$\hat{\Sigma}_{\mathbf{xx}} = Z_{(c)}{}^{T}Z_{(c)}/(n-1)$

$$=(n-1)\hat{\Sigma}_{yy}-(n-1)\hat{\Sigma}_{y\mathbf{x}}\hat{\Sigma}_{\mathbf{xx}}^{-1}\hat{\Sigma}_{\mathbf{x}y}=(n-1)\hat{\Sigma}_{yy\bullet\mathbf{x}}$$

$\hat{\boldsymbol{\gamma}}=\hat{\Sigma}_{\mathbf{xx}}^{-1}\hat{\Sigma}_{\mathbf{x}y}$

所以$\hat{\sigma}^2=\dfrac{1}{n-p}RSS=\dfrac{n-1}{n-p}\hat{\Sigma}_{yy\bullet\mathbf{x}}$

(3) 当$0 \notin J, \hat{\boldsymbol{\beta}}_J = \hat{\boldsymbol{\gamma}}_J$. 下面记$W = Z_{(c)}$，$\hat{\boldsymbol{\gamma}} = (W^{\top}W)^{-1}W^{\top}\mathbf{y}$,

由命题1，$\hat{\boldsymbol{\gamma}}_J = (W_J^{\perp\top}W_J^{\perp})^{-1}W_J^{\perp\top}\mathbf{y}$，其中$W_J^{\perp} = W_J - P_{W_{-J}}W_J$.

$$W_J^{\perp\top}W_J^{\perp} = \left(W_J - P_{W_{-J}}W_J\right)^{\top}\left(W_J - P_{W_{-J}}W_J\right) = W_J^{\top}W_J - W_J^{\top}P_{W_{-J}}W_J$$

$$= W_J^{\top}W_J - W_J^{\top}\left(W_{-J}(W_{-J}^{\top}W_{-J})^{-1}W_{-J}^{\top}\right)W_J$$

$$= (n-1)\hat{\Sigma}_{JJ} - (n-1)\hat{\Sigma}_{J,-J}\left(\hat{\Sigma}_{-J,-J}\right)^{-1}\hat{\Sigma}_{-J,J} = (n-1)\hat{\Sigma}_{JJ\bullet-J}$$

类似地，$W_J^{\perp\top}\mathbf{y} = (n-1)\hat{\Sigma}_{Jy\bullet-J}$

所以 $\hat{\boldsymbol{\beta}}_J = \hat{\Sigma}_{JJ\bullet-J}^{-1}\hat{\Sigma}_{Jy\bullet-J}$，$\mathrm{var}(\hat{\boldsymbol{\beta}}_J \mid X) = \dfrac{\sigma^2}{(n-1)\hat{\Sigma}_{JJ\bullet-J}}$

拟合优度: 复相关系数平方R²

模型： $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + Z\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim (0, \sigma^2 I_n)$

平方和分解：

$$\boxed{\begin{array}{ccccc} \|\mathbf{y}-\mathbf{1}\bar{y}\|^2 & = & \|\mathbf{y}-\hat{\mathbf{y}}\|^2 & + & \|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2 \\ \mathrm{SS}_{总} & = & RSS & + & \mathrm{SS}_{回} \end{array}}$$

$\mathbf{1} \in L(X)$

$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \perp \hat{\mathbf{y}} - \mathbf{1}\bar{y}$

$\hat{\mathbf{y}}$的样本均值

$\bar{\hat{y}} = \mathbf{1}^{\mathrm{T}}\hat{\mathbf{y}}/n = \mathbf{1}^{\mathrm{T}}\mathrm{P}_X\mathbf{y}/n$

$= \mathbf{1}^{\mathrm{T}}\mathbf{y}/n = \bar{y}.$

y的拟合值向量（**y**在$L(X)$ 上的投影）的样本方差占**y**的样本方差的百分比定义为R^2(复相关系数平方)。

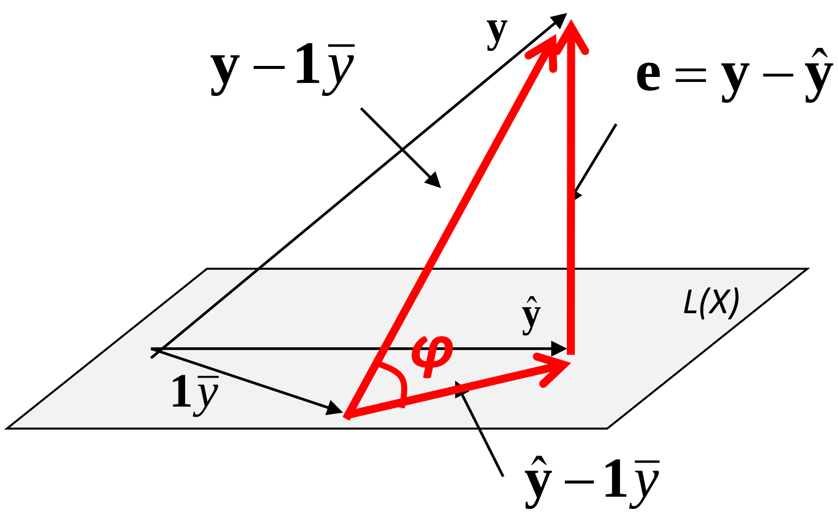
定义：复相关系数平方 $R^2 = \dfrac{\mathrm{SS}_{回}}{\mathrm{SS}_{总}} = \dfrac{\|\hat{\mathbf{y}}-\mathbf{1}\bar{\mathrm{y}}\|^2}{\|\mathbf{y}-\mathbf{1}\bar{\mathrm{y}}\|^2}$，$0 \le R^2 \le 1$,

度量**y**与X的相关性，

度量以X拟合**y**的好坏程度(拟合优度goodness of fit).

图示：总平方和分解为红色直角三角形的勾股定理：

$$\mathbf{y} = \hat{\mathbf{y}} \oplus \mathbf{e}，\ \mathbf{e} \perp \hat{\mathbf{y}}.$$



对于中心化了的红色三角形：

$$\mathbf{y}-\mathbf{1}\bar{y} = (\hat{\mathbf{y}}-\mathbf{1}\bar{y}) \oplus \mathbf{e}，\ \mathbf{e} \perp \hat{\mathbf{y}}-\mathbf{1}\bar{y}.$$

$$\begin{aligned} \|\mathbf{y}-\mathbf{1}\bar{y}\|^2 &= \|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2 + \|\mathbf{e}\|^2 \\ SS_{总} &= SS_{回} + RSS \end{aligned}$$

$$R^2 = \frac{SS_{回}}{SS_{总}} = \frac{\|\hat{\mathbf{y}}-\mathbf{1}\bar{\mathrm{y}}\|^2}{\|\mathbf{y}-\mathbf{1}\bar{\mathrm{y}}\|^2} = \cos^2(\varphi) = \begin{cases} (r_{\hat{\mathbf{y}},\mathbf{y}})^2，\text{几何意义} \\ \max\limits_{\mathbf{u}\in L(X)} (r_{\mathbf{u},\mathbf{y}})^2，\text{最小二乘} \end{cases}$$

命题2. 复相关系数平方$R^2=[r_{\mathbf{y}\hat{\mathbf{y}}}]^2=\max\limits_{\mathbf{u}\in L(X)}(r_{\mathbf{u},\mathbf{y}})^2$，

最大值在$\mathbf{u}\propto\hat{\mathbf{y}}=X\widehat{\boldsymbol{\beta}}$处达到，其中$\widehat{\boldsymbol{\beta}}=(X^{\top}X)^{-1}X^{\top}\mathbf{y}$。

证明：$[r_{Y\hat{Y}}]^2=\dfrac{\left((\mathbf{y}-\mathbf{1}\overline{y})^{\top}(\hat{\mathbf{y}}-\mathbf{1}\overline{y})\right)^2}{\|\mathbf{y}-\mathbf{1}\overline{y}\|^2\|\hat{\mathbf{y}}-\mathbf{1}\overline{y}\|^2}.$

因为$(\hat{\mathbf{y}}-\mathbf{1}\overline{y})\in L(X)$, 所以$\mathbf{e}^{\top}(\hat{\mathbf{y}}-\mathbf{1}\overline{y})=0$,

$$(\mathbf{y}-\mathbf{1}\overline{y})^{\top}(\hat{\mathbf{y}}-\mathbf{1}\overline{y})=(\hat{\mathbf{y}}+\mathbf{e}-\mathbf{1}\overline{y})^{\top}(\hat{\mathbf{y}}-\mathbf{1}\overline{y})$$
$$=(\hat{\mathbf{y}}-\mathbf{1}\overline{y})^{\top}(\hat{\mathbf{y}}-\mathbf{1}\overline{y})=\|\hat{\mathbf{y}}-\mathbf{1}\overline{y}\|^2$$

$$\Rightarrow[r_{\hat{\mathbf{y}}\mathbf{y}}]^2=\frac{\|\hat{\mathbf{y}}-\mathbf{1}\overline{y}\|^2}{\|\mathbf{y}-\mathbf{1}\overline{y}\|^2}=R^2$$

下面证明$R^2 = \max_{\mathbf{u}\in L(X)} (r_{\mathbf{u},\mathbf{Y}})^2$：记$\mathbf{y}_{(c)} = \mathbf{y} - P_{\mathbf{1}}\mathbf{y} = \mathbf{y} - \mathbf{1}\bar{y}$, $\quad \hat{\mathbf{y}}_{(c)} = \hat{\mathbf{y}} - \mathbf{1}\bar{y}$.

对任何$\mathbf{u} = X\boldsymbol{\beta} = \mathbf{1}\beta_0 + Z\boldsymbol{\gamma} \in L(X)$, 令 $\mathbf{u}_{(c)} = \mathbf{u} - P_{\mathbf{1}}\mathbf{u}$

$$\mathbf{u}_{(c)} = (\mathbf{1}\beta_0 + Z\boldsymbol{\gamma}) - P_{\mathbf{1}}(\mathbf{1}\beta_0 + Z\boldsymbol{\gamma}) = Z\boldsymbol{\gamma} - P_{\mathbf{1}}Z\boldsymbol{\gamma} = Z_{(c)}\boldsymbol{\gamma}$$

则 $(r_{\mathbf{u},\mathbf{y}})^2 = \dfrac{(\mathbf{u}_{(c)}{}^{\mathsf{T}}\mathbf{y}_{(c)})^2}{\|\mathbf{u}_{(c)}\|^2\|\mathbf{y}_{(c)}\|^2} = \dfrac{(\boldsymbol{\gamma}^{\mathsf{T}}Z_{(c)}{}^{\mathsf{T}}\mathbf{y}_{(c)})^2}{\boldsymbol{\gamma}^{\mathsf{T}}Z_{(c)}{}^{\mathsf{T}}Z_{(c)}\boldsymbol{\gamma}\cdot\|\mathbf{y}_{(c)}\|^2}$,

令$\mathbf{w} = (Z_{(c)}{}^{\mathsf{T}}Z_{(c)})^{1/2}\boldsymbol{\gamma}$

$$= \frac{\left(\mathbf{w}^{\mathsf{T}}\left[(Z_{(c)}{}^{\mathsf{T}}Z_{(c)})^{-1/2}Z_{(c)}{}^{\mathsf{T}}\mathbf{y}_{(c)}\right]\right)^2}{\mathbf{w}^{\mathsf{T}}\mathbf{w}\cdot\|\mathbf{y}_{(c)}\|^2}$$

应用CS不等式

$$\le \frac{\mathbf{w}^{\mathsf{T}}\mathbf{w}\times\mathbf{y}_{(c)}{}^{\mathsf{T}}Z_{(c)}(Z_{(c)}{}^{\mathsf{T}}Z_{(c)})^{-1}Z_{(c)}{}^{\mathsf{T}}\mathbf{y}_{(c)}}{\mathbf{w}^{\mathsf{T}}\mathbf{w}\cdot\|\mathbf{y}_{(c)}\|^2} = \frac{\|\hat{\mathbf{y}}_{(c)}\|^2}{\|\mathbf{y}_{(c)}\|^2} = R^2$$

当$\mathbf{w} \propto (Z_{(c)}{}^{\top} Z_{(c)})^{-1/2} Z_{(c)}{}^{\top} \mathbf{y}_{(c)} \Leftrightarrow$

$\boldsymbol{\gamma} = (Z_{(c)}{}^{\top} Z_{(c)})^{-1/2} \mathbf{w} \propto (Z_{(c)}{}^{\top} Z_{(c)})^{-1} Z_{(c)}{}^{\top} \mathbf{y}_{(c)}$时，等号成立.

(注意极大化问题的解不唯一：若$\mathbf{v}$是最优解，则$\mathbf{v}a + \mathbf{b}$也是)

此时 $\mathbf{u}_{(c)} = Z_{(c)}\hat{\boldsymbol{\gamma}}$，$\mathbf{u} = Z_{(c)}\hat{\boldsymbol{\gamma}} + \mathbf{1}\bar{\mathrm{u}}$，特别地在$\mathbf{u} = Z_{(c)}\hat{\boldsymbol{\gamma}} + \mathbf{1}\bar{y} = X\hat{\boldsymbol{\beta}}$达到极大。

注：

• 在X列空间上寻找与响应$\mathbf{y}$相关系数绝对值最大的组合$\mathbf{u} = X\boldsymbol{\beta}$

（这称为典则相关分析），LS估计$\hat{\boldsymbol{\beta}} = (X^{\top} X)^{-1} X^{\top} \mathbf{y}$为最优解。

R^2 可理解为为$\mathbf{y}$与X所有列的一种相关性度量。

命题3. 对于模型

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + Z\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim (0, \sigma^2)$$

记样本方差-协方差矩阵：$\hat{\Sigma}_{\mathbf{xx}} = Z_{(c)}{}^{\top} Z_{(c)}/(n-1)$，$\hat{\Sigma}_{\mathbf{xy}} = Z_{(c)}{}^{\top}\mathbf{y}/(n-1)$,

以及$\hat{\Sigma}_{yy\bullet\mathbf{x}} = \hat{\Sigma}_{yy} - \hat{\Sigma}_{y\mathbf{x}}\hat{\Sigma}_{\mathbf{xx}}^{-1}\hat{\Sigma}_{\mathbf{x}y}$，则$R^2 = \dfrac{\hat{\Sigma}_{y\mathbf{x}}\hat{\Sigma}_{\mathbf{xx}}^{-1}\hat{\Sigma}_{\mathbf{x}y}}{\hat{\Sigma}_{yy}}$.

证明：推论1(2）表明$RSS = (n-1)\hat{\Sigma}_{yy\bullet\mathbf{x}}$

所以$SS_{回} = SS_{总} - RSS = (n-1)\hat{\Sigma}_{yy} - (n-1)\hat{\Sigma}_{yy\bullet\mathbf{x}} = (n-1)\hat{\Sigma}_{y\mathbf{z}}\hat{\Sigma}_{\mathbf{zz}}^{-1}\hat{\Sigma}_{\mathbf{z}y}$.

细节如下：$\hat{\mathbf{y}} = P_{\mathbf{1}}\mathbf{y} + P_{Z_{(c)}}\mathbf{y} = \mathbf{1}\bar{y} + Z_{(c)}\hat{\boldsymbol{\gamma}}$，所以$\hat{\mathbf{y}} - \mathbf{1}\bar{y} = Z_{(c)}\hat{\boldsymbol{\gamma}}$，

所以$\mathrm{SS}_{回} = \|\hat{\mathbf{y}} - \mathbf{1}\bar{y}\|^2 = \|Z_{(c)}\hat{\boldsymbol{\gamma}}\|^2 = \mathbf{y}^{\top} Z_{(c)}(Z_{(c)}{}^{\top} Z_{(c)})^{-1} Z_{(c)}{}^{\top}\mathbf{y} = (n-1)\hat{\Sigma}_{y\mathbf{z}}\hat{\Sigma}_{\mathbf{zz}}^{-1}\hat{\Sigma}_{\mathbf{z}y}$

而$\mathrm{SS}_{总} = \|\mathbf{y} - \mathbf{1}\bar{y}\|^2 = (n-1)\hat{\Sigma}_{yy} \Rightarrow R^2 = \dfrac{\hat{\Sigma}_{y\mathbf{z}}\hat{\Sigma}_{\mathbf{zz}}^{-1}\hat{\Sigma}_{\mathbf{z}y}}{\hat{\Sigma}_{yy}}$。