Formula Sheet

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_1 + \frac{\sum u_i (X_i - \bar{x})}{(\sum (X_i - \bar{x})^2)} \text{ i.e. } SST_x$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$SST = \sum (Y_i - \bar{Y})^2 \qquad SSE = \sum (\hat{Y}_i - \bar{Y})^2 \qquad SSR = \sum (Y_i - \hat{Y}_i)^2$$

$$R^2 = \frac{SSE}{SST}$$

$$\tilde{\delta}_1 = \frac{\sum (x_{i,1} - \bar{x}_1)(x_{i,2} - \bar{x}_2)}{\sum (x_{i,1} - \bar{x}_1)^2}$$

$$\tilde{\beta}_1 = \hat{\beta}_1 + \tilde{\delta}_1 \hat{\beta}_2$$

slope from regression of X_2 on X_1

$$r_{x,y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2}\sqrt{\sum (Y_i - \bar{Y})^2}}.$$

$$\tilde{\beta}_1 = \beta_1 + \frac{\sum (x_{1,i} - \bar{x}_1) u_i}{\sum (x_{1,i} - \bar{x}_1)^2} + \beta_2 \tilde{\delta}_1$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum X_i^2}{n \sum (X_i - \bar{X})^2}$$

$$\hat{\sigma}^2 = \frac{SSR}{n-2} \qquad \hat{\sigma}^2 = \frac{SSR}{n-k-1}$$

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} \qquad t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

$$\widehat{Var(\hat{\beta}_1)} = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2}$$

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/n-k-1} \sim F_{q,\,n-k-1}$$

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \bar{X})^2}}$$

$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)}$$

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

overall: $F = \frac{R^2/q}{(1-R^2)/(n-k-1)}$

Probability density function of a standard normal random variable:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Cumulative distribution function of a standard logistic random variable:

$$\Lambda(z) = \frac{e^z}{1 + e^z}$$

Probability density function of a standard logistic random variable:

$$\lambda(z) = \frac{e^z}{(1 + e^z)^2}$$

MLR.1 (Linear in parameters): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$.
MLR.2 (Random sampling)
MLR.3 (No perfect collinearity)
MLR.4 : (Zero conditional mean) $E(u \mid x_1, \ldots, x_k) = 0$
MLR.5 (Homoskedasticity) $Var(u \mid x_1, \ldots, x_k) = \sigma^2$

$\Big\}$ Gauss-Markov Theorem $\Rightarrow$ OLS estimations are BLUE

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j (1 - R_j^2)}$$

SST_j: total sample variation in x_j

R_j^2: R-squared from a regression of x_j on all other independent variables

MLR.6 : $u_i \sim N(0, \sigma^2)$ independently of $x_{i1}, x_{i2}, \ldots, x_{ik}$.

$\Rightarrow y \mid x \sim N(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k, \sigma^2)$.

using $\log(y)$ will be more closely than y.

MLR1-6 $\Rightarrow$ classical linear model assumption.

α (given) = prob of type I error. (we can stand)
= significance level size of the test.

p-value $= P(T > t_{\hat{\beta}_j})$

$H_0: \beta_1 = \beta_2 \Rightarrow$ ① $t = \dfrac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)} \sim t_{n-k-1}$ ② Let $\theta = \beta_1 - \beta_2 \Rightarrow H_0: \theta = 0,\ H_1: \theta < 0$

$\beta_1 = \theta + \beta_2$

OLS estimators are consistent. $plim\ \hat{\beta}_j = \beta_j \quad j = 0, 1, 2, \ldots, k$

$$y = \beta_0 + \beta_1 x_1 + \boxed{\beta_2 x_2 + v}\ u. \quad \tilde{\beta}_1 = \beta_1 + \frac{\sum u_i (x_{1i} - \bar{x}_1)}{\sum (x_{1i} - \bar{x}_1)^2}$$

$$plim\ \tilde{\beta}_1 = \beta_1 + \frac{Cov(x_1, u)}{Var(x_1)} = \beta_1 + \frac{Cov(x_1, x_2)}{Var(x_1)} \beta_2$$

$$Cov(x_1, x_2) = 0 \rightleftharpoons \tilde{\beta}_1 \text{ consistent}$$

$$x_1, x_2 \text{ indep} \rightarrow E\left(\beta_1 + \frac{\sum u_i (x_{1,i} - \bar{x}_1)}{\sum (x_{1,i} - \bar{x}_1)^2}\right) = \beta_1 + \underbrace{E(u)}_{=0} E\left(\frac{\sum (x_{1,i} - \bar{x}_1)}{\sum (x_{1,i} - \bar{x}_1)^2}\right) = \beta_1$$

i.e. unbiased.

average partial effect: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_1 \cdot x_2 + u$

$$\frac{\Delta y}{\Delta x_1} = \beta_1 + 2\beta_3 x_1 + \beta_4 x_2 \qquad APE_{x_1} = \hat{\beta}_1 + 2\hat{\beta}_3 \bar{x}_1 + \hat{\beta}_4 \bar{x}_2$$

Non-nested Model

if neither model is a special case of the other

$y = \beta_0 + \beta_1 \log x_1 + u$.

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + u$.

CHOW Test

将数据分为两部分
检验两部分参数是否相等 $\left\{ \begin{array}{l} \text{分男女} \\ \text{两组} \end{array} \right.$
由此判断结构是否发生变化.

$y = \beta_{g,0} + \beta_{g,1} x_1 + \beta_{g,2} x_2 + \cdots + \beta_{g,k} x_k + u$.

$g = 1, 2.$ $\quad n_1 / n_2$.

$$\boxed{\begin{aligned} H_0:\ & \beta_{1,0} = \beta_{2,0} = \beta_0 \\ & \beta_{1,1} = \beta_{2,1} = \beta_1 \\ & \cdots\cdots \\ & \beta_{1,k} = \beta_{2,k} = \beta_k \\ H_1:\ & H_0 \text{ not true.} \end{aligned}}$$

$SSR_{ur} = SSR_1 + SSR_2$

$SSR_r = SSR_P$ pooled 将两组数据混合评估 与原 SSR_r 相同. 即 restricted

$$F = \frac{[SSR_p - (SSR_1 + SSR_2)] / (k+1)}{(SSR_1 + SSR_2) / (n - 2k - 2)} \sim F_{k+1,\, n-2k-2}.$$

Binary dependent variable: Linear probability model (LPM). : $y = \begin{cases} 1 \\ 0 \end{cases}$, $E(y|x) = P(y=1|x)$

$P(y=1|X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \Rightarrow y = P(y=1|X)$, $\hat{y} = \widehat{P(y=1|X)}$ (No guarantee that $0 \le \hat{y} \le 1$).

$$\beta_j = \frac{\Delta P(y=1|X)}{\Delta x_j}$$

LPM violates one of the Gauss-Markov assumptions (MLR1-5)
(MLR.5 : Homoskedasticity).
$Var(y|x) = p(y=1|x)(1 - p(y=1|x))$.

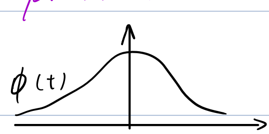
nonLinear model for binary reponse. $P(y=1|X) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) = G(\beta X)$

Candidates for G:

① logit model: $G(z) = \dfrac{e^z}{1 + e^z} = \Lambda(z)$ c.d.f of standard logistic distribution

② probit model: $G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(z)\,dz$

c.d.f. of standard normal distribution.



$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

estimation: $L(\beta) = \prod_{i=1}^{n} f(y_i | x_i ; \beta) \Rightarrow \max \log(L(\beta))$.

heteroskedasticity

one regressor $Var(\hat{\beta}_j) = \frac{1}{SST_x^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \sigma_i^2$

multiple regressors $\widehat{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$

- **Weighted least squares estimation**
- Heteroskedasticity is known up to a multiplicative constant

$Var(u_i | \mathbf{x}_i) = \sigma^2 h(\mathbf{x}_i), \; h(\mathbf{x}_i) = h_i > 0$ ← The functional form of the heteroskedasticity is known

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$$

$$\Rightarrow \left[\frac{y_i}{\sqrt{h_i}}\right] = \beta_0 \left[\frac{1}{\sqrt{h_i}}\right] + \beta_1 \left[\frac{x_{i1}}{\sqrt{h_i}}\right] + \cdots + \beta_k \left[\frac{x_{ik}}{\sqrt{h_i}}\right] + \left[\frac{u_i}{\sqrt{h_i}}\right]$$

$\Leftrightarrow \; y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \cdots + \beta_k x_{ik}^* + u_i^*$ ← Transformed model