

- 一、(1) 无偏性: 估计值等于真实值或估计值的平均值等于真实值
 (2) 异方差
 (3) 拟合优度: 一个模型对观测数据的拟合程度 ($R^2 = 1 - \frac{\text{残差平方和}}{\text{总平方和}}$)
 (4) 异方差: 一个变量的方差在不同条件下不恒定
 (5) 虚拟变量: 把某种信息表达成数值

- 二、(1) 假设: ① 线性关系 因变量与自变量关系线性
 ② 同方差性 误差项的方差不随自变量变化
 ③ 正态性 误差项正态分布
 ④ 独立性 误差项与自变量相互独立

- (2) 简述: ① 建立假设
 ② 选择适当的检验统计量
 ③ 根据检验统计量的值和显著性水平, 判断假设是否成立
 原理: 对总体参数假设, 搜集样本数据, 计算假设多大程度可靠

- (3) 影响: ① 方差过大
 ② 显著性检验失效
 ③ 精度下降, 可靠度降低

原理: 检验残差平方和是否与自变量显著相关

过程: ① 将残差平方和作为因变量, 重新拟合模型

② 计算 R^2

③ 判断是否存在异方差问题

三、证明

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

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

四、(1) $\ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{abil} + u$

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} \pm t_{0.1} \frac{s_y}{\sqrt{n}} \\ &= 5000 \pm \frac{1333}{\sqrt{50}} \\ &= 5000 \pm 2.9 \end{aligned}$$

$$\begin{aligned} \hat{\beta}_1 &= \bar{x}_1 \pm t_{0.1} \frac{s_1}{\sqrt{n}} \\ &= 1200 \pm \end{aligned}$$

$$n - k - 1 = 17$$

$$\alpha = 0.1$$

$$t_{\frac{\alpha}{2}} = t_{0.05} = 1.740$$

$$\begin{aligned} \hat{\beta}_0 &: 5000 \pm 1.740 \times 1000 \\ & \quad * (3260, 6740) \end{aligned}$$

$$\begin{aligned} \hat{\beta}_1 &: 1200 \pm 1.740 \times 400 \\ & \quad (504, 1896) \end{aligned}$$

$$\begin{aligned} \hat{\beta}_2 &: 2000 \pm 1.740 \times 200 \\ & \quad (1652, 2348) \end{aligned}$$

(2) ① 增加样本量 n

② 减少自变量 k

③ ~~增加~~ 减少置信水平

④ 选择更优的模型

202300300269 谭绍庭

五、(1) 每多受教育一年，工资的自然对数会改变多少的期望值增加 0.078

(2) $t = \frac{0.078}{0.007} = 11.14 > 1.96$

拒绝原假设，教育对工资有显著影响