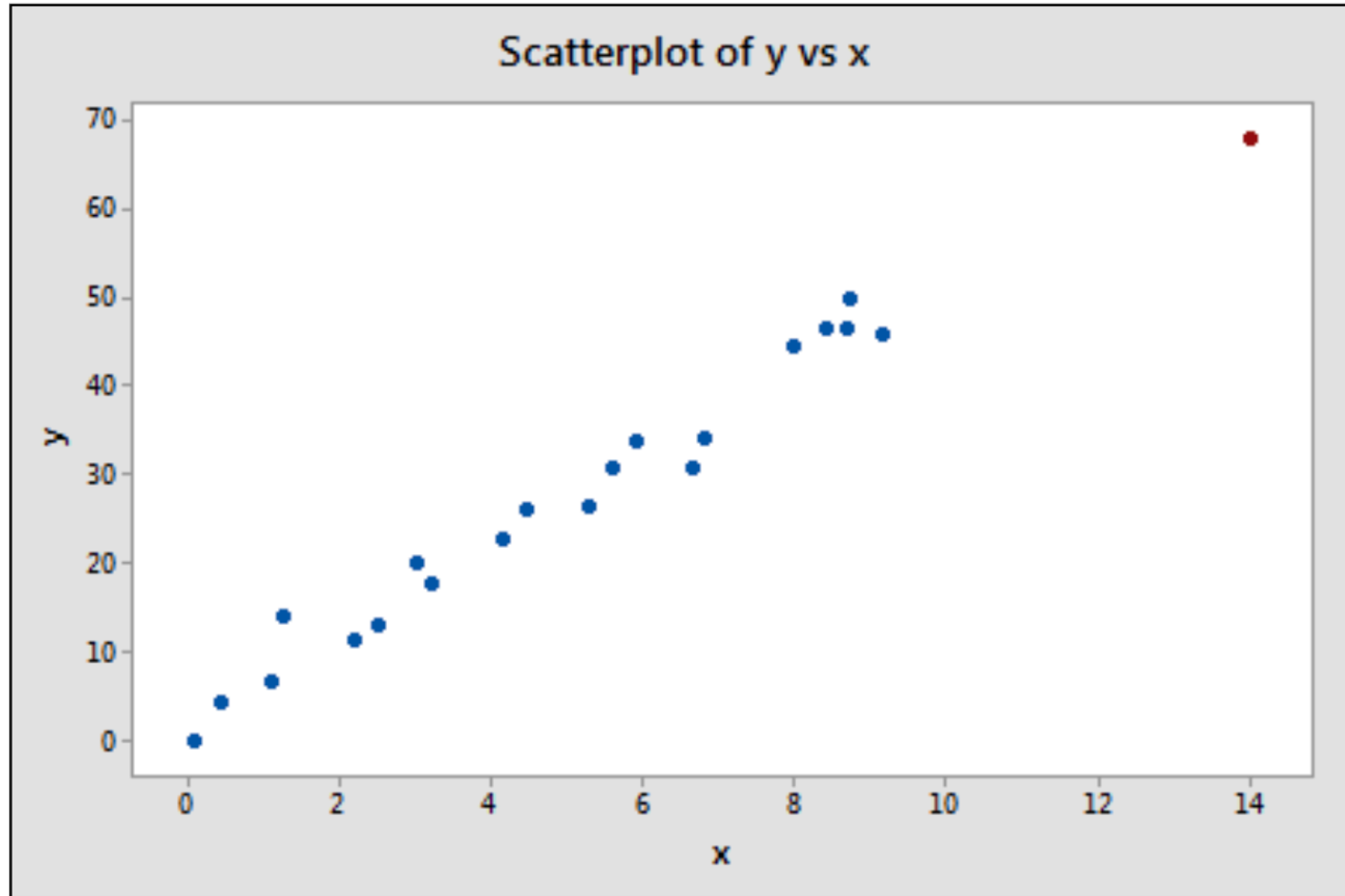


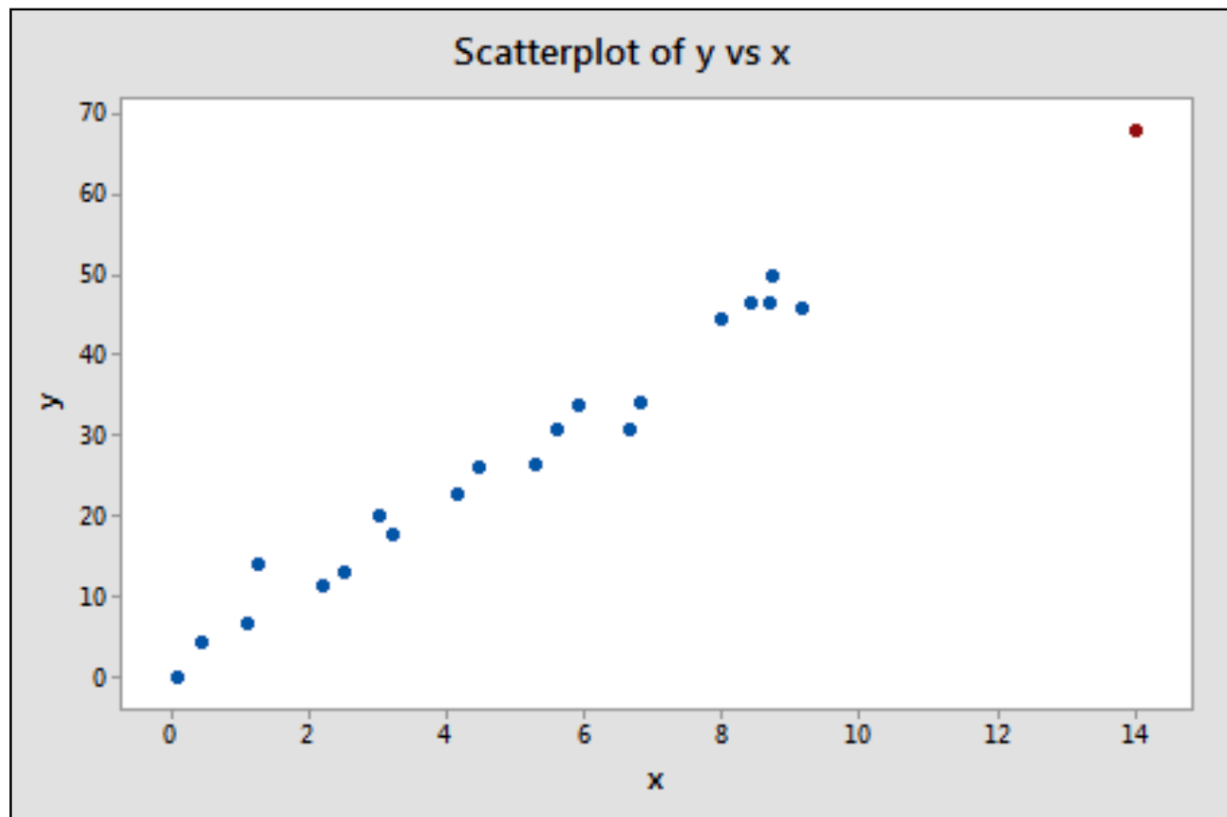
# 回归分析

强影响点，异常点，高杠杆点

# 高杠杆点 (high leverage point)



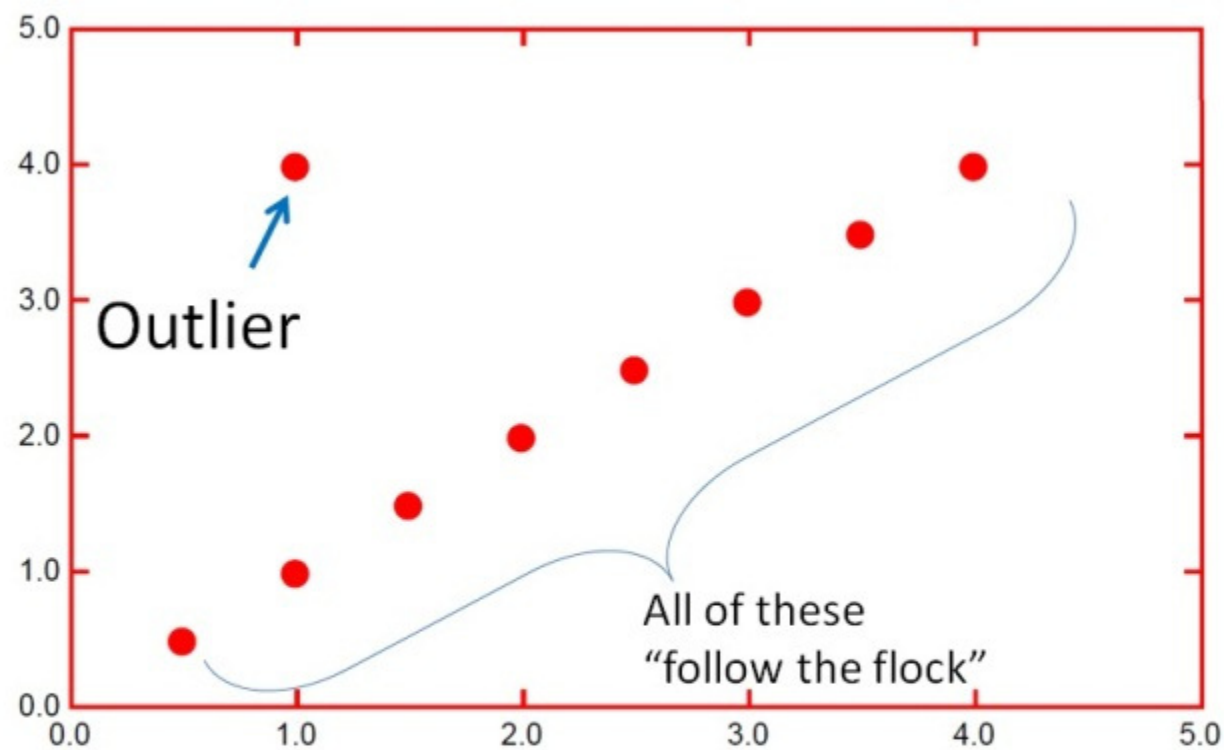
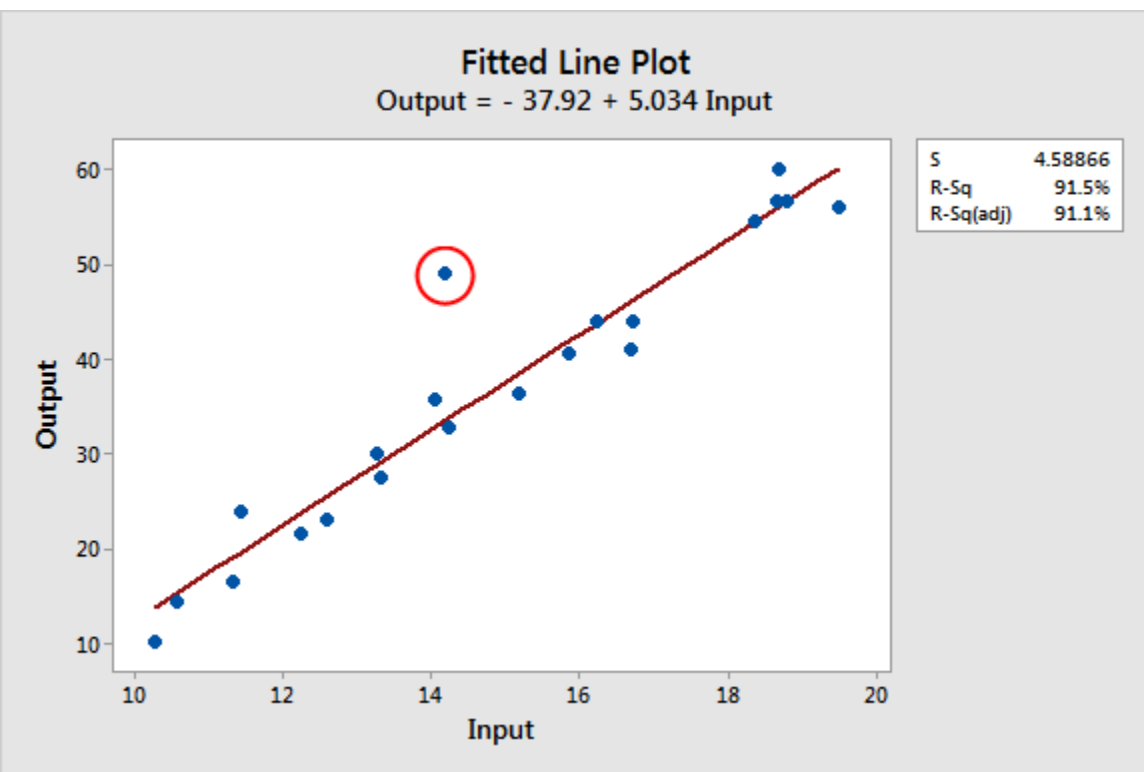
# 高杠杆点 (high leverage point)



$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{1}{n} + \frac{1}{n-1} \left( \frac{x_i - \bar{x}}{s_x} \right)^2$$

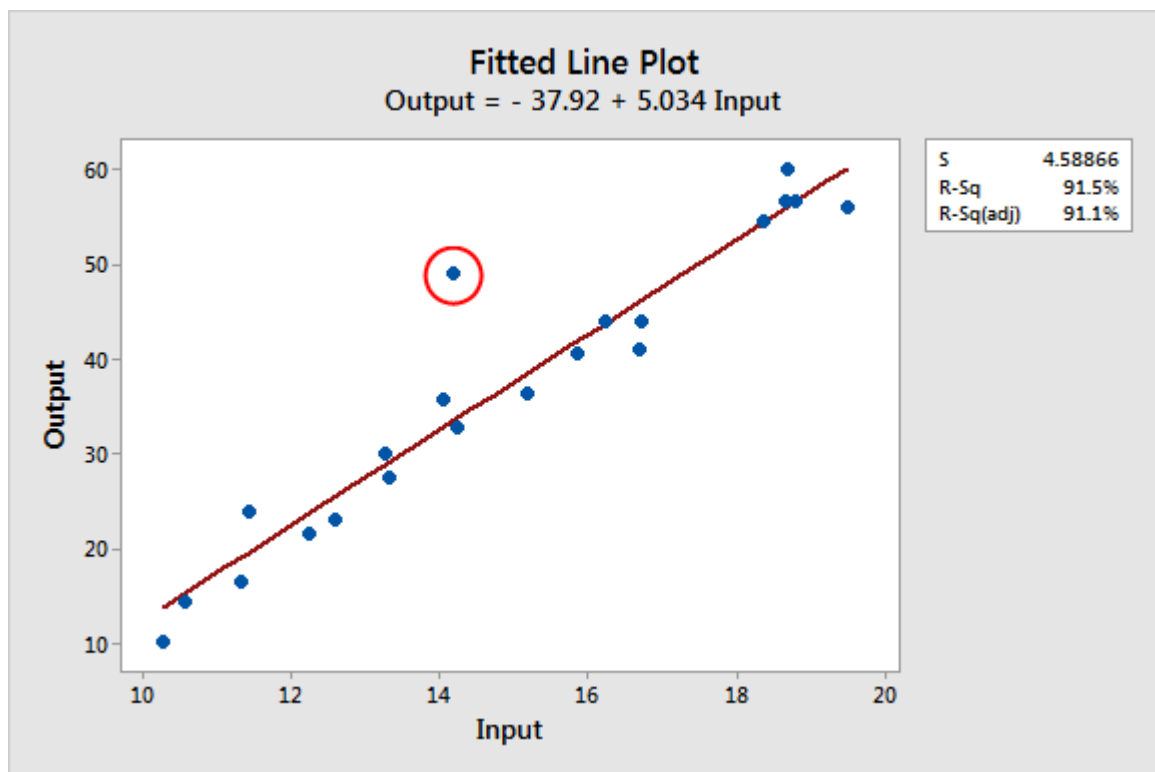
如果  $h_{ii} > 2\frac{p}{n}$  , 那就是高杠杆点。

# 异常点(outlier)



Never mind what the axes mean...

# 异常点(outlier)



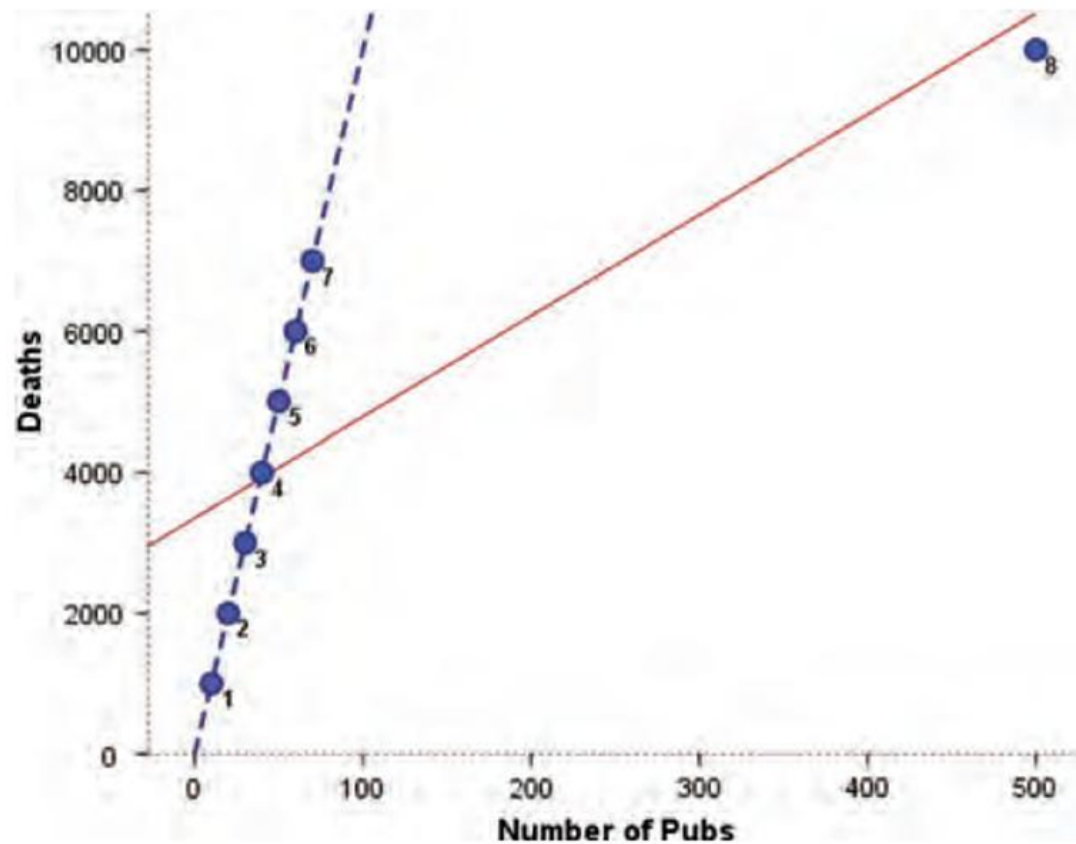
标准化残差：

$$r_i = \frac{\hat{\epsilon}_i}{\hat{\sigma} \sqrt{1 - h_{ii}}}$$

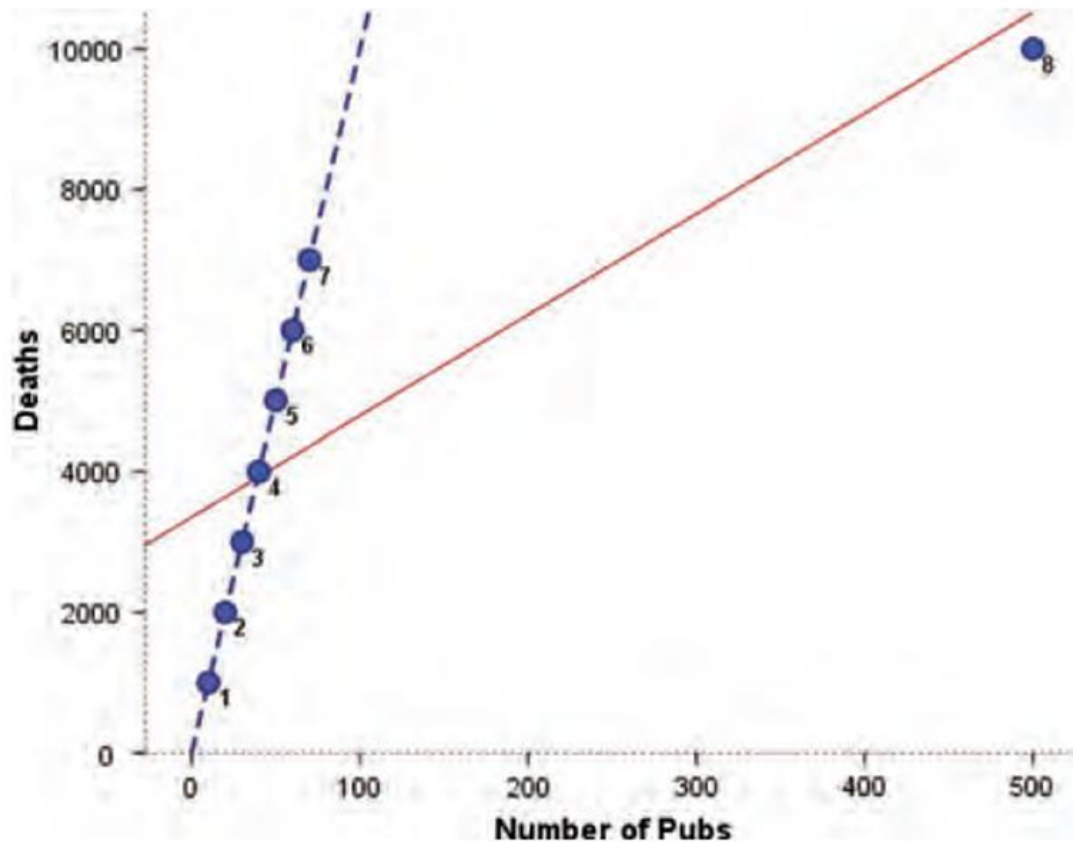
Handwritten annotations:  $\sqrt{MSE}$  points to  $\hat{\sigma}$ , and a blue box highlights  $\sqrt{1 - h_{ii}}$ .

如果大于1.96，就是异常点，潜在影响点。

# 强影响点



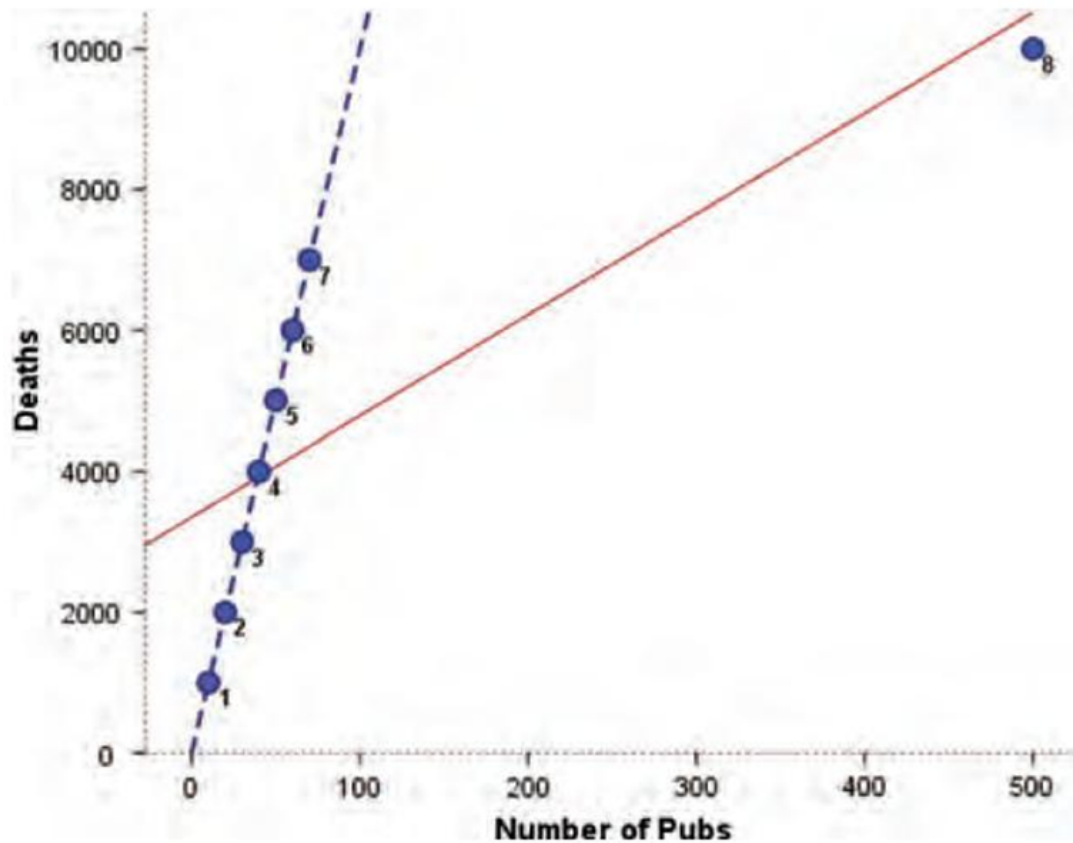
# Cook distance(库克距离)



$$D_i = \frac{\sum_{j=1}^n (\hat{Y}_j - \hat{Y}_{j(i)})^2}{(p+1) \hat{\sigma}^2}$$

如果Cook distance >1或者相对其他较高就是强影响点。

# 看斜率改变多少



$$\text{斜率的改变率} = \frac{\hat{\beta} - \hat{\beta}_{(i)}}{\hat{\beta}}$$