

## 7. inference for 线性回归

a. 你的数据和模型是否合理, 可信.

b. slope — 假设检验.

x 是否良好预测了 y.

$H_0: \beta_1 = 0$ : x 对 y 没有意义  $\rightarrow$  slope = 0

$H_1: \beta_1 \neq 0$ : x 对 y 有意义

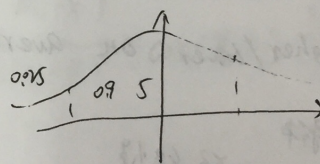
使用 t-统计推断.  $T = \frac{\text{point estimate} - \text{null value}}{SE}$

$$T = \frac{k - 0}{SE_k}$$

df = n - 2 自由变量除 k, b, 估计两个参数

c. confidence 区间 slope

$$k \pm t_{df} SE_k$$



$$gt(0.05, df=25) = -0.2059 \quad \therefore t^* = 2.06$$

$$0.9014 \pm 2.06 \times 0.0943 = (0.7, 1.1)$$

我们 95% 确信 x, y are expected on average to be higher by 0.7 to 1.1 points

1.1 points

d. 识别 data 类型: 随机抽样, population, 非-随机抽样

统计推断和 P-value 对于 population 无意义.

Caution 如果 data 非随机 (biased), 结果是不可信的

目标: have 相关性 — 并且你知道如何检查它们









