第十一章：进行预测

① 样本均值：

$\bar{x} = \frac{\sum x}{n}$ ，$\hat{\mu} = \bar{x}$.

② 样本方差：

$\hat{\sigma}^2 = \frac{\sum (x-\bar{x})^2}{n-1}$. （因使用正常方法时，样本方差偏小，所以使用 $n-1$）.

$\hat{\sigma}^2 = s^2$.

③ 预测总体比例.

p_s：样本的成功比例.

$\hat{p} = p_s$ ，$p_s = \frac{成功数目}{样本数目}$

③ 为样本计算概率 .（p_s 的分布）.

$p = 0.25$

$n = 100$

$X \sim B(n, p)$.

$p_s = X/n$

$E(p_s) = E(\frac{X}{n}) = \frac{E(X)}{n} = \frac{np}{n} = p$.

$Var(p_s) = Var(\frac{X}{n}) = \frac{Var(X)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}$.

$p_s \sim N(p, \frac{pq}{n})$.

④ 求样本均值的概率

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$ （将所有的样本均值形成一个分布）

$$E(\bar{X}) = E\left(\frac{1}{n}X_1\right) + E\left(\frac{1}{n}X_2\right) + \cdots + E\left(\frac{1}{n}X_n\right)$$

$$= \frac{1}{n}\left(E(X_1) + E(X_2) + \cdots + E(X_n)\right)$$

$$= \frac{1}{n}(n\mu) = \mu$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n\right)$$

$$= \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + \cdots + \sigma^2\right)$$

$$= n \cdot x \cdot \frac{1}{x^2} \cdot \sigma^2 = \frac{\sigma^2}{n}$$

（中心极限定理：从一个非正态总体X中取出一个样本，且样本很大，则$\bar{X}$近似于正态分布）

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

第十二章：置信区间

① 求95%置信区间（均值）

$E(\bar{x})=\mu$，$Var(\bar{x})=\frac{s^2}{n}$

↓

$P(\bar{x}<a)=0.025$ 和 $P(\bar{x}>b)=0.025$

$\downarrow \bar{x}\sim N(\mu,0.25)$

$z=\frac{\bar{x}-\mu}{\sqrt{0.25}}$，$z\sim N(0,1)$

↓ 查表

$P(-1.96<\frac{\bar{x}-\mu}{0.5}<1.96)=0.95$ ⟶ $P(\bar{x}-0.98<\mu<\bar{x}+0.98)=0.95$

$\bar{x}=62.7$

↓

置信区间为 $(61.72, 63.68)$

② 求99%置信区间（概率）

$\left(p_s-c\sqrt{\frac{p_s q_s}{n}},\ p_s+c\sqrt{\frac{p_s q_s}{n}}\right)$

$c=2.58$，$p_s=0.25$，$q_s=0.75$，$n=50$

③ 当样本太小时

↓

t分布 ⟶ 当样本很大时，t近似正态分布
当样本很小时，t曲线较为扁平

$T \sim t(v)$ ，$v=n-1$.
正在使用自由度为v的t分布

$$T = \frac{\bar{x}-\mu}{s/\sqrt{n}} ,\quad \left(\bar{x} - t\frac{s}{\sqrt{n}} ,\ \bar{x} + t\frac{s}{\sqrt{n}}\right)$$

$$P(-t \le T \le t) = 0.95$$

第十三章：假设检验：

①确定假设 ②选择检验统计量
③确定拒绝域 ④求出统计量P值
⑤与拒绝域比较 ⑥作出决策

$H_0: p = 90\%$
$H_1: p < 90\%$

$P(X \le 11)$ 小于0.05 则拒绝H_0
$= 0.0555$
不能拒绝

单尾检验 vs 双尾检验

↓ ↓

< 或 > ≠

1: $X \sim B(n,p)$ 如果 n 很大，$np>5$ 且 $nq>5$

则可用正态分布近似

$X \sim N(\mu, \sigma^2)$，已知 μ 和 σ^2，则

$$\bar{X} \sim N(\mu, \sigma^2/N)$$

$X \sim N(\mu, \sigma^2)$ 已知 μ，但 σ^2 未知，则

$$\bar{X} \sim N(\mu, s^2/n)$$

$X \sim N(\mu, \sigma^2)$，已知 μ，但 σ^2 未知，样本很小，则

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

① Type I Error：

原假设实际正确下拒绝原假设。

$p = \alpha$。

② Type II Error：

当原假设错误的时候，却接受了原假设。

$p = \beta$。

$$Z = \frac{X-90}{3}$$

$P(\text{Type I Error}) = 0.05$，与显著性水平相同

$P(\text{Type II Error})$

$H_0: p = 0.9$

$H_1: p = 0.8$

$$\frac{X-90}{3} \geq -1.64$$

$$X \geq 85.08$$

当 X大于 85.08 则接受 H_0

需算出 $P(X \geq 85.08)$，其中 $X \sim N(80, 16)$

$$Z = \frac{85.08-80}{\sqrt{16}} = 1.27$$

$P(Z \geq 1.27) = 0.102$

$P(\text{Type II Error}) = 0.102$

↓

实际能治愈80%的情况，接受治愈90%患者

③ 认识功效

功效 $= 1-\beta$（在H_0为假时拒绝H_0）

功效 $= 1-\beta = 1-0.102 = 0.898$