区间估计

Guangyao Zhao

2022-08-18

Contents

点估计是估计出一个具体的值，但是很粗糙，并不能反映估计精确程度。区间估计则是给出一个范围，并给出此范围包含参数 θ 真值的可信度。相比于点估计，区间估计就要合理的多，因为在现实生活中，我们往往需要的是一个相对感性的把握，即一个范围就足够了，而并不需要一个具体的值。

比如估计『明天的温度八成在 $27 \sim 30$ 之间』，那么 $27 \sim 30$ 就是置信区间，其中 27 是置信下限，八成表示有 80% 的可信度。除了可信度之外，还需要一定的精确度满足足够的需求，比如『明天的温度八成在 $-50 \sim 50$ 之间』就是一个非常没有意义的推断，可信度极其高，但是精度也是极其差。

置信区间

假设总体 X 的分布函数为 $F(x;\theta,\theta \in \Theta)$。对于给定值 $\alpha(0<\alpha<1)$，若由来自 X 的样本 $X_1,X_2,...,X_n$ 确定的两个统计量 $\underline{\theta}=\theta(X_1,X_2,...,X_n),\overline{\theta}=\theta(X_1,X_2,...,X_n),(\underline{\theta}<\overline{\theta})$，对于任意的 $\theta \in \Theta$ 满足:

$$P\{\underline{\theta}=\theta(X_1,X_2,...,X_n)<\theta<\overline{\theta}=\theta(X_1,X_2,...,X_n)\}\geq 1-\alpha$$

则称随机区间 $(\underline{\theta},\overline{\theta})$ 是 θ 的置信水平为 $1-\alpha$ 的置信区间。

置信区间的思想是中心极限定理，样本的均值的也好，方差也好，都应该大概率的集中在真实的全体均值和方差周围，就是我们常说的 σ，2σ，3σ 定理。

置信区间和置信度是两个可灵活变动的指标，通过调置信度即 $100(1-\alpha)\%$ 可改变置信区间，反之亦然。

估计正态分布中的 μ

假设总体 $X\sim N(\mu,\sigma^2)$，μ 未知，方差已知，求 μ 的置信水平为 $1-\alpha$ 的置信区间。寻找一个和 μ 有关的统计量，该统计量的分布是确定的，用以确定 $(\underline{\mu},\overline{\mu})$，使得 $P\{\underline{\mu}<\mu<\overline{\mu}\}=1-\alpha$

1. 已知 $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$，将其转化为标准正态分布 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$
2. 根据正态分布图形：

$$P\{-z_{\frac{\alpha}{2}}<\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<z_{\frac{\alpha}{2}}\}=1-\alpha$$

3. 解上式得：

$$P\{\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}<\mu<\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\}=1-\alpha$$

寻求一个样本 $X_1,X_2,...,X_n$ 和 θ 的函数 W，要求 W 的分布不依赖于 θ 以及其它参数，具有此种性质的函数 W 为枢轴量：比如上例子中枢轴量为 $\bar{X}$，其分布已知为 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，只有 μ 为未知量，n 和 σ 为已知量。

常用枢轴量

1. $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$
2. $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$
3. $\frac{(n-1)S^2}{\sigma^2}\sim \chi^2(n-1)$
4. $\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

正态总体均值与方差的区间估计

μ 的置信区间

σ 已知（正态分布）

若 σ 已知，则正常使用 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ 即可

σ 未知（t 分布）

若 σ 未知，则需要用 $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ ，其中 $E(S^2)=\sigma^2$，即

$$P\{-t_{\frac{\alpha}{2}}(n-1) < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\frac{\alpha}{2}}(n-1)\} = 1-\alpha$$

求解得：

$$P\{\bar{X} - \frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1) < \mu < \bar{X} + \frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)\} = 1-\alpha$$

σ 的置信区间

μ 已知（正态分布）

这个比较简单，使用 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ 即可

μ 未知（t 分布）

同估计 μ 的置信区间 σ 未知的情况类似，只是由反解 μ 转化为反解 σ

单侧置信区间

在一些实际问题中，往往只关心上限或下限某一侧的问题。比如元器件的寿命的下限，化学样品中杂质的上限。