浙江大学
ZHEJIANG UNIVERSITY

第55讲：单个正态总体均值的假设检验

（标准差已知，Z检验）

设总体$X \sim N(\mu, \sigma^2)$, σ^2已知

$X_1, X_2, \ldots, X_n$是来自总体$N(\mu, \sigma^2)$样本.

$x_1, \ldots, x_n$是$X_1, \ldots, X_n$的样本观测值.

考虑假设问题(显著水平为α)

$$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0,$$

$$H_0: \mu = \mu_0, H_1: \mu < \mu_0,$$

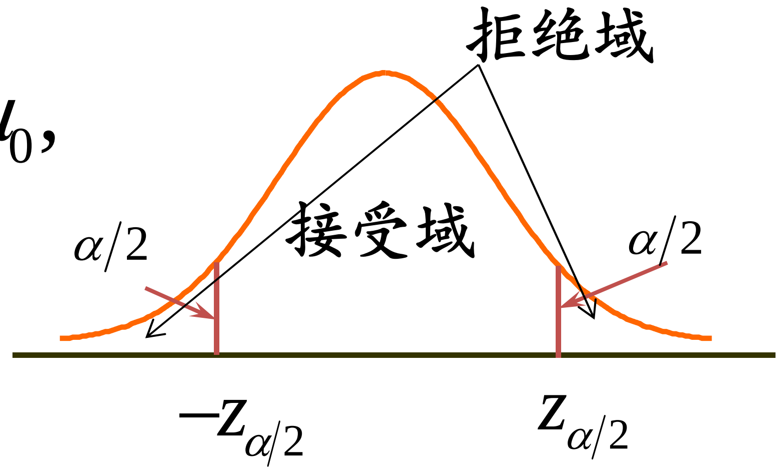
$$H_0: \mu = \mu_0, H_1: \mu > \mu_0,$$

其中μ_0是已知的常数.

双边假设问题

$$H_0:\mu=\mu_0,\ H_1:\mu\neq\mu_0,$$

其中μ_0是已知的常数.



检验统计量为 $Z=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

检验拒绝域$W=\left\{|Z|=\left|\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right|\geq z_{\alpha/2}\right\}.$

P_值的计算

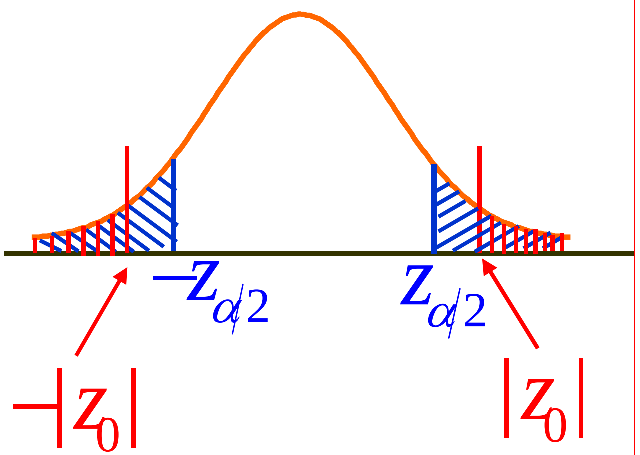
对给定的样本观察值$x_1,\cdots,x_n$,记检验统计量Z的取值

$$z_0=\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}},$$

$$P\_=P_{H_0}\{|Z|\geq|z_0|\}=2(1-\Phi(|z_0|)).$$

当$P_\leq\alpha$时，拒绝原假设，

当$P_>\alpha$时，接受原假设.



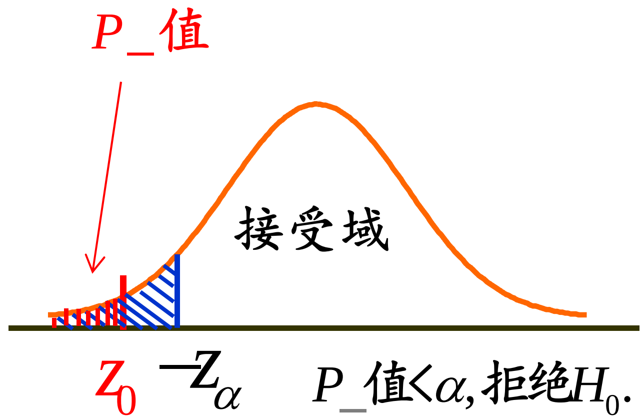
红色区域概率值：$P_$值

蓝色区域概率值：α

$P_$值$<\alpha$,拒绝H_0.

左边假设问题：$H_0:\mu=\mu_0, H_1:\mu<\mu_0,$

其中μ_0是已知的常数.



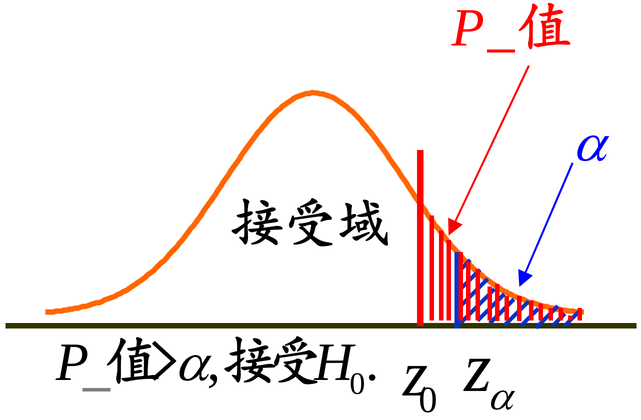
检验统计量为 $Z=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

检验拒绝域$W=\left\{Z=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\le -z_\alpha\right\}.$

$P\_=P_{H_0}\{Z\le z_0\}=\Phi(z_0).$ 其中$z_0=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}.$

右边假设问题： $H_0: \mu = \mu_0, H_1: \mu > \mu_0,$

其中μ_0是已知的常数.



检验统计量为 $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

检验拒绝域 $W = \left\{ Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq z_\alpha \right\}.$

$P\_ = P_{H_0}\{Z \geq z_0\} = 1 - \Phi(z_0).$ 其中 $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$

例1：为了了解A高校学生的消费水平，随机抽取225位学生调查其月消费(近6个月的消费平均值)，得到该225位学生的平均月消费为1530元. 假设学生月消费服从正态分布，标准差为σ=120.

已知B高校学生的月平均消费为1550 元，是否可以认为A高校学生的消费水平要低于B高校?

本例的Excel计算见实验18.

步骤1：提出检验假设

$$H_0: \mu = 1550,\ H_1: \mu < 1550$$

步骤2：确定检验规则

检验统计量为 $Z = \dfrac{\bar{X} - 1550}{\sigma/\sqrt{n}}$. 取显著水平$\alpha = 0.05$,

由备择假设的形式知，这是左边检验，因此检验规则为：当$Z \le -z_\alpha = -z_{0.05} = -1.645$时，拒绝$H_0$.

步骤3:计算检验统计量的值

将样本均值$\bar{x}=1530, \sigma=120, n=225,$

代入检验统计量，计算得

$$Z=\frac{\bar{X}-1550}{\sigma/\sqrt{n}}=\frac{1530-1550}{120/\sqrt{225}}=-2.5<-1.645.$$

步骤4：根据实际情况作出判断

因此，根据检验规则，做出拒绝原假设H_0的判断.

即认为A高校学生的生活水平低于B高校.

利用P_值进行假设检验

步骤3’：计算P_值

$$P\_ = P\left(\frac{\bar{X}-1550}{\sigma/\sqrt{n}} \le \frac{1530-1550}{120/\sqrt{225}} \middle| \mu = 1550\right)$$

$$= P(Z \le -2.5) = 0.006$$

步骤4’：根据显著水平作出判断

$$P\_ = 0.006 < \alpha = 0.05,$$

同样做出拒绝原假设$H_0 : \mu = 1550$的判断.

例2：据健康统计中心报告35至44岁的男子平均心脏收缩压为128，标准差为15. 现根据某公司在35至44岁年龄段的72位员工的体检记录，计算得平均心脏收缩压为126.07(mm/hg). 问该公司员工的心脏收缩压与一般人群是否存在差异呢?（假设该公司员工的心脏收缩压与一般中年男子的心脏收缩压具有相同的标准差）。(α=0.05)

步骤1：提出检验假设

$$H_0:\mu=128, H_1:\mu\neq 128$$

步骤2:计算检验统计量的观测值.

$$z_0=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}=\frac{126.07-128}{15/\sqrt{72}}=-1.09$$

步骤3：计算P_值

$$P_{-}=2(1-\Phi(|z_0|))=2(1-\Phi(1.09))=0.2758.$$

步骤4：根据实际情况作出判断

P_=0.2758>0.05，因此，没有充分理由拒绝原假设。

假设检验与区间估计

- 作区间估计时，对参数是未知，并且没有先验的认识，但参数是固定不变的，所以区间估计的目的是：根据样本对参数进行估计；
- 作假设检验时，对参数有一个先验的认识（例如 $\mu=\mu_0$），但由于某种情形的出现（如工艺改良等），猜测真实参数值可能发生了变化，所以假设检验的目的是：根据样本确认参数是否真的发生了改变。

但置信区间与假设检验的拒绝域之间又有密切的关系。

考虑总体$X \sim N(\mu,\sigma^2)$,σ^2已知时μ的统计推断

设$X_1, X_2,\ldots, X_n$是来自总体$N(\mu,\sigma^2)$样本.

μ的枢轴量为$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$,

μ的置信水平为$1-\alpha$的置信区间由下式得到,

$$\frac{\left|\bar{X}-\mu\right|}{\sigma/\sqrt{n}} < z_{\alpha/2}$$

等价为 $\bar{X}-\dfrac{\sigma}{\sqrt{n}}z_{\alpha/2} < \mu < \bar{X}+\dfrac{\sigma}{\sqrt{n}}z_{\alpha/2}$.

假设检验问题 $H_0: \mu = \mu_0 \; H_1: \mu \neq \mu_0,$

显 著 性 水 平 为 α 的 检 验 拒 绝 域 为

$$W = \left\{ \frac{|\bar{X} - \mu_0|}{\sigma / \sqrt{n}} \geq z_{\alpha/2} \right\},$$

接受域为

$$\bar{W} = \left\{ \frac{|\bar{X} - \mu_0|}{\sigma / \sqrt{n}} < z_{\alpha/2} \right\}$$

$$= \left\{ \bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu_0 < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right\}$$

将μ_0改为参数μ
就是置信区间！