检验的p值(p-value)

假设检验的结论通常是：在给定的显著性水平下，不是拒绝原假设就是保留原假设.

然而有时也会出现这样的情况：在一个较大的显著性水平($\alpha = 0.05$)下得到拒绝原假设的结论，而在一个较小的显著性水平($\alpha = 0.01$)下却会得到相反的结论.

这种情况在理论上很容易解释：因为显著性水平变小后会导致检验的拒绝域变小，于是原来落在拒绝域中的观测值就可能落入接受域.

这种情况在应用中会带来一些麻烦: 假如这时一个人主张选择显著性水平$\alpha = 0.05$, 而另一个人主张选显著性水平$\alpha = 0.01$, 则对于同一组样本, 可能第一个人的结论是拒绝H_0, 而后一个人的结论是接受H_0.

Example

例1: 一支香烟中的尼古丁含量X 服从正态分布$N(\mu, 1)$, 质量标准规定μ不能超过1.5毫克.现从某厂生产的香烟中随机抽取20支, 测得其中平均每支香烟的尼古丁含量为$\bar{x} = 1.97$毫克, 试问该厂生产的香烟尼古丁含量是否超过了质量标准的规定?

解: 这是一个假设检验问题

$$H_0 : \mu \le 1.5 \leftrightarrow H_1 : \mu > 1.5.$$

现在$\sigma = 1$已知，故采用u检验，取检验统计量为

$$u(\widetilde{X}) = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X} - 1.5}{1/\sqrt{20}},$$

拒绝域为

$$D = \{\widetilde{x} : u(\widetilde{x}) \ge u_\alpha\}.$$

计算得，$u(\widetilde{x}) = \frac{\overline{x}-1.5}{1/\sqrt{20}} = \frac{1.97-1.5}{1/\sqrt{20}} \approx 2.10.$

对一些不同的显著性水平，下表列出了相应的拒绝域和检验结论.

显著性水平	拒绝域	$u=2.10$对应的结论
$\alpha=0.05$	$u\geq 1.645$	拒绝H_0
$\alpha=0.025$	$u\geq 1.96$	拒绝H_0
$\alpha=0.01$	$u\geq 2.33$	接受H_0
$\alpha=0.005$	$u\geq 2.58$	接受H_0

我们看到, 对于同一个样本进行分析, 在不同的显著性水平α下, 可能得到不同的结论.

现在换一个角度来看，当$\mu = 1.5$时，$u(\widetilde{X})$的分布是$N(0,1)$. 此时可算得，

$$\mathsf{P}_{\mu=1.5}(u(\widetilde{X}) \geq 2.10) = 0.0179.$$

若以0.0179为基准来看上述检验问题,可得

- 当$\alpha < 0.0179$时，$u_\alpha > 2.10$. 于是样本就不在$\{u \geq u_\alpha\}$中，此时应接受原假设H_0;
- 当$\alpha \geq 0.0179$时，$u_\alpha \leq 2.10$. 于是样本就落入$\{u \geq u_\alpha\}$中，此时应拒绝原假设H_0;

由此可以看出,0.0179是能用此组样本(检验统计量的值为2.10)做出"拒绝H_0"的最小的显著性水平，这就是p值.

Definition

定义:

In statistical hypothesis testing, the p-value is the probability of obtaining a result at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.

(From Wikipedia, the free encyclopedia)

当原假设成立时，检验统计量取比观察到的结果更为极端的数值的概率.

引进检验的p 值有明显的好处:

- 它比较客观,避免了事先确定显著性水平;
- 由检验的p 值与人们心目中的显著性水平α进行比较，可以很容易作出检验的结论:

  如果$\alpha \geq p$，则在显著性水平α下拒绝H_0;

  如果$\alpha < p$，则在显著性水平α下保留H_0.

检验的p值在应用中很方便，如今的统计软件如:

R/Python/SPSS/SAS/Minitab/Excel等对检验问题一般都是对于给定的样本给出检验的p 值.

处理假设检验问题的一般步骤(一)

1. 根据问题提出原假设H_0和备择假设H_1;
2. 确定检验统计量$T(X_1, X_2, \cdots, X_n)$, 并根据原假设和备择假设确定拒绝域D的形式.(拒绝域D的形式主要依赖于备择假设的形式)
3. 选取适当的显著性α, 并求出临界值, 使得

$$\sup \mathsf{P}(\widetilde{X} \in D | H_0\text{为真}) \le \alpha\text{并尽可能地接近}\alpha.$$

在总体为连续型随机变量时, 往往要使得

$$\sup \mathsf{P}(\widetilde{X} \in D | H_0\text{为真}) = \alpha.$$

4. 由样本$\widetilde{X}$的观察值$\widetilde{x}$算出检验统计量$T(\widetilde{X})$的值$t = T(\widetilde{x})$, 并与临界点进行比较. 若观察值$\widetilde{x}$落入拒绝域D, 则拒绝原假设H_0, 否则接受原假设.

处理假设检验问题的一般步骤(二)

1. 根据问题提出原假设H_0和备择假设H_1;
2. 确定检验统计量$T(X_1, X_2, \cdots, X_n)$, 并根据原假设和备择假设确定拒绝域D的形式;
3. 计算检验统计量的值, 并由此得到p值;
4. 将p值与显著性水平进行比较, 得出相应的结论.

Example

例2: 某工厂两位化验员甲,乙分别独立地用相同方法对某种聚合物的含氯量进行测定.甲测了9次，得样本方差为0.7292；乙测了11次，得样本方差为0.2114. 假定测量数据服从正态分布，试对两总体方差作一致性检验(双侧):

$$H_0 : \sigma_{甲}^2 = \sigma_{乙}^2 \leftrightarrow H_1 : \sigma_{甲}^2 \neq \sigma_{乙}^2.$$

取检验统计量为

$$F = S_{甲}^2 / S_{乙}^2,$$

在原假设成立的条件下,$F \sim F(8,10)$,则拒绝域为

$$D = \{\widetilde{x} : F > F_{\alpha/2}(8,10) \ \text{ 或 } \ F < F_{1-\alpha/2}(8,10)\}.$$

现在我们换种方法，不把拒绝域具体化，而是由观测值算得检验统计量的值

$$F = 0.7292/0.2114 = 3.4494,$$

由此再去计算该检验的p值.

在这种双边检验情况下，如何由检验统计量的值$F=3.4494$来计算得p值呢？

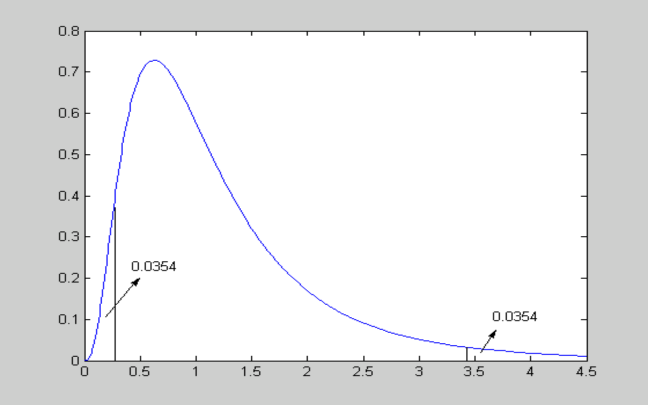
首先，我们用F分布算得

$$\mathsf{P}_{\sigma^2_{甲}=\sigma^2_{乙}}(F\geq 3.4494)=\mathsf{P}_{\sigma^2_{甲}=\sigma^2_{乙}}(S^2_{甲}/S^2_{乙}\geq 3.4494)=0.0354.$$

其次考虑到双边检验的拒绝域分散在两端，且两端尾部概率相等(见图A)，据此可定出p 值为

$$p=2\mathsf{P}_{\sigma^2_{甲}=\sigma^2_{乙}}(F\geq 3.4494)=0.0708.$$

此p值不算很小,若$\alpha=0.05$,则接受两方差相等的假设.



图A： 观测值$F=3.4494$对应的p值
由两端尾部概率之和确定

Example

例3: 设$x_1, x_2, \cdots, x_n$是来自$B(1,\theta)$的样本观测值,要检验如下(单侧)假设:

$$H_0: \theta \le \theta_0 \leftrightarrow H_1: \theta > \theta_0.$$

取检验统计量为$T(\widetilde{X}) = \sum_{i=1}^n X_i$, 则拒绝域形式为$D = \{T(\widetilde{x}) \ge C\}$. 则在得到观测值$\sum_{i=1}^n x_i = t_0$后,我们只需要计算概率$p = \mathsf{P}_{\theta_0}\{\sum_{i=1}^n X_i \ge t_0\}$. 这就是检验的$p$值.

譬如: 当$n = 40, \theta_0 = 0.1, t_0 = 8$, 则

$$p = 1 - 0.9^{40} - \binom{40}{1} \times 0.1 \times 0.9^{39} - \cdots - \binom{40}{7} \times 0.1^7 \times 0.9^{33} = 0.0419.$$

若取$\alpha = 0.05$, 由于$p < \alpha$, 则应拒绝原假设.

例：两台机床生产同一个型号的滚珠，从甲机床生产的滚珠中抽取8个，从乙机床生产的滚珠中抽取9个，测得这些滚珠的直径(毫米)如下：

甲机床	15.0	14.8	15.2	15.4	14.9	15.1	15.2	14.8	
乙机床	15.2	15.0	14.8	15.1	14.6	14.8	15.1	14.5	15.0

设两机床生产的滚珠直径分别为X,Y,且$X\sim N\left(\mu_1,\sigma_1^2\right),Y\sim N\left(\mu_2,\sigma_2^2\right)$

检验假设$H_0:\sigma_1^2=\sigma_2^2,H_1:\sigma_1^2\neq\sigma_2^2(\alpha=0.1)$

3

解一： 当μ_1,μ_2未知时， 检验 $H_0: \sigma_1^2=\sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2$

取检验统计量为：$\frac{S_1^2}{S_2^2}$，拒绝域形为$D = \left\{ \frac{s_1^2}{s_2^2} < C_1 \text{ 或 } \frac{s_1^2}{s_2^2} > C_2 \right\}$

由于$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

故在原假设成立时，即$\sigma_1^2=\sigma_2^2$时， $\frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1)$

故得拒绝域为:$\left\{ \frac{s_1^2}{s_2^2} < F_{1-\alpha/2}(n_1-1, n_2-1), \text{或 } \frac{s_1^2}{s_2^2} > F_{\alpha/2}(n_1-1, n_2-1) \right\}$.

查表得：$F_{0.05}(7,8)=3.50, F_{0.95}(7,8)=\frac{1}{F_{0.05}(8,7)}=\frac{1}{3.73}=0.268,$

故拒绝域为:$\left\{ \frac{s_1^2}{s_2^2} < 0.268, \text{或 } \frac{s_1^2}{s_2^2} > 3.50 \right\}$

本题中 $n_1=8$, $\bar{x}=15.05$, $s_1^2=0.0457$； $n_2=9$, $\bar{y}=14.9$, $s_2^2=0.0575$

计算得： $\frac{s_1^2}{s_2^2}=0.795 \in (0.268, 3.50)$

故接受原假设，认为方差没有显著差异.

解二：　当μ_1,μ_2未知时，检验 $H_0:\sigma_1^2=\sigma_2^2, H_1:\sigma_1^2\neq\sigma_2^2$

取检验统计量为：$\dfrac{S_1^2}{S_2^2}$，拒绝域形为$D=\left\{\dfrac{s_1^2}{s_2^2}<C_1 \text{ 或 } \dfrac{s_1^2}{s_2^2}>C_2\right\}$

由于$\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

故在原假设成立时，即$\sigma_1^2=\sigma_2^2$时，　$\dfrac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1)$

本题中 $n_1=8$，$\bar{x}=15.05$，$s_1^2=0.0457$；$n_2=9$，$\bar{y}=14.9$，$s_2^2=0.0575$

计算得：$\dfrac{s_1^2}{s_2^2}=0.795$

P_值$=2P_{\sigma_1^2=\sigma_2^2\text{时}}\left(\dfrac{S_1^2}{S_2^2}\leq 0.795\right)=2P(F(7,8)\leq 0.795)=2\times 0.39=0.78$

『说明：p_值$=2P(F(7,8)\leq 0.795)$或，$2P(F(7,8)\geq 0.795)$，哪一个小于1，哪一个就是p值。』

现在$\alpha=0.1<0.78$，故接受原假设，认为方差没有显著差异.

补充一例

- 例：两台机床生产同一个型号的滚珠，从甲机床生产的滚珠中抽取8个，从乙机床生产的滚珠中抽取9个，测得这些滚珠的直径(毫米)如下：

甲机床	15.0	14.8	15.2	15.4	14.9	15.1	15.2	14.8	
乙机床	15.2	15.0	14.8	15.1	14.6	14.8	15.1	14.5	15.0

设两机床生产的滚珠直径分别为X,Y,且$X\sim N\left(\mu_1,\sigma_1^2\right),Y\sim N\left(\mu_2,\sigma_2^2\right)$

若现在检验假设 $H_0:2\sigma_1^2\le\sigma_2^2, H_1:2\sigma_1^2>\sigma_2^2(\alpha=0.1)$

(其实也就是检验假设 $H_0:\sigma_1^2/\sigma_2^2\le 1/2, H_1:\sigma_1^2/\sigma_2^2>1/2$)

此检验为单边检验

解一： 当μ_1,μ_2未知时， 检验 $H_0:\sigma_1^2/\sigma_2^2\le 1/2, H_1:\sigma_1^2/\sigma_2^2>1/2$

取检验统计量为：$\dfrac{S_1^2/S_2^2}{1/2}$， 拒绝域形为$D=\left\{\dfrac{s_1^2/s_2^2}{1/2}>C\right\}$

由于$\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

故在$2\sigma_1^2=\sigma_2^2$时， $\dfrac{S_1^2/S_2^2}{1/2}\sim F(n_1-1,n_2-1)$

故得拒绝域为:$\left\{\dfrac{s_1^2/s_2^2}{1/2}>F_\alpha(n_1-1,n_2-1)\right\}$.

查表得：$F_{0.1}(7,8)=2.62$,故拒绝域为:$\left\{\dfrac{s_1^2/s_2^2}{1/2}>2.62\right\}$

本题中 $n_1=8,\ \bar{x}=15.05,\ s_1^2=0.0457$； $n_2=9,\ \bar{y}=14.9,\ s_2^2=0.0575$

计算得：$\dfrac{s_1^2/s_2^2}{1/2}=1.59\le 2.62$

故接受原假设，认为$2\sigma_1^2\le\sigma_2^2$.

解二： 当μ_1,μ_2未知时，检验 $H_0:\sigma_1^2/\sigma_2^2 \le 1/2, H_1:\sigma_1^2/\sigma_2^2 > 1/2$

取检验统计量为：$\dfrac{S_1^2/S_2^2}{1/2}$，拒绝域形为$D=\left\{\dfrac{S_1^2/S_2^2}{1/2}>C\right\}$

由于$\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1,n_2-1)$

故在$2\sigma_1^2=\sigma_2^2$时，$\dfrac{S_1^2/S_2^2}{1/2} \sim F(n_1-1,n_2-1)$

本题中 $n_1=8,\ \bar{x}=15.05,\ s_1^2=0.0457;\ n_2=9,\ \bar{y}=14.9,\ s_2^2=0.0575$

计算得：$\dfrac{s_1^2/s_2^2}{1/2}=1.59$

P_值$=P_{2\sigma_1^2=\sigma_2^2时}\left(\dfrac{S_1^2/S_2^2}{1/2}\ge 1.59\right)=P(F(7,8)\ge 1.59)=0.2644$

现在$\alpha=0.1<0.2644$,

故接受原假设，认为$2\sigma_1^2 \le \sigma_2^2$.