生物统计学

Z-test

侯乃侨 迟小函 袁恩铭

应用的条件

- 总体为服从正态分布或近似正态分布的随机变量
- 知道总体均值
- 已知总体方差或者样本标准误

应用的场景

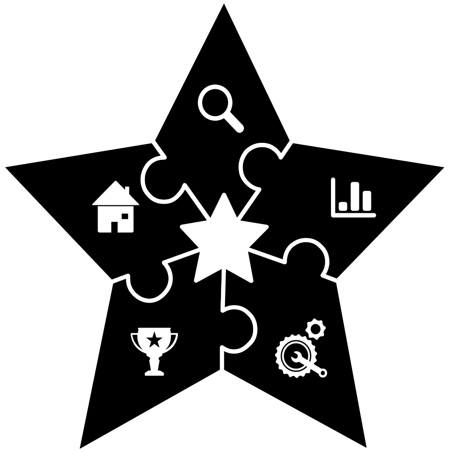
1.检验单个样本与总体均值/proportion

2.检验两个样本的均值/proportion

单样本与总体

When σ is known:

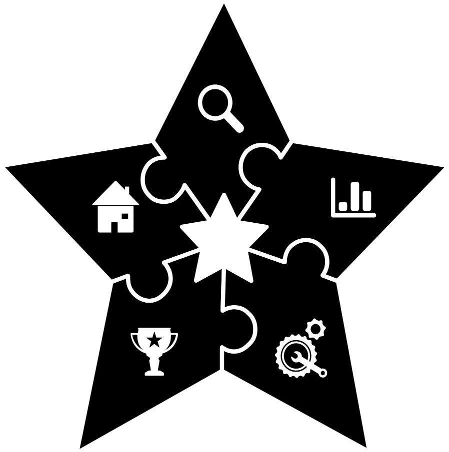
$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$



When the Population σ is not known, large samples ≥ 100 , use the following formula:

$$Z = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

单个样本的proportion



Test Statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

两个样本的Z-test

$$Z = \frac{(mean1 - mean2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

两个样本的Z-test——proportions

- Z-test (recall lecture 4)

$$z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\left(1-\hat{p}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Total number of successes in both samples

$$SE_{D_P} = \sqrt{\hat{p}\left(1-\hat{p}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

Z-test using R

install.packages(“BSDA”)/ install.packages("UsingR")

BSDA包提供了函数z.test()，它可以对基于正态分布的单样本和双样本进行假设检验，其使用方法如下：

```
z.test(x, y = NULL, alternative = "two.sided", mu = 0, sigma.x = NULL, sigma.y = NULL, conf.level = 0.95)
```

If y is NULL, a one-sample z-test is carried out with x. If y is not NULL, a standard two-sample z-test is performed.

使用> help(“z.test”)查看更多

Z-test using R

- 输出结果：

```
> x <- rnorm(50, 0, 5)
> BSDA::z.test(x,sigma.x=5)

        One-sample z-Test

data:  x
z = -0.75437, p-value = 0.4506
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -1.9193214  0.8524863
sample estimates:
 mean of x
-0.5334175
```

Z-test using R

Description

Density, distribution function, quantile function and random generation for the normal distribution with mean equal to mean and standard deviation equal to sd.

Usage

dnorm(x, mean = 0, sd = 1, log = FALSE)

pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)

qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)

rnorm(n, mean = 0, sd = 1)

```
> dnorm(0)
[1] 0.3989423
> qnorm(0.975)
[1] 1.959964
> pnorm(3)
[1] 0.9986501
> rnorm(5,0,5)
[1]  6.480355 -2.459721 -1.050214 -2.096472 -0.842141
```

Z-test using R

单样本z检验函数

```
z.test=function(x,mu,sigma,alternative="two.sided"){
+   n=length(x)
+   result=list()  #构造一个空的list，用于存放输出结果
+   mean=mean(x)
+   z=(mean-mu)/(sigma/sqrt(n))  #计算z统计量的值
+   options(digits=4)  #结果显示至小数点后4位
+   result$mean=mean;result$z=z  #将均值、z值存入结果
+   result$P=2*pnorm(abs(z),lower.tail=FALSE)  #根据z计算P值
+   #若是单侧检验，重新计算P值
+   if(alternative=="greater") result$P=pnorm(z,lower.tail=FALSE)
+   else if(alternative=="less") result$P=pnorm(z)
+   result
+ }
```

Z-test using R

双样本z检验函数

```
z.test2=function(x,y,sigma1,sigma2,alternative="two.sided"){
+   n1=length(x);n2=length(y)
+   result=list()  #构造一个空的list，用于存放输出结果
+   mean=mean(x)-mean(y)
+   z=mean/sqrt(sigma1^2/n1+sigma2^2/n2)  #计算z统计量的值
+   options(digits=4)  #结果显示至小数点后4位
+   result$mean=mean;result$z=z  #将均值、z值存入结果
+   result$P=2*pnorm(abs(z),lower.tail=FALSE)  #根据z计算P值
+   #若是单侧检验，重新计算P值
+   if(alternative=="greater") result$P=pnorm(z,lower.tail=FALSE)
+   else if(alternative=="less") result$P=pnorm(z)
+   result
+ }
```

例题一：

- 有两种方法可用于制造某种以抗拉强度为重要特征的产品。根据以往的资料得知，第一种方法生产出的产品抗拉强度的标准差为8千克，第二种方法的标准差为10千克。从两种方法生产的产品中个抽一个随机样本，样本量分别为n1 = 32,n2 = 40,测得样本一的均值为50千克，样本二的均值为44千克。问这两种方法生产出来的产品平均抗拉强度是否有显著差别（α = 0.05）？

答案：

由于检验两种方法生产出的产品在抗拉强度上是否存在显著差别，并未涉及方向，所以是双侧检验。建立假设并进行检验：

H0：$\mu1 - \mu2 = 0$ 没有显著差别

H1：$\mu1 - \mu2 \neq 0$ 有显著差别

本题中$\sigma1^2$，$\sigma2^2$已知，应选用Z作为检验统计量，由：

$$Z = \frac{(mean1 - mean2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

计算得，Z = 2.83

$\alpha = 0.05$时，$Z_{\alpha/2} = 1.96$

因为$Z > Z_{\alpha/2}$,所以拒绝H0，即两种方法生产出来的产品其抗拉强度有显著差别。

如果计算P值，方法与一个正态总体均值检验中计算P值的方法相同。经计算，此题中的P值为0.004654.在双侧检验中，由于$P<\alpha/2$,故拒绝H0，得到与前面相同的结论。

例题二：

- 2012年各月北京市的新建住宅价格指数是否服从均值为102.4、方差为0.45(标准差为0.67)的正态分布？
- 一房产机构估计2012年各月北京市的新建住宅价格指数：102.5,102.4,102.0,101.8,101.8,102.1,102.3,102.5,102.6,102.8,103.4,104.2

答案：

- bj <- c(102.5,102.4,102.0,101.8,101.8,102.1,102.3,102.5,102.6,102.8,103.4,104.2)
- > BSDA::z.test(x=bj,mu=102.4,sigma.x=0.67,alternative="two.sided")

One-sample z-Test

data: bj

z = 0.68937, p-value = 0.4906

alternative hypothesis: true mean is not equal to 102.4

95 percent confidence interval:

102.1543 102.9124

sample estimates:mean of x 102.5333

THANKS