

KS-检验 (Kolmogorov-Smirnov test)

检验数据是否符合某种分布

KS 是比较一个频率分布 $f(x)$ 与理论分布 $g(x)$ 或者两个观测值分布的检验方法。其原假设 H_0 :两个数据分布一致或者数据符合理论分布。 $D = \max|f(x) - g(x)|$, 当实际观测值 $D > D(n, \alpha)$ 则拒绝 H_0 , 否则则接受 H_0 假设。KS 检验与 t -检验之类的其他方法不同是:

- KS 检验不需要知道数据的分布情况, 可以算是一种非参数检验方法。
- KS 检验的灵敏度没有相应的检验来的高。在样本量比较小的时候。
- KS 检验最为非参数检验在分析两组数据之间是否不同时相当常用。

PS: t -检验的假设是检验的数据满足正态分布, 否则对于小样本不满足正态分布的数据用 t -检验就会造成较大的偏差, 虽然对于大样本不满足正态分布的数据而言 t -检验还是相当精确有效的手段。

参考资料 <https://www.cnblogs.com/arkenstone/p/5496761.html>.

R 语言实现

1.检验指定的数列是否服从正态分布

```
ks.test(rnorm(100), "pnorm")
```

```
##  
## One-sample Kolmogorov-Smirnov test  
##  
## data:  rnorm(100)  
## D = 0.13368, p-value = 0.05608  
## alternative hypothesis: two-sided
```

p 值为 0.5093 大于 0.05 接受原假设, 故该总体服从正态分布。## 2.检验指定的两个数列是否服从相同分布

```
ks.test(rnorm(100), rnorm(50))
```

```
##  
## Two-sample Kolmogorov-Smirnov test  
##  
## data:  rnorm(100) and rnorm(50)  
## D = 0.1, p-value = 0.8846  
## alternative hypothesis: two-sided
```

p 值为 0.6137>0.05 接受原假设, 故两总体服从相同分布。

python 语言实现

加载相关包

```
from scipy.stats import kstest
import numpy as np
```

1.检验指定的数列是否服从正态分布

```
x = np.random.normal(0,1,1000)
test_stat = kstest(x, 'norm')
print(test_stat)
```

```
## KstestResult(statistic=0.03157672969624237, pvalue=0.26657851993623344)
```

可得 p 值为 $0.7 > 0.05$ 接受原假设，故该分布服从正态分布，

2.检验指定的两个数列是否服从相同分布

```
from scipy.stats import ks_2samp
beta=np.random.beta(7,5,1000)
norm=np.random.normal(0,1,1000)
ks_2samp(beta,norm)
```

```
## Ks_2sampResult(statistic=0.585, pvalue=1.0217401968903905e-150)
```

可得 p 值为很小，故拒绝原假设，两分布不是相同的分布。