Unidade 2 – Testes para uma amostra

Teste de Kolmogorov - Smirnov

É um teste de aderência, tendo maior utilização em testar normalidade.

Procura especificar a distribuição de freqüência **acumulada** teórica e compará-la com a distribuição de freqüência **acumulada** observada. Em algum ponto essa diferença será maior e então utilizada para o teste. Este teste pode ser aplicado para pequenas amostras

Hipóteses:

H_0: não há diferença nas frequências esperadas em relação as observadas;

H_1: as frequências esperadas diferem-se das observadas.

Procedimento:

a) Especificar F(x): frequência observada teórica acumulada sob H_0. Isso depende da distribuição que testamos (Uniforme, Poisson, Binomial, Normal...);
b) Distribuir os escores observados S(x) também de forma acumulada e correspondente com cada valor de F(x). S(x) pode ser calculado como k/n, sendo k o numero de observações iguais a x;
c) Para cada caso da distribuição acumulada, calcular:

$$D = \max\{|F(\text{x}) - S(x)|\}$$

d) Mediante tabela, concluir sobre o teste. Rejeitaremos H_0 se o D que calculamos exceder o valor tabelado.

Exemplo: O contrato de um restaurante que fornece refeições para uma determinada instituição garante uma média de 290 gramas de carne por refeição, com um desvio padrão de 56 gramas. Em uma amostra de 10 refeições foram aferidas as seguintes quantidades de carne:

198	254	262	272	275	278	285	287	287	292

Podemos afirmar que essa amostra é proveniente de uma população com distribuição normal com média 290 gramas e desvio 56 gramas?

Resolução:

$$\begin{cases} H_0: \text{ é proveniente de uma normal} \\ H_1: \text{não é proveniente de uma normal} \end{cases}$$

O que queremos nesse caso, é simplesmente testar a normalidade. As frequências observadas já temos com os dados. Observe que é uma distribuição contínua, logo cada dado será 1 ocorrencia em k que é 10 (1/10). Dessa forma o primeiro S(x) é 0.1. Se segundo 0.2, pois aqui devemos sempre ir acumulando. Observe como ficará no Quadro 1 abaixo.

Precisaremos encontrar a distribuição acumulada teórica para depois calcularmos as diferenças. Isso pode ser feito padronizando os valores e obtendo as probabilidades acumuladas na tabela da distribuição normal.

Para o primeiro valor teríamos:

$$Z_{198} = \frac{(198-290)}{56} = -1.64$$

Olhando na tabela, o total acumulado até -1.64 é F(x) = 0.5-0.44950 = 0.0505

Para o segundo valor:

$$Z_{254} = \frac{(254-290)}{56} = -0.64$$

Na tabela, o acumulado será F(x) = 0.5-0.23891 = 0.2611.

Repetimos o processo até a ultima observação. No fim teremos:

Quadro 1: Teste de Kolmogorov Smirnov para Normalidade

X	Z	F(x)	S(x)*	$\lvert F(x)-S(x)\rvert$
198	-1.64	0.0505	0.1	0.0495
254	-0.64	0.2611	0.2	0.0611
262	-0.50	0.3085	0.3	0.0085
272	-0.32	0.3745	0.4	0.0255
275	-0.27	0.3936	0.5	0.1064
278	-0.21	0.4168	0.6	0.1832
285	-0.09	0.4641	0.7	0.2359
287	-0.05	0.4801	0.8	0.3199
287	-0.05	0.4801	0.9	0.4199
292	0.04	0.5160	1	0.4840

*Note que vamos acumulando de 0.1 e 0.1. Porque temos 1 ponto no total de 10 em cada linha.

Agora fazemos:

$$D = \max\{|F(x) - S(x)|\}$$

Analisando o Quadro 1, percebemos que D = 0.4840.

Pela Tabela do teste para N = 10 e α=0.05:

Teste de Kolmogorov-Smirnov – Uma Amostra

Valores Críticos de D

N \ α	0,20	0,15	0,10	0,05	0,01
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,326	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,250	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,231	0,246	0,264	0,294	0,356
25	0,210	0,220	0,240	0,270	0,320
30	0,190	0,200	0,220	0,240	0,290
35	0,180	0,190	0,210	0,230	0,270
> 35	$\frac{1,07}{\sqrt{N}}$	$\frac{1,14}{\sqrt{N}}$	$\frac{1,22}{\sqrt{N}}$	$\frac{1,36}{\sqrt{N}}$	$\frac{1,63}{\sqrt{N}}$

Logo, como 0.48 (calculado) é maior que 0.41 (tabelado), rejeitamos H_0 e não podemos concluir que existem evidencias de que a amostra seja proveniente de uma distribuição normal com média 290 e desvio padrão 56.