Mood's median test

Function of the test

Median test is used to determine wheather two or more samples drawn from their respective popⁿs have same median or not

Two independent sample ($k=2$)	More than two independent samples ($k \geq 3$)
Median test is alternative test to Mann-Whitney U test.	Median test is alternative to Kruskal Wallis H test.

Median test is less powerful than its alternative test

Test Assumptions

1. Samples are drawn randomly and independently from their respective popⁿs.
2. Measurement scale in at least ordinal
3. The popⁿs have some basic shape

Hypothesis to test

H_0 : $M_1 = M_2 = \cdots = M_k$ (Median of k ind. sample are same)

H_1 : At least median one popⁿ is different from the median of at least one popⁿ.

Test Statistics

Step 1: Combine k independent samples into one. (Pooling of samples)

Step 2: Find the grand median of combined sample.

Step 3: Construct 2×k contingency table.

No. of obs.	Sample 1	Sam. 2	...	Sample k	Total
Greater than GM	O_{11}	O_{21}	...	O_{1k}	R_1
Lesser or equal to GM	O_{21}	O_{22}	....	O_{2k}	R_2
Total	C_1	C_2	....	C_k	N

Now, the test becomes chi-square test of independence. Test statistics is given by,

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad i = 1, 2. \quad j = 1, 2, \text{——}, k$$

$$d.f = (R-1)(C-1) = (2-1)(k-1) = k-1$$

Numerical

\# A large store accepts card, cheque or cash as payment for merchandise purchased at the store. The manager of the store wants to know if

any form of purchase is higher per order than others.
. A random samples from sales gave the following data.

Card	140	162	95	125	130	132	110	98		
Cheque	135	126	101	110	140	150	103			
Cash	80	95	82	105	103	112	78	68	70	90

Use the median test to know whether the median amount of purchase is same for three mode of payment.
Take $\alpha = 5\%$

Soln:-

Step 1: Null and Alternative Hypothesis

Let M_1, M_2 and M_3 be the median amount of purchase using card, cheque and cash respectively.

$H_0: M_1 = M_2 = M_3$ (Median amount of purchase is same for three modes of payment)

H_1: Median amount of purchase for at least one mode of payment is diff. from median of at least one other mode.

Step 2: $\alpha = 5\%$

Step 3: Test Statistics

We have 2×3 table, $k = 3$,

No. of obs.	Sample 1	Sample 2	Sample 3	Total
Greater than G.M	O_{11}	O_{12}	O_{13}	R_1
Less than or equal to G.M	O_{21}	O_{22}	O_{23}	R_2
Total	C_1	C_2	C_3	H

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad , i = 1, 2, \quad j = 1, 2, 3 \ldots$$

Step 4 :- Calculated χ^2

Combined sample

140 152 95 125 130 132 110 98 138
126 101 110 140 150 108 80 95 82 105
108 112 78 68 70 90

Sorted Data

68 70

Grand median = size of $\left(\frac{n+1}{2}\right)^{th}$ ordered data

$= \text{size of } \left(\frac{25+1}{2}\right)^{th}$ ordered data

$= 108$

Now, we have,

No. of obs.	Sample 1	Sample 2	Sample 3	Total
Greater than GM	6	5	1	12
Lesser than or equal to GM	2	2	9	13
Total	8	7	10	n = 25

Now,

SN	Row, col	O_{ij}	$E_{ij} = \frac{R_i \times C_j}{n}$	$\chi^2 = (O_{ij} - E_{ij})^2 / E_{ij}$
1	(1,1)	6	0.1667	
2	(1,2)	5	0.4	
3	(1,3)	1	3	
4	(2,1)	2	1	
5	(2,2)	2	2	
6	(2,3)	9	0.667	cal $\chi^2 = 9.66$
		$\Sigma O_{ij} = 25$	$\Sigma E_{ij} = 7.2337$	

Step 5 :- Tabulated χ^2

$\chi^2_{0.05}(2) = 5.99$

$d.f = k-1 = 3-1 = 2$

Step 6 :- Statistical Decision

Since, cal $\chi^2 =$ Tab $\chi^2 = 5.99$. So, we reject Ho.

Step 7:- Conclusion

Median amount of purchase for at least one mode of payment is diff^n from median of at least one other mode of payment.

Date
Page

Kolmogorov Smirnov Test

Function of the test

The Kolmogorov Smirnov test is also called as K-S 'D' test. It is basically goodness of fit test.

K-S test can be used in,

(1) One-sample situation
(Discrete OR continuous)

(2) Two sample situation
(Discrete or Continuous)

In One sample situation, it determines how well a hypothesized frequency distribution $F_T(x)$ fits to the observed or empirical frequency distribution $F_o(x)$.

K-S test is more powerful and stingent test than chi-square goodness of fit test.

Test Assumptions

(1) Sample is drawn randomly from population
(2) Hypothesized frequency distn is either discrete or continuous.

3. Hypothesized frequency distn is specified in advance.

Hypothesis to test

H_0: The data follow the specified distribution.
$(H_0 : F_T(x) = F_0(x))$

H_1: The data do not. " " "
$(H_1 : F_T(x) \neq F_0(x))$

Test Statistics

Continuous data

We find D statistic as follows:-

$D_i \leq | F_0(x_i) - F_T(x_i) |$

$D_i' = | F_0(x_{i-1}) - F_T(x_i) |$

$i = 1, 2, \ldots, n$

$D = \max \{ \mathrm{Max}\ D_i, \mathrm{Max}\ D_i' \}$

Discrete data

$D_i = \mathrm{Max} \{ | F_0(x_i) - F_T(x_i) | \} \ \forall\ i = 1, 2, \ldots, k$

Statistical Decision

Numerical

⇒ 1 sample discrete case

A dice is rolled for 60 times and following outcomes were obtained.

Side	Frequency
1	8
2	9
3	13
4	7
5	15
6	8
Total	$n = 60$

Test the hypothesis that the die is fair. (All sides have equal chance of appearing). Use K–S test at 5% level of significance.

Soln:-

Step 1: Null and Alternative Hypothesis

H_0 : Dice is fair (Frequencies are according to uniform distribution)

H_1 : Dice is not fair (freq. are not " " ")

Step 2: Level of significance

$\alpha = 5\%$ (given)

Step 3: Test Statistic

The D statistic is given,

Discrete Data

$D_i = \text{Max}\{|F_o(x_i) - F_T(x_i)|\} \ \forall \ i = 1, 2, \ldots, k$

Step 4: Calculated D

Side	F_o	F_e	$F_o(x)$	$F_T(x)$	$D = \|F_o(x) - F_T(x)\|$
1	8	10	8	10	2
2	9	10	17	20	3
3	13	10	30	30	0
4	7	10	37	40	3
5	15	10	52	50	2
6	8	10	60	60	0
Total	$\Sigma F_o =$ 60	$\Sigma f_e =$ 60			

$\therefore$ Cal D = 3

Step 5:- Tabulated D

The test is right sided.

$k = 6$

$n = 60$

$D_{0.05}(6, 60) = 9$

Step 6: Statistical Decision

Since, cal $D = 3 <$ critical $D = 9$,

We donot reject H_0 at 5% level of significance.

Step 7: Conclusion

Dice is fair.

For interpolation case, (linear)

$a = 30$ $c = 40$ $b = 60$

$A = 6$ $C = 7$ $B =$

$D(k=6, n=60) = 9$

$D(k=6, n=30) = 6$

$D(k=6, n=40) = ?$

$$C = A + \frac{c-a}{b-a}(B-A)$$

$$= 6 + \frac{40-30}{60-30}(9-6)$$

$$= 7$$

Sample

$\chi^2_{0.05}(40) = 55.76$

$\chi^2_{0.05}(60) = 79.08$

$\chi^2_{0.05}(45) = ?$

$a = 40$ $c = 45$ $b = 60$

$A = 55.76$ $C = ?$ $B = 79.08$

$$C = A + \frac{c-a}{b-a}(B-A)$$

$$= 55.76 + \frac{5}{20}(79.08 - 55.76)$$

$$= 61.59$$

Two Sample Kolmogorov Smirnov test

K-S test can also be used in two sample situations and the objective is to compare two distributions, and determines whether two samples come from the same distribution or not. Comparing distributions can be either both continuous or both discrete.

Test Assumptions

1. Two samples are drawn randomly from $pop^{n}s$.
2. Two comparing distribution can be discrete or continuous.

Hypothesis

H_0: Both sample come from a pop^{n} with the same distribution i.e. $PDF_1 = PDF_2$

H_1: Two samples come from $diff^{n}$ distributions i.e. $PDF_1 \neq PDF_2$

classmate
Date
Page

Test Statistics

Suppose that the first sample has size m with an observed cumulative distribution function of F(x) and that the second sample has size n with an observed cumulative distribn fxn of G(x). Test statistic is the maximum absolute difference betn two cumulative prob. distribn's.

So, the test statistic is given by,

$$D = Max\,|F(x) - F(y)|$$

i)
ii)
(iii)

Numerical

Following reasons were obtained in the survey regarding the questions of confidence level in purchasing good from online store and retail store.

Answer Categories	Online Store	Retail Store
non-confident	20	4
slightly confident	30	27
somewhat confident	13	28
Confident	20	18
very confident	41	47
Total	124	124

Are these two distn's. significantly diffn from each other? Use K-S test at 5% level of significance.