Testing of Hypothesis

A quick reference to work on the applications of various **Tests of Significance**

Department of Statistics
P.R. Government College (A)
Kakinada

Parametric Tests

Tests for Mean(s) (for variables)

Large Samples ($n \geq 30$)

Single Mean: (Z-test)

NH: The difference between hypothetical mean and sample mean is not significant.

Test formula: $Z = \dfrac{\bar{x}-\mu}{[\sigma/\sqrt{n}]}$

Small Samples ($n<30$)

Single Mean: (t-test)

NH: The difference between hypothetical mean and sample mean is not significant.

Test formula: $t = \dfrac{\bar{x}-\mu}{[s/\sqrt{(n-1)}]}$

$df = n-1$

Application areas:

Testing average marks, height, sales, yield of a crop, etc in a population

Tests for Mean(s) (for variables)

Large Samples (n≥30)

Two Means: (Z-test)

NH: The difference between two sample means is not significant.

Test formula:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{[\sigma_1^2/n_1 + \sigma_2^2/n_2]}}$$

Small Samples (n<30)

Two Means: (t-test)

NH: The difference between two sample means is not significant.

Test formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{[1/n_1 + 1/n_2]}}$$

Where,

$$S = \sqrt{[(n_1 s_1^2 + n_2 s_2^2)/(n_1 + n_2 - 2)]}$$

$df = n_1 + n_2 - 2$

Application areas:

Testing the significance difference between average marks, heights, sales, yields of a crop, etc in two different populations.

Tests for Mean(s) (for variables)

Small Samples (n<30)

Two Dependent Means: (paired t-test)

NH: The difference between means of samples of two populations of same units before and after some activity is not significant.

Test formula:

$$t = d' \sqrt{(n-1)}/s_d$$

where, $d = x_1 - x_2$

df = n-1

Application areas:

Testing the significance difference between average marks of students before and after extra coaching, , average heights of children before and after extra vitamin diet , average sales before and after advertisement , average yields of a crop before and after applying a special fertilizer, etc

Tests for proportions(s) (for attributes)

Large Samples ($n \geq 30$)

Single Proportion: (Z-test)

NH: The difference between hypothetical proportion and sample proportion is not significant.

Test formula: $Z = \dfrac{p - P}{\sqrt{[PQ/n]}}$

Application areas:

Testing the proportion with respect to an attribute like, gender, drinking, attacking a disease, mode of education, etc in a population

Two Proportions: (Z-test)

NH: The difference between two sample proportions is not significant.

Test formula:

$$Z = \frac{p_1 - p_2}{\sqrt{[pq(1/n_1 + 1/n_2)]}}$$

where, $p = (x_1+x_2)/(n_1+n_2)$

Application areas:

Testing the significance difference between proportions with respect to an attribute like, gender, drinking, attacking a disease, mode of education, etc in two populations

Tests for variance(s) (for variables)

Large Samples ($n \geq 30$)

Single S.D: (Z-test)

NH: The difference between hypothetical SD and sample SD is not significant.

Test formula: $Z = \dfrac{s-\sigma}{[\sigma/\sqrt{2n}]}$

Small Samples ($n<30$)

Single Variance: (χ^2-test)

NH: The difference between hypothetical variance and sample variance is not significant.

Test formula: $\chi^2 = \dfrac{ns^2}{\sigma^2}$

df = n-1

Application areas:

Testing the variance in marks, height, sales, yield of a crop, etc in a population

Tests for Variances(s) (for variables)

Large Samples (n≥30)

Two SD's: (Z-test)

NH: The difference between two sample SD's is not significant.

Test formula:

$$Z = \frac{s_1 - s_2}{\sqrt{[s_1^2/2n_1 + s_2^2/2n_2]}}$$

Small Samples (n<30)

Two Variances: (F-test)

NH: The difference between two sample variances is not significant.

Test formula:

$$F = \frac{S_1^2}{S_2^2}$$

where,

$$S^2 = ns^2/(n-1)$$

df = $(n_1 - 1, n_2 - 1)$

Application areas:

Testing the significance difference between variances in marks, heights, sales, yields of a crop, etc in two different populations.

χ^2-tests (for attributes)

Small Samples (n<30)

Goodness of fit:

NH: The difference between observed and expected frequencies is not significant.

Test formula: $\chi^2 = \sum[(Oi-Ei)^2/Ei]$

df = n-1

Application areas:

Testing the theoretical assumptions like the figures from any probability law, figures related to genetics , etc

Independence of attributes:

NH: The two attributes are independent

Test formula:

$$\chi^2 = \sum\sum\{[(AB)_o-(AB)_e]^2/(AB)_e\}$$

df = (m-1)(n-1)

Application areas:

Testing the association between any two attributes like vaccination and curing a disease, lockdown and controlling corona, education and improvement in technology, etc

Non-Parametric Tests

(Appropriate for only nominal and ordinal data)

Sign test

One Sample Sign Test:

NH: The median of the population is M_o (a specified value).

Test formula:

$$p = (1/2^n)\sum {}^nc_u$$

where, $u = \min\{u^+, u^-\}$

u^+ = no. of observations $> M_o$

u^- = no. of observations $< M_o$

$n = u^+ + u^-$

If $n>30$, May apply Z-test (approx)

$$Z = (u-n/2)/\surd(n/4)$$

Application areas:

Testing the average (median) marks, height, weight, income, etc (testing the data with two levels)

Paired Sample Sign Test:

NH: The distributions of two populations are equal.

Test formula:

$$p = (1/2^n)\sum {}^nc_u$$

where, $u = \min\{u^+, u^-\}$

u^+ = no. of d's > 0

u^- = no. of d's < 0

$d = x-y$

$n = u^+ + u^-$

If $n>30$, May apply Z-test (approx)

$$Z = (u-n/2)/\surd(n/4)$$

Application areas:

Testing the significance difference between distributions of marks, heights, income, etc of two populations of same units

Signed Rank test

(an alternative to Sign test)

One Sample Signed Rank Test:

NH: The median of the population is M_o (a specified value).

Test formula:

$$R = \text{Min}\{R^+, R^-\}$$

where,

R^+ = Sum of ranks of +ve deviations (x-Mo)

R^- = Sum of ranks of -ve deviations(x-Mo)

If n>30, May apply Z-test (approx)

$$Z = \frac{R - n(n+1)/4}{\sqrt{[n(n+1)(2n+1)/24]}}$$

Application areas:

Testing the average (median) marks, height, weight, income, etc (testing the data with two levels)

Paired Sample Signed Rank Test:

NH: The distributions of two populations are equal.

Test formula:

$$R = \text{Min}\{R^+, R^-\}$$

where,

R^+ = Sum of ranks of +ve deviations (x-y)

R^- = Sum of ranks of -ve deviations(x-y)

If n>30, May apply Z-test (approx)

$$Z = \frac{R - n(n+1)/4}{\sqrt{[n(n+1)(2n+1)/24]}}$$

Application areas:

Testing the significance difference between distributions of marks, heights, income, etc of two populations of same units

Run test

(by Wald-Wolfowitz)

One Sample Run Test(or Test for Randomness) :

NH: The sample is drawn at random.

Test formula:

r = no. of runs

here, runs are counted from the sequence of sample in given order by observing that they are > or < median

If $n>25$, May apply Z-test (approx)

$$Z = \frac{r - (n+2)/2}{\sqrt{[n(n-2)/4(n-1)]}}$$

Application areas:

Testing the randomness of any observed sample which is an essential principle in all statistical analyses

Two Samples Run Test:

NH: The distributions of two populations are equal.

Test formula:

r = no. of runs

here, runs are counted from the sequence of combined ordered sample by observing that they are of 1st sample or 2nd sample

If $n_1,n_2>10$, May apply Z-test (approx)

$$Z = \frac{r - \{[2n_1n_2/(n_1+n_2)]+1\}}{\sqrt{[2n_1n_2(2n_1n_2-n_1-n_2)/(n_1+n_2)^2(n_1+n_2-1)]}}$$

Application areas:

Testing the significance difference between distributions of marks, heights, income, etc of two populations

U- test and Median test

U-Test for two samples

(Wilcoxon-Mann-Witney U-test) :

NH: The distributions of two populations are equal.

Test formula:

$$U = \min\{U_1, U_2\}$$

where, $U_1 = n_1 n_2 + n_1(n_1+1)/2 - R_1$

$U_2 = n_1 n_2 + n_2(n_2+1)/2 – R_2$

R_1 , R_2 = Sum of the ranks of 1st and 2nd sample values in combined ordered sample.

If $n_1, n_2 > 8$, May apply Z-test (approx)

$$Z = \frac{U – n_1 n_2/2}{\sqrt{[n_1 n_2(n_1+n_2+1)/12]}}$$

Application areas:

Testing the significance difference between distributions of marks, heights, income, etc of two populations

Median test for two Samples:

NH: The distributions of two populations are equal.

Test formula:

$$p = \sum_{m1=1} \{ {}^{n1}c_{m1} * {}^{n2}c_{m2} / {}^{(n1+n2)}c_{(m1+m2)} \}$$

where, m1, m2 = no. of observations of 1st and 2nd sample > median of combined ordered sample

If $n_1, n_2 > 10$, May apply χ^2-test (approx)

$$\chi^2 = \frac{N[|ad-bc|-N/2]^2}{(a+b)(c+d)(a+c)(b+d)}$$

here, $a = m_1$, $b=m_2$, $c=n_1-m_1$, $d=n_2-m_2$

Application areas:

Testing the significance difference between distributions of marks, heights, income, etc of two populations

Notations:

-> NH – Null Hypothesis

-> n = sample size

-> $\overline{x}$ = sample mean=$\sum x/n$

-> s = sample SD = $\sqrt{[\sum x^2/n - \overline{x}^2]}$

-> √ - square root

-> μ= population mean

-> σ= population SD

-> p = sample proportion = x/n

-> P = population proportion

-> Q = 1-P

-> O = $(AB)_o$= observed frequency

-> E = $(AB)_e$ =expected frequency