some fixed number, then the calculated sample median. In the latter case, any observation equal to this fixed number are ignored in the analysis and the sample size is adjusted accordingly. The run test can be used for one sided or two sided alternatives. If the alternative hypothesis is simply non randomness a two sided test should be used since the presence of a trend ~~color~~ would usually be indicated by a clustering of similar objects which is reflected by an unusually small no. of runs. A one sided test is more appropriate for trend alternatives.

~~H~~ Example

The following are the marks secured by two batches of salesman in the final test taken after the completion of training.

Use U-test and Run test for the null hypothesis that the samples are drawn from identical distribution against the alternative that the distributions differ in location only.

Batch A (x): 26, 27, 31, 26, 19, 21, 20, 25, 30

Batch B (Y) 23, 28, 26, 24, 22, 19

Seq	19	19	20	21	22	23	24	25	26	26	26	27	28	30	31	run
1	X	Y	X	X	Y	Y	Y	X	Y	X	X	X	Y	X	X	9
2	X	Y	X	X	Y	Y	Y	X	X	Y	X	X	Y	X	X	9
3	X	Y	X	X	Y	Y	Y	X	X	X	Y	X	Y	X	X	9
4	Y	X	X	X	Y	Y	Y	X	Y	X	X	X	Y	X	X	8
5	Y	X	X	X	Y	Y	Y	X	X	Y	X	Y	Y	X	X	8
6	Y	X	X	X	Y	Y	Y	X	X	X	Y	X	Y	X	X	8

run = 9

$\therefore$ $n_1 = 9$

$n_2 = 6$

critical value at $n_1 = 9$, $n_2 = 6$ is 4

so, $9 > 4$

we accept the null hypothesis

$U = 1+1+4+4+4+5+6+6$

$= 31$

$U' = n_1 n_2 - U$

$= 54 - 31 = \cancel{24} = 23$

Critical value of U test at 0.02 level of significance 7

$\therefore$ $\cancel{24 > 7}$ $\underline{23 > 7}$

So, we accept the null hypothesis.

Then, samples are drawn

Median Test

Let $X_1, X_2, \ldots, X_{n_1}$ be a random sample of size n_1 drawn from a population with continuous dist^{n} func^{n} $F(\cdot)$ and $Y_1, Y_2, \ldots, Y_{n_2}$ be a random sample of size n_2 drawn from a popu^{n} with continuous distribution func^{n} $G(\cdot)$, $G(x) = F(x-\delta)$, $\delta \in \mathbb{R}$.
The samples are drawn independently of each other.
Consider the problem of testing,

$$H_0 : \delta = 0 \quad \text{against} \quad H_1 : \begin{cases} \delta > 0 \\ \delta < 0 \\ \delta \neq 0 \end{cases}$$

- **Test Procedure** : Let $\underset{\sim}{Z} = (X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, \ldots, Y_{n_2})$ denote the combined sample. Let R_i denote the rank of the ith individual in the combined sample. R_i, $(i = 1(1)\, n_1 + n_2)$

Now consider a indicator function

$$\psi(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x < 0 \end{cases}$$

Now let us consider the statistic,

T = No. of 2nd sample observations exceeding the combined sample median, $\tilde{\theta}$.

$$= \sum_{j=1}^{n_2} \psi(Y_j - \tilde{\theta}) = \sum_{i=n_1+1}^{n} \psi\left(R_i - \left[\frac{n+1}{2}\right]\right),$$ where $n = n_1 + n_2$, total no of observations

and $\left[\frac{n+1}{2}\right]$ = Rank of the combined sample median.

Thus T can be interpreted as the no. of second sample ranks exceeding $\left[\frac{n+1}{2}\right]$. T is called the median test statistic.

Non-parametric justification :

Under H_0, $\underset{\sim}{R} = (R_1, R_2, \ldots, R_n)'$ has a uniform distribution over the set of $n!$ realisation of $(1, 2, \ldots n)$. For this we consider the following result.

$\boxed{\text{Result}}$: Let $X_1, X_2, \ldots, X_n$ be iid with continuous d.f. $(F(\cdot))$. Let R_i = Rank of X_i = No. of observations $\leq X_i$

$\underset{\sim}{R} = (R_1, R_2, \ldots, R_n)'$ = Vector of ranks.

Then
$$P(\underset{\sim}{R} = \underset{\sim}{r}) = \begin{cases} \frac{1}{n!} & \forall\ \underset{\sim}{r} = \{r_1, r_2, \ldots r_n\} \text{ of } \{1, 2, \ldots n\} \\ 0 & \text{o.w} \end{cases}$$

Proof >> Let $X_{(1)} < X_{(2)} \cdots < X_{(n)}$ be the order statistics.

The $\{X_1, X_2, \ldots, X_n\} \Leftrightarrow \{X_{(1)}, X_{(2)}, \ldots, X_{(n)}, R_1, R_2 \ldots R_n\}$

$\underset{\sim}{R} = (r_1, r_2, \ldots, r_n)$ is nothing but a permutation of $(1, 2, \ldots, n)$.

Given $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$, $(X_1, X_2, \ldots, X_n)$ has $n!$ possible permutations.

So since $X_1, X_2, \ldots X_n$ are iid and $F(\cdot)$ is continuous, these permutations are equally likely.

Example

7	$R_1 = 4$
11	$R_2 = 2$
15	$R_3 = 1$
18	$R_4 = 3$

original sample

18, 11, 7, 15

So, the conditional distribution of $(X_1, X_2, \dots, X_n)$ given $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ is $\frac{1}{n!}$

$\therefore P(\underset{\sim}{R} = \underset{\sim}{r} \mid X_{(1)}, X_{(2)}, \dots, X_{(n)}) = \frac{1}{n!}$, which is independent of the conditioning variables.

Hence, $P(\underset{\sim}{R} = \underset{\sim}{r}) = \begin{cases} \frac{1}{n!} & \forall\ \underset{\sim}{r} = \{r_1, r_2, \dots, r_n\} \text{ of } \{1, 2, \dots, n\} \\ 0, & \text{otherwise.} \end{cases}$

Therefore the distn of $\underset{\sim}{R}$ is independent of F, whenever F is continuous. Hence T, being a function of $\underset{\sim}{R}$ has its distribution independent of F under Ho. So the test provided by T is exactly distn free under Ho and hence non-parametric.

Critical Region : $\delta > 0$ implies that the second sample observations are expected to be larger than the first sample observations, i.e. the second sample ranks are expected to be larger under H_1, $\delta > 0$ than that under H_0, $\delta = 0$. So T is expected to be larger under H_1, $\delta > 0$ than that under H_0. Therefore a right tailed test is appropriate for testing $H_0 : \delta = 0$ against $H_1 : \delta > 0$

Similarly a left tail test is appropriate for testing $H_0 : \delta = 0$ against $H_1 : \delta < 0$

and a two tailed test is appropriate for testing $H_0: \delta = 0$ against $\delta \neq 0$

Alternative hypothesis	Critical Region	p-value
$\delta > 0$ or $\theta_2 > \theta_1$	$T \geq c_\alpha$	$P(T \geq T_0)$, T_0 :- observed value of the test statistic
$\delta < 0$ or $\theta_2 < \theta_1$	$T \leq c'_\alpha$	$P(T \leq T_0)$
$\delta \neq 0$ or $\theta_2 \neq \theta_1$	$T \geq c$ or $T' \leq c'$	2. Minimum of the above two

Null distribution, Expectation and Variance

$T = t$ = test statistic

$\longleftarrow n_2 - t \longrightarrow \longleftarrow t \longrightarrow$

$\left[\frac{n+1}{2}\right]$

No. of Y obs. $= n_2$

$\longleftarrow \left[\frac{n+1}{2}\right] \longrightarrow \longleftarrow n - \left[\frac{n+1}{2}\right] \longrightarrow$

obs.   obs.

$$P_{H_0}(T=t) = \frac{\text{Number of choices of } (R_{n_1+1}, R_{n_1+2}, \ldots, R_n) \text{ out of } (1, 2, \ldots, n) \text{ such that } \sum_{j=n_1+1}^{n} \psi\left(R_j - \left[\frac{n+1}{2}\right]\right) = t}{\binom{n}{n_2}}$$

$$= \frac{\binom{\left[\frac{n+1}{2}\right]}{n_2 - t} \cdot \binom{n - \left[\frac{n+1}{2}\right]}{t}}{\binom{n}{n_2}}$$

[Above the combined sample median, there are $n - \left[\frac{n+1}{2}\right]$ observations and t Y observations can be arranged there in $\binom{n - \left[\frac{n+1}{2}\right]}{t}$ ways. Remaining $(n_2 - t)$ Y observation can be arranged among $\left[\frac{n+1}{2}\right]$ observations (less than or equal to the combined median) is $\binom{\left[\frac{n+1}{2}\right]}{n_2 - t}$ ways]

$\therefore$

$$\left[\frac{n+1}{2}\right] \geq n_2 - t$$

$$t \geq n_2 - \left[\frac{n+1}{2}\right]$$

$$t \leq n - \left[\frac{n+1}{2}\right]$$

$$t \leq n_2$$

Range

$$t = \text{Max}\left(0,\ n_2 - \left[\frac{n+1}{2}\right]\right) \quad (1)$$

$$\text{Min}\left(n_2,\ n - \left[\frac{n+1}{2}\right]\right)$$

$$\frac{\binom{NP}{x}\binom{N-NP}{n-x}}{\binom{N}{n}} \text{ pmf}$$

$$E(x) = nP$$

$$E_{H_0}(T) = \frac{\left(n - \left[\frac{n+1}{2}\right]\right) n_2}{n} \simeq \frac{n_2}{n} \text{ for large } n$$

$N = n$, $n = n_2$

$$NP = n - \left[\frac{n+1}{2}\right]$$

$$P = \frac{n - \left[\frac{n+1}{2}\right]}{n}$$

$$E(x) = n_2 \cdot \frac{n - \left[\frac{n+1}{2}\right]}{n}$$

$$V_{H_0}(T) = \frac{(n - n_2)}{(n-1)} \left[\frac{n - \left[\frac{n+1}{2}\right]}{n}\right] \left[\frac{\left[\frac{n+1}{2}\right]}{n}\right] n_2$$

$$\simeq \frac{n_1 n_2}{4n} \text{ for large } n.$$

Large Sample distribution and test :

Suppose for each n, there are $n_1 = n_1(n)$ and $n_2 = n_2(n)$ such that as $n \to \infty$,

(i) $n_1, n_2 \to \infty$

(ii) $\frac{n_1}{n} \to \lambda$ & $\frac{n_2}{n} \to 1-\lambda$, $\lambda \in (0,1)$.

Then under H_0, as $n \to \infty$

$$\frac{T - \frac{n_2}{n}}{\sqrt{\frac{n_1 n_2}{4n}}} \xrightarrow{\mathcal{D}} N(0,1)$$

H_1	critical region (for asymtotic size-α test)
$\delta > 0$	$\frac{T - n_2/n}{\sqrt{\frac{n_1 n_2}{4n}}} > \tau_\alpha$
$\delta < 0$	$\frac{T - \frac{n_2}{n}}{\sqrt{\frac{n_1 n_2}{4n}}} < -\tau_\alpha$
$\delta \neq 0$	$\left\lvert \frac{T - \frac{n_2}{n}}{\sqrt{\frac{n_1 n_2}{4n}}} \right\rvert > \tau_{\alpha/2}$

18.20 : X : 13, 10, 9, 12, 11, 10, 8

Y : 10, 11, 12, 13, 9, 11, 14, 12, 13

Combined sample

8	9	9	10	10	10	11	11	11	12	12	12	13	13	13	14
X	X	Y	X	X	Y	X	Y	Y	Y	Y	X	X	Y	Y	Y

$n_1 = 7$, $n_2 = 9$

$n = 16$

Combined sample median

$= \left[\frac{n+1}{2}\right]$

$= \left[\frac{16+1}{2}\right] = 8$

$T_0 = 5$

$H_0 : \delta = 0$

against

$H_1 : \delta > 0$

Critical region : $T \geq C_\alpha$

p value.

$P(T \geq T_0)$

$$P(T=5) = \frac{\binom{8}{4}\binom{8}{5}}{\binom{16}{9}}$$

$\cancel{= 0.343}$

$$P(T=6) = \frac{\binom{8}{3}\binom{8}{6}}{\binom{16}{9}} \quad \cancel{= 0.137}$$

$$P(T=7) = \frac{\binom{8}{2}\binom{8}{7}}{\binom{16}{9}} \cancel{= 0.0195}$$

$$P(T=8) = \frac{\binom{8}{1}\binom{8}{8}}{\binom{16}{9}} \cancel{= 0.0006}$$

$$\text{P-value} = \cancel{0.0343 + 0.137 + 0.0195 + 0.006}$$

$$\cancel{= 0.1914} = P_{H_0}(T=5) + P_{H_0}(T=6) + P_{H_0}(T=7) + P_{H_0}(T=8)$$

$$= 0.5$$

$\alpha = 0.05$

p-value = 0.5 > 0.05

We ~~reject~~ accept null hypothesis i.e. ~~not~~ drawn from identical distn.

Example 2 :

X : 26, 27, 31, 26, 19, 21, 20, 25, 30

Y : 23, 28, 26, 24, 22, 19

$n_1 = 9$

$n_2 = 6$

Combined sample

19	19	20	21	22	23	24	25	26	26	26	27	28	30	31
X	Y	X	X	Y	Y	Y	X	Y	X	X	X	Y	X	X

$T_0 = 2$

Combined sample median $= \left[\frac{15+1}{2}\right] = 8$

$\cancel{\delta}$ $H_0 : \delta = 0$

against

$H_1 : \delta \neq 0$

$\cancel{\Rightarrow}$ $P_{H_0}(T \geq 2)$

$P_{H_0}(T \leq 2)$

$$P_{H_0}(T \leq 2) = P_{H_0}(T=0) + P_{H_0}(T=1) + P_{H_0}(T=2)$$

$$= \frac{\binom{8}{6}\binom{7}{0}}{\binom{15}{6}} + \frac{\binom{8}{5}\binom{7}{1}}{\binom{15}{6}} + \frac{\binom{8}{4}\binom{7}{2}}{\binom{15}{6}}$$

$$= 0.37761$$

$$P_{H_0}(T \geq 2) = 1 - P(T < 2)$$

$$= 1 - P(T=0) - P(T=1)$$

$$= 0.91609$$

$$\text{p-value} = 2 \cdot \text{Min}\left[P_{H_0}(T \leq 2), P_{H_0}(T \geq 2)\right]$$

$$= 2 \times 0.037761$$

$$= 0.7552 > 0.05$$

p-value $> \alpha \Rightarrow$ Accept H_0

Wilcoxon's Rank Sum Test

Let $X_1, X_2, \ldots, X_{n_1}$ be i.i.d with distribution function $F(\cdot)$ and $Y_1, Y_2, \ldots, Y_{n_2}$ be i.i.d with d.f. $G(\cdot)$ with $G(x) = F(x-\delta)$, $\delta \in \mathbb{R}$. The samples are drawn independently of each other and $F(\cdot)$ is univariate continuous. Let $\underset{\sim}{Z} = (X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, Y_{n_2})$ be the combined sample. Let R_i be the rank of the i^{th} observation in the combined sample. The wilcoxon's rank sum test statistic is defind as

$$W = \sum_{i=n_1+1}^{n_1+n_2} R_i = \text{Sum of the ranks corresponding to the second sample observations.}$$

- Non-parametric justification :

Under H_0
$\underset{\sim}{R} = (R_1, R_2, \ldots, R_n)'$ has uniform distribution over the set of $n!$ realization of $1, 2, \ldots, n$ so the distribution of $\underset{\sim}{R}$ is independent of F whenever F is continuous. W, being a function of $\underset{\sim}{R}$ has its distribution independent of F under H_0. Hence the test provided by W is exactly distribution free and non-parametric.

- Critical Region : $\delta > 0$ implies that any Y_j is expected to be larger than any X_i i.e. Y_j's are expected to be larger under $H_1 : \delta > 0$ than that under $H_0 : \delta = 0$.

In other words R_j's $(j = n_1+1(1)n_1+n_2)$ are expected to be larger under $H_1 : \delta > 0$ than than under H_0 i.e. W is expected to be larger under $H_1 : \delta > 0$ than under H_0. So a right tail test is appropriate for testing H_0 against $H_1 : \delta > 0$. Similarly a left tail test is appropriate for testing H_0 against $H_1 : \delta < 0$. It can be shown that the distribution of W under H_0 is symmetric about its expectation. So an equal tail test is appropriate for testing H_0 against $H_1 : \delta \neq 0$. Exact size α test for testing H_0 against $H_1 : \delta \neq 0$ is given by

$$\phi = \begin{cases} 1 & \text{if } \left| W - \frac{n_2(n+1)}{2} \right| > W_{\alpha/2} \\ \gamma & \text{if } \left| W - \frac{n_2(n+1)}{2} \right| = W_{\alpha/2} \\ 0 & \text{if } \left| W - \frac{n_2(n+1)}{2} \right| < W_{\alpha/2} \end{cases}$$

γ and $W_{\alpha/2}$ are such that $E_{H_0}[\phi] = \alpha$

- <u>Expectation and Variance of W, under H_0</u>

$$W = \sum_{i=n_1+1}^{n_1+n_2} R_i$$

under H_0, $R_{n_1+1}, R_{n_1+2}, \ldots, R_{n_1+n_2}$ has $\binom{n}{n_2}$ possible realizations and all of them are equally likely.

So, under H_0, $(R_{n_1+1}, R_{n_1+2}, \ldots, R_{n_1+n_2})$ may be considered as a SRSWOR of size n_2 from $(1, 2, \ldots, n)$

$$\therefore E_{H_0}\left[\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2} R_i\right] = \frac{n+1}{2} \quad \left[\because E(\bar{y}) = \bar{Y} \text{ in SRSWOR}\right]$$

$$\text{or, } E_{H_0}\left(W/n_2\right) = \frac{n+1}{2}$$

$$\text{or, } E_{H_0}(W) = \frac{n_2(n+1)}{2}$$

Again,

$$\text{Var}_{H_0}\left[\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2} R_i\right] = \frac{n^2-1}{12 n_2}\cdot\frac{n_1-n_2}{n-1} \quad \left[\text{using } V(\bar{y}) = \frac{\sigma_Y^2}{n}\cdot\frac{N-n}{N-1} \text{ in SRSWOR}\right]$$

[Here, $\sigma_Y^2 = \frac{n^2-1}{12}$ = variance of the first n natural numbers]

$$= \frac{(n+1)\,n_1}{12 n_2} \quad \left[\because n = n_1+n_2\right]$$

$$\therefore \text{Var}_{H_0}\left(\frac{W}{n_2}\right) = \frac{n_1(n+1)}{12 n_2}$$

$$\text{Var}_{H_0}(W) = \frac{n_1(n+1).n_2}{12}$$

$$= \frac{n_1 n_2 (n_1+n_2+1)}{12}$$

Asympotic test

Suppose for each n, there exists $n_1 = n_1(n)$ & $n_2 = n_2(n)$ such that as $n \to \infty$

(i) $n_1 \to \infty,\ n_2 \to \infty$

(ii) $\frac{n_1}{n} \to \lambda,\ \frac{n_2}{n} \to 1-\lambda$

Then under H_0, as $n \to \infty$

$$\frac{W - \frac{n_2(n+1)}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}} \xrightarrow{\mathcal{D}} N(0,1)$$

Alternative	Critical region
$\delta > 0$	$\frac{W - \frac{n_2(n+1)}{2}}{\boxed{\cancel{\sqrt{\frac{n_1 n_2 (n+1)}{12}}}}} > \tau_\alpha \sqrt{\frac{n_1 n_2 (n+1)}{12}}$
$\delta < 0$	$W - \frac{n_2(n+1)}{2} < -\tau_\alpha \sqrt{\frac{n_1 n_2 (n+1)}{12}}$
$\delta \neq 0$	$\left\lvert W - \frac{n_2(n+1)}{2} \right\rvert > \tau_{\alpha/2} \sqrt{\frac{n_1 n_2 (n+1)}{12}}$

Relationship between U & W

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \phi(X_i, Y_j) \qquad \phi(X_i, Y_j) = \begin{cases} 1 & \text{if } X_i < Y_j \\ 0 & \text{o.w} \end{cases}$$

$$\sum_{i=1}^{n_1} \phi(X_i, Y_j) = \text{No. of } X_i\text{'s} > Y_j$$

$= R(Y_j) = $ No. of observations $\leq Y_j$

$= n -$ No. of observations $> Y_j$

Define, q_j = No. of Y observations $> Y_j = (n_2 - j)$

No. of observations $> Y_j$ = No. of X obs $> Y_j$ + No. of Y obs $> Y_j$

$$n - R(Y_j) = \sum_{i=1}^{n_1} \phi(X_i, Y_j) + (n_2 - j)$$

$$R(Y_j) = (n - n_2) + j - \sum_{i=1}^{n_1} \phi(X_i, Y_j)$$

$$= n_1 + j - \sum_{i=1}^{n_1} \phi(X_i, Y_j)$$

$$W = \sum_{j=1}^{n_2} R(Y_j)$$

$$W = n_1 n_2 + \sum_{j=1}^{n_2} j - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \phi(X_i, Y_j)$$

$$W = n_1 n_2 + \frac{n_2(n_2+1)}{2} - U$$

$$E(W) = n_1 n_2 + \frac{n_2(n_2+1)}{2} - E(U)$$

$$= \frac{n_1 n_2}{2} + \frac{n_2(n_2+1)}{2} = \frac{n_2}{2}(n_1 + n_2 + 1)$$

$$= \frac{n_2}{2}(n+1)$$

One sample Kolmogorov - Smirnov Test

Assume that we have a random sample $X_1, X_2, \ldots, X_n$ and we want to test

$$H_0 : F_X(x) = F_0(x) \quad \forall x$$

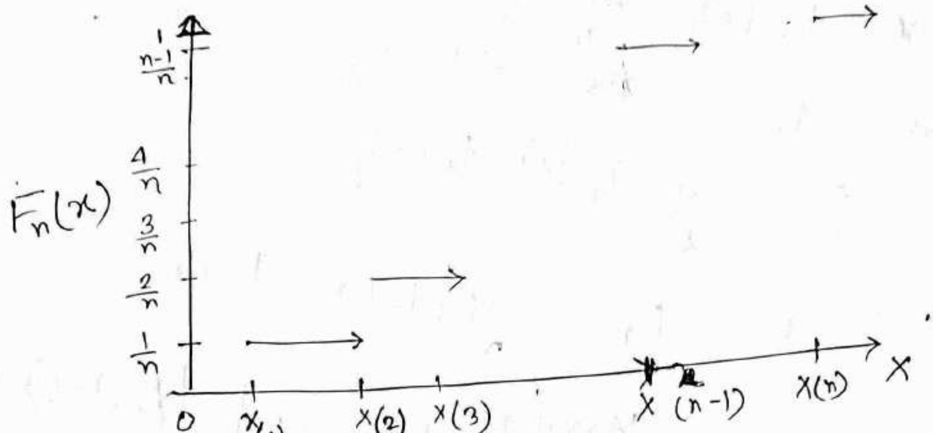
where $F_0(x)$ is a completely specified continuous distribution function, against the usual two sided goodness of fit

$$H_1 : F_X(x) \neq F_0(x) \text{ for some } x$$

Test Statistic

Define $F_n(x) = \dfrac{\text{No. of } X_i\text{ s} \leq x \text{ in the sample}}{n \text{ (sample size)}}$

~~Emp.~~ (it is called the empirical d.f.)



Expectation & Variance of $F_n(x)$

$X_i \leq x \rightarrow$ success

$X_i > x \rightarrow$ Failure

$nF_n(x) =$ No. of X_i s in the sample $\leq x$

$nF_n(x)$ takes values $0, 1, 2, \ldots, n$

$$nF_n(x) \sim Bin\left(n,\ p = P(X_i \le x) = F_X(x)\right)$$

$$E\left(nF_n(x)\right) = nF_X(x)$$

or $$E\left(F_n(x)\right) = F_X(x)$$

$$V\left(F_n(x)\right) = \frac{F_X(x)\left(1 - F_X(x)\right)}{n}$$

By Glivenko - Cantelli's theorem, as n increases $F_n(x)$ with jumps occuring at the values of the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ for the sample, approaches the true distribution function $F_X(x)$ $\forall$ x. Therefore for large n the deviations between the true function and its statistical image, $|F_n(x) - F_X(x)|$ should be small for all values of x. This suggests the use of following statistics,

$$D_n = \sup_x \left|F_n(x) - F_X(x)\right|$$

Under Ho, $$D_n = \sup_x \left|F_n(x) - F_0(x)\right|$$

$$= \max_x \left(\left|F_n(x) - F_0(x)\right|, \left|F_n(x-\epsilon) - F_0(x)\right| \right)$$

where ϵ is a very small positive number/quantity.

D_n is called one sample Kolmogorov - Smirnov test statistic.

The directional derivatives

$$D_n^+ = \sup_x \left[F_n(x) - F_X(x)\right] \text{ and}$$

$$D_n^- = \sup_x \left[F_X(x) - F_n(x)\right]$$

are called One-sided Kolmogorov - Smirnov test statistics.

Theorem : The statistics D_n, D_n^+, and D_n^- are completely distribution free for any continuous F_X.

Proof >> $D_n = \sup_x \left| F_n(x) - F_X(x) \right|$

(For some x, $F_n(x) - F_X(x)$ is positive
for some, $F_n(x) - F_X(x)$ is negative)

$$= \max_x \left[D_n^+ , D_n^-\right]$$

Let us define additional order statistics ;

$$X_{(0)} = -\infty$$

$$X_{(n+1)} = \infty$$

$$F_n(x) = \frac{i}{n} \text{ if } X_{(i)} \le x < X_{(i+1)} \quad , i = 0, 1, 2, \ldots, n$$

Now,

$$D_n^+ = \sup_x \left[F_n(x) - F_X(x)\right]$$

$$= \max_{1 \le i \le n} \sup_{X_{(i)} \le x < X_{i+1}} \left[F_n(x) - F_X(x)\right]$$

$$= \max_{1 \le i \le n} \sup_{X_{(i)} \le x < X_{i+1}} \left(\frac{i}{n} - F_X(x)\right)$$

$X_{(0)} \le x < X_{(1)}$

$F_n(x) = 0$

$X_{(1)} \le x < X_{(2)}$

$F_n(x) = \frac{1}{n}$

$X_{(2)} \le x < X_{(3)}$

$F_n(x) = \frac{2}{n}$

$X_{(n-1)} \le x < X_{(n)}$

$$= \max_{1\le i\le n}\left(\frac{i}{n} - \inf_{X_{(i)} \le x < X_{(i+1)}} F_X(x)\right)$$

$$= \max_{1\le i\le n}\left(\frac{i}{n} - F_X(X_{(i)})\right)$$

$$= \max\left[\max_{1\le i\le n}\left(\frac{i}{n} - F_X(X_{(i)})\right), 0\right]$$

Similarly, $D_n^- = \max\left[\max_{1\le i\le n}\left(F_X(X_{(i)}) - \frac{i-1}{n}\right), 0\right]$

$$D_n = \max\left[\max_{1\le i\le n}\left(F_X(X_{(i)}) - \frac{i-1}{n}\right), \max_{1\le i\le n}\left(\frac{i}{n} - F_X(X_{(i)})\right), 0\right]$$

The probability distribution of D_n^+, D_n^- and D_n depend on the random variable $F_X(X_{(i)})$, $i = 1(1)n$. These are the order statistics from the $U(0,1)$ distⁿ regardless of the original $F_X(x)$ as long as it is continuous. Thus D_n, D_n^+, D_n^- have distributions which are independent of the particular $F(x)$.

Theorem

$$P\left(D_n < \frac{1}{2n} + v\right) = \begin{cases} 0, & \text{if } v \le 0 \\ \displaystyle\int_{\frac{1}{2n}-v}^{\frac{1}{2n}+v}\int_{\frac{1}{3n}-v}^{\frac{1}{3n}+v}\cdots\int_{\frac{2n-1}{2n}-v}^{\frac{2n-1}{2n}+v} f(u_1, u_2, \ldots u_n)\, du_1\, du_2 \ldots du_n & \text{if } 0 < v < \frac{2n-1}{2n} \\ 1 & \text{if } v \ge \frac{2n-1}{2n} \end{cases}$$

provided

$F_X(\cdot)$ is continuous.

$\boxed{\text{Theorem}}$ If $F_X(\cdot)$ is continuous, then for every $d>0$

$$\lim_{n\to\infty} P\left(D_n \le \frac{d}{\sqrt{n}}\right) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1} e^{-2i^2d^2}$$

Problem

The 20 observations below were chosen randomly from $U(0,1)$ distribution, recorded to four significant figures and rearrange in asscending order of magnitude. Test the null hypothesis that the square root of this ~~is~~ numbers also have $U(0,1)$ distn.

0.0123, 0.1039, 0.1954, 0.2621, 0.2802, 0.3217, 0.3645, 0.3919, 0.4240, 0.4814, 0.5139, 0.5846, 0.6275, 0.6541, 0.6889, 0.7621, 0.8320, 0.8871, 0.9249, 0.9634

$\rightarrow$ $$D_n = \sup_x \left| F_n(x) - F_0(x) \right|$$

$$= \max_x \left[\left| F_n(x) - F_0(x) \right|, \left| F_n(x-\epsilon) - F_0(x) \right| \right]$$

$$F_n(x) = \frac{\text{No. of obs} \le x}{\text{Total no. of observation}}$$

x is the square root of numbers

$F_0(x) = x,\ x \in (0,1)$, H_0: samples are from $U(0,1)$

$|F_n(x-\epsilon) - F_0(x)| \longrightarrow$ 1st $|0 - F_0(x)|$

2nd $\to$ $|0.05 - F_0(x)|$

R	x	$F_n(x)$	$F_0(x)$	$\lvert F_n(x) - F_0(x)\rvert$	$\lvert F_n(x-\epsilon) - F_0(x)\rvert$
~~0.0129~~ 0.0001	0.1109	0.05	0.1109	0.0609	0.1109
0.0107	0.3223	0.1	0.3223	0.2223	0.2723
0.0382	0.4420	0.15	0.4420	0.292	0.342
0.0687	0.5119	0.2	0.5119	0.3119	0.3619
0.0785	0.5293	0.25	0.5293	0.2793	0.3293
0.1035	0.5672	0.3	0.5672	0.2672	0.3172
0.1329	0.6037	0.35	0.6037	0.2537	0.3037
0.1536	0.6260	0.4	0.6260	0.226	0.276
0.1798	0.6512	0.45	0.6512	0.2012	0.2512
0.2317	0.6938	0.5	0.6938	0.1938	0.2438
0.2641	0.7169	0.55	0.7169	0.1669	0.2169
0.3418	0.7646	0.6	0.7646	0.1646	0.2146
0.3938	0.7921	0.65	0.7921	0.1421	0.1921
0.4278	0.8088	0.7	0.8088	0.1088	0.1588
0.4746	0.8300	0.75	0.8300	0.08	0.13
0.5808	0.8729	0.8	0.8729	0.0729	0.1229
0.6922	0.9121	0.85	0.9121	0.0621	0.1121
.7869	0.9419	0.9	0.9419	0.0419	0.0919
.8554	0.9617	0.95	0.9617	0.0117	0.0617
	0.9815	1	0.9815	0.0185	~~0.0618~~ 0.0315

$D_n = 0.3619$

for $\alpha = 0.05$, $n = 20$

If observed $D_n > 0.294$ (for $\alpha = 0.05$),

then reject H_0

Conclusion: Therefore we reject H_0 i.e. samples are not coming from $U(0,1)$.

$X \sim U(0,1)$

$P(\sqrt{X} \le x) = P(X \le x^2) = x^2$

c.d.f of the rv $= \cancel{P(X \le x^2)}$

$Y = \sqrt{X}$

$F_Y(y) = y^2$

p.d.f of y, $f_Y(y) = \frac{d}{dy}(y^2) = 2y$, $0 < y < 1$

$\downarrow$

Beta $(2,1)$

Prob 2: 1.5, 2.3, 4.2, 7.1, 10.4, 8.4, 9.3, 6.5, 2.5, 4.6. Test whether the data comes from exponential distribution.

$\rightarrow$ For $Exp(\theta)$ dist^n

$\hat{\theta} = \frac{1}{\bar{x}}$ $\qquad F_0(x) = 1 - e^{-\hat{\theta} x}$

Since the parameter θ of the exponential dist^n is not specified in the problem we take the MLE of θ, $\hat{\theta} = \frac{1}{\bar{x}}$ as an estimate of θ.

x	$F_n(x)$	$F_0(x)$	$\lvert F_n(x) - F_0(x) \rvert$	$\lvert F_n(x-\epsilon) - F_0(x) \rvert$
1.5	0.1	0.232	0.132	0.132
2.3	0.2	0.333	0.133	0.233
2.5	0.3	0.356	0.056	0.156
4.2	0.4	0.523	0.123	0.223
4.6	0.5	0.555	0.055	0.155
6.5	0.6	0.681	0.081	0.181
7.1	0.7	0.713	0.013	0.113
8.4	0.8	0.772	0.028	0.072
9.3	0.9	0.805	0.095	0.0005
10.4	1	0.82	0.16	0.06

and use the formula $F_0(x) = 1 - e^{-\hat{\theta} x}$ to compute Kolmogorov smirnov test statistic.

$$\hat{\theta} = \frac{1}{5.68} = 0.176$$

$$D_n = 0.233$$

at $\alpha = 0.05$, $n = 10$, $D_n = 0.409$

$$D_n < 0.409$$

Conclusion: accept H_0 i.e. data comes from exponential distribution.

Two Sample Kolmogorov Smirnov Test

Let $X_1, X_2, \ldots, X_{n_1}$ be a random sample of size n_1 from a population with distribution function $F(\cdot)$ and $Y_1, Y_2, \ldots, Y_{n_2}$ be a random sample of size n_2 from a population with distⁿ funcⁿ $G(\cdot)$. We want to test $H_0 : F(x) = G(x) \;\forall x$

against

$H_1 : F(x) \neq G(x)$ for some x

The order statistics corresponding to these two continuous random samples of size n_1 & n_2 from the popuⁿ F and G are.

$X_{(1)}, X_{(2)}, \ldots, X_{(n_1)}, Y_{(1)}, Y_{(2)}, \ldots, Y_{(n_2)}$.

~~They~~ Their respective empirical distribution functions are defined and denoted by $F_{n_1}(x)$ and $G_{n_2}(x)$

by,
$$F_{n_1}(x) = \begin{cases} 0 & \text{if } x < X_{(1)} \\ \frac{k}{n_1} & \text{if } X_{(k)} \leq x < X_{(k+1)}, \quad k = 1, 2, \ldots, n_1 - 1 \\ 1 & \text{if } x \geq X_{(n_1)} \end{cases}$$

&
$$G_{n_2}(x) = \begin{cases} 0 & \text{if } x < Y_{(1)} \\ \frac{k}{n_2} & \text{if } Y_{(k)} \leq x < Y_{(k+1)}, \quad k = 1(1)n_2 - 1 \\ 1 & \text{if } x \geq Y_{(n_2)} \end{cases}$$

In a Combined ordered arrangement of the $n_1 + n_2$ sample observations $F_{n_1}(x)$ and $G_{n_2}(x)$ are the respective proportions of X and Y observations which do not ~~exist~~ exceed the specified value x. If the null hypothesis is true the population distributions are identical and we have two samples from the same population. The empirical distribution functions for the X and Y samples are reasonable estimates of their respective population cdf. Therefore allowing for sampling variation there should be reasonable agreement between the two empirical distributions if H_0 is true, otherwise the data suggest that H_0 is not true and should be reject.

The two sided Kolmogorov Smirnov test statistics denoted by D_{n_1,n_2} is defined as

$$D_{n_1,n_2} = \max_{x} \left| F_{n_1}(x) - G_{n_2}(x) \right|$$

Since here only the magnitudes of the deviations are considered D_{n_1,n_2} is appropriate for a general two sided alternative. The critical region is given by, $D_{n_1,n_2} \geq C_\alpha$ where C_α is such that $P_{H_0}(D_{n_1,n_2} \geq C_\alpha) \leq \alpha$

When $n_1, n_2 \to \infty$ in such a way that $\frac{n_1}{n_2}$ remains constant, $\lim\limits_{n_1, n_2 \to \infty} P\left(\sqrt{\frac{n_1 n_2}{n_1 n_2}}\, D_{n_1,n_2} \leq d \right)$

$$= 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 d^2}$$

<u>Problem</u> : Two mutually independent random sample each of size 8. where generated, first one from the standard Normal dist^n and the second one from χ^2 - dist^n with d.f 18

$N(0,1)$	χ^2_{18}	Y
-1.91	4.90	-2.18
-1.22	7.25	-1.79
-0.96	8.04	-1.66
-0.72	14.10	-0.65
0.14	18.3	0.05
0.82	21.21	0.535
1.45	23.1	0.85
1.86	28.12	1.69

Based on these samples investigate whether the standarised χ^2-distribution/variable $\frac{\chi^2 - n}{\sqrt{2n}}$ approches a $N(0,1)$ distribution even for moderate degrees of freedom.

$X \to$ Normal samples

- For large n,

$$\frac{\chi^2 - n}{\sqrt{2n}} \xrightarrow{\mathcal{D}} N(0,1) \qquad Y \to \frac{\chi^2 \text{ samples} - 18}{\sqrt{36}}$$

Table

Combined ordered observations (t) (1)	# X ≤ t (2)	# Y ≤ t (3)	$F_{n_1}(t)$ $\frac{(2)}{8}$	$G_{n_2}(t)$ $\frac{(3)}{8}$	$\lvert F_{n_1}(t) - G_{n_2}(t) \rvert$
−2.18 (Y)	0	1	0	1/8	+ 0.125
−1.91 (X)	1	1	1/8	1/8	0
−1.79 (Y)	1	2	1/8	1/4	0.125
−1.66 (Y)	1	3	1/8	3/8	0.25
−1.22 (X)	2	3	1/4	3/8	0.125
−0.96 (X)	3	3	3/8	3/8	0
−0.72 (X)	4	3	1/2	3/8	1/8.
−0.65 (Y)	4	4	1/2	1/2	0
0.05 (Y)	4	5	1/2	5/8	1/8
0.14 (X)	5	5	5/8	5/8	0
0.54 (Y)	5	6	5/8	3/4	1/8
0.82 (X)	6	6	3/4	3/4	0
0.85 (Y)	6	7	3/4	7/8	1/8
1.45 (X)	7	7	7/8	7/8	0
1.69 (Y)	7	8	7/8	1	1/8
1.86 (X)	8	8	1	1	0

$D_{n_1, n_2} = 0.25$ (max of $\lvert F_{n_1}(t) - G_{n_2}(t) \rvert$)

$\cancel{n_1 n_2 D} = n_1 n_2 D = 16.25$

n_1	n_2	$n_1 n_2 D$	P
		64	0.000
8	8	56	0.002
		48	0.019
		40	0.087
		32	0.283

③ $P_{H_0}(n_1 n_2 D_{n_1, n_2} \geq 32)$

$= 0.283$

② $P_{H_0}(n_1 n_2 D_{n_1, n_2} \geq 16)$

$\geq P_{H_0}(n_1 n_2 D_{n_1, n_2} \geq 32)$

$P_{H_0}(64 D_{n_1, n_2} \geq 16) \rightarrow$ P value

$= P_{H_0}(D_{n_1, n_2} \geq \frac{1}{4})$

---

① p-value $= P_{H_0}(D_{n_1, n_2} \geq \frac{1}{4}) = P_{H_0}(n_1 n_2 D_{n_1, n_2} \geq 16)$

④ $\therefore$ P-value $\geq 0.283 > \alpha$.

accept the null hypothesis.

~~Those samples~~

$\boxed{\text{Kruskal - Wallis Test}}$ (Nonparametric competitors of ONE-WAY ANOVA)

The Kruskal-Wallis test is the natural extension of the Wilcoxon Rank sum test for ~~l~~location with two independent samples to the situation of K mutually independent samples from continuous populations. The null hypothesis is that the K populations are the same but when we assume the location model ~~is~~ ~~to~~ this hypotheses can be written in terms of the respective location parameters as follows

$H_0 : \theta_1 = \theta_2 = \dots = \theta_K$

against

H_1 : at least two of them differ.

Test procedure: To perform the test all $n_1 + n_2 + \dots + n_K = N$ raw observations are combined into a simple array and rank from $1, 2, \dots, N$. The test statistic is defined as,

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{1}{n_i} \left[R_i - \frac{n_i(N+1)}{2} \right]^2$$

$$= \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - 3(N+1)$$

where R_i is the sum of the ranks of the observations corresponding to the K^{th} group or K^{th} sample. The statistic is asymtotically distributed as χ^2 with d.f. $(K-1)$. The approximation is generally satisfactory except when $K=3$ and the sample sizes are less than equal to 5. The critical region is the large values of H. If H_0 is rejected then we go for the pairwise comparison test and we conclude that the location parameters of i^{th} and j^{th} population differ significantly if .

$$\frac{|\bar{R}_i - \bar{R}_j|}{\left[\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)\right]^{1/2}} \geq \tau_{\alpha/2}$$

where $\bar{R}_i = \frac{R_i}{n_i}$ = average rank of the observations corresponding to group i.

If $n_i = \frac{N}{K} \ \forall \ i = 1(1)K$

Then $\frac{|\bar{R}_i - \bar{R}_j|}{\left[\frac{K(N+1)}{6}\right]^{1/2}} \geq \tau_{\alpha/2}$ is the rejection criterion.

Problem: In a comparison of the clinic action of four detergents, 20 pieces of white cloths where first soiled with ink. The cloths were then washed under controlled condition with 5 pieces washed by each of the detergents. whiteness readings are showing

Detergent

A	B	C	D
77	74	73	76
81	66	78	85
61	58	57	77
76	63	69	64
69	61	63	80

Test the hypothesis of no difference between the four ranks of detergent regarding the average whiteness reading after washing.

Combined observations in ascending order

	Rank
57 (C)	1
58 (B)	2
61 (B)	3.5
61 (A)	3.5
63 (B)	5.5
63 (C)	5.5
64 (D)	7
66 (B)	8
69 (A)	9.5
69 (C)	9.5
73 (C)	11
74 (B)	12
76 (A)	13.5
76 (D)	13.5
77 (A)	15.5
77 (D)	15.5
78 (C)	17
80 (D)	18
81 (A)	19
85 (D)	20

$R_1 = 3.5 + 9.5 + 9.5 + 13.5 + 15.5 = 61$

$R_2 = 2 + 3.5 + 5.5 + 8 + 12 = 31$

$R_3 = 1 + 5.5 + 11 + 9.5 + 17 = 44$

$R_4 = \frac{20(20+1)}{2} - 61 - 31 - 44 = 74$

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - 3(N+1) \quad , K = 4, \; N = 20, \; n_i = 5 \; \forall i$$

$= 6.109$

$\chi^2_{3, 0.05} = 7.815$

Reject H_0 if $H > \chi^2_{3, 0.05}$

Here, $7.815 \not> 6.109$

accept H_0

i.e. no difference b/w the four ranks of detergent regarding the avg. whiteness reading after washing.

$\boxed{\text{NOTE}}$ For $K = 3$

and $n_1, n_2, n_3 \leq 5$

Use a special table of Kruskal-Wallis test.

For $K > 3$, use χ^2 table.

(2) The following table shows the life times in hours in excess of thousand hours, of the samples of 60 watt electric light bulbs of three different brands. Test at 5% level the hyp. that there is no diff b/w the three brands w.r.t any life tm.

Brand I	II	III
16	18	26
15	22	31
13	20	24
21	16	30
15	24	24

Combined samples

13(I) (1), 15(I) (2.5), 15(I) (2.5), 16 (I) (4.5), 16 (II) (4.5), 18 (II) 6, 20 (II) (7), 21 (I) (8), 22 (II) (9), 24 (II) (11), 24 (III) (11), 24 (III) (11), 26 (III) (13), 30 (III) (14), 31 (III) (15)

$$R_1 = 18.5 , \quad R_2 = 37.5 , \quad R_3 = 64$$

$$H = \frac{12}{15 \times 16} (1168.9) - 3 \times 16 = \cancel{7.925} \; 10.445$$

$H > 5.78$ (for, $n_1, n_2, n_3 = 5$. at 0.05 $H = 5.78$)

Reject Ho

$$\bar{R}_1 = \frac{R_1}{5} = 3.7 , \quad \bar{R}_2 = 7.5 , \quad \bar{R}_3 = 12.4$$

$$\frac{|\bar{R}_i - \bar{R}_j|}{\left[\frac{K(N+1)}{6}\right]^{1/2}} , \quad \frac{|\bar{R}_1 - \bar{R}_2|}{\left(\frac{16 \times 3}{6 \times}\right)^{1/2}} = \frac{3.8}{0.943} = 4.029$$

$$\frac{|\bar{R}_1 - \bar{R}_3|}{\left(\frac{16 \times 3}{6 \times}\right)^{1/2}} = 9.226$$

$$\frac{|\bar{R}_2 - \bar{R}_3|}{\left(\frac{16 \times 3}{6 \times}\right)^{1/2}} = 5.196$$

$|\bar{R}_1 - \bar{R}_2| = 3.8$

$|\bar{R}_2 - \bar{R}_3| = 4.9$

$|\bar{R}_1 - \bar{R}_3| = 8.7$

$$\frac{|\bar{R}_i - \bar{R}_j|}{\left[\frac{K(N+1)}{6}\right]^{1/2}} \geq z_{0.025} = 1.96 \quad \text{is rejection}$$

rejection

$\therefore$ accept for (1,2), (2,3)

Therefore brand I & III differs significantly.

$\Rightarrow |\bar{R}_i - \bar{R}_j| > \cancel{1.8143}$ 5.543 $\Rightarrow$ Reject or else accept.

1 → 3
2 → 4
3 → 4
4 → ~~0~~ 3
5 → 4
6 → 4
7 → 0
8 → 4
9 → 4
10 → 1
11 → 0
12 → 0
13 → 3
14 → 2
15 → 3
16 → 4
17 → 1
18 → 0
19 → 3
20 → ~~0~~ 1

21 → 2
22 → 4
23 → 2
24 → 0
25 → 1
26 → 3
27 → 2
28 → 4
29 → 0
30 → ~~1~~ 3
31 → 4
32 → 3
33 → 4
34 → 1
35 → 2
36 → 4

4 → very frequently
3 → frequently
2 → sometimes
1 → rarely
0 → almost never

87

(X) Popular washer	Rank	New washer (Y)	Rank
13	14	10	5
10	5	11	8
9	1.5	12	11
12	11	13	14
11	8	9	1.5
10	5	11	8
8	1	14	16
		12	11
		13.	14
	$\boxed{40.5}$		$\boxed{88.5}$

	Rank
8	1
9	1.5
9	1.5
10	5
10	5
10	5
11	8
11	8
11	8
12	11
12	11
12	11
13	14
13	14
13	14
14	16

$$U_{PW} = 40.5 - \frac{7(7+1)}{2}$$

$$= 12.5$$

$$U_{NW} = 88.5 - \frac{9(9+1)}{2}$$

$$= 43.5$$

$n_1 = 7.$

$n_2 = 9$

$n_1 + n_2 = 16$

9	9
X	Y
Y	X

10	10	10
X	X	Y
X	Y	X
Y	X	X

11	11	11
X	Y	Y
Y	X	Y
Y	Y	X

12	12	12
X	Y	Y
Y	X	Y
Y	Y	X

13	13	13
X	Y	Y
Y	X	Y
Y	Y	X

Seq	8	9	9	10	10	10	11	11	11	12	12	12	13	13	13	14
	X	Y	X	X	Y	X	Y	X	Y	X	Y	Y	Y	X	Y	Y

$0+1+1+2+4+6 = 14$ → ~~Run~~

$U = 14$

$W = 14 + \frac{n_2(n_2+1)}{2}$

$= 59$

$U' = n_1 n_2 - U = 63 - 14$

$= 49$

From table VIII of Appendix C, we find that for $n_2 = 9$, $n_1 = 7$, for a two tail test at the level 0.02, the critical value is 9. Since 49 is greater than 9, we have no reason to beleive that the samples are ~~not~~ drawn from identical ~~test~~ dist$\underline{\underline{n}}$.

26 26 26 26

Y Y X X

8	9	9	10	10	10	11	11	11	12	12	12	13	13	13	14
X	X	Y	X	X	Y	X	Y	Y	X	Y	Y	X	Y	Y	Y

$1 + 1 + 2 + 4 + 6 = 14$

$\underline{U = 14}$

$U' = 7 \times 9 - 14$

$\boxed{U' = 49}$

$H_0 : \delta = 0$

$H_1 : \delta > 0$

$W = U + \frac{n_2(n_2+1)}{2}$

For $n_1 = 7$, $n_2 = 9$

$C_\alpha = 9$

$= 14 + \frac{9 \times 10}{2} = 59$

$U < 9$ → reject null hypothesis

X Y

X − 17

Y

Run test

X X Y Y X X X Y Y Y Y X X Y Y Y → Run = 6

Thus we have $n_1 = 7$, $n_2 = 9$, $N = 16$ and 3 runs of x and 3 runs of y's giving $r = 6$. The critical value of r at the 5 % level from the table is 4. Since the observed value is greater than the critical value, we accept the null hypotheses (that the popu^n s are dist^n are identical) at 5 %. level.