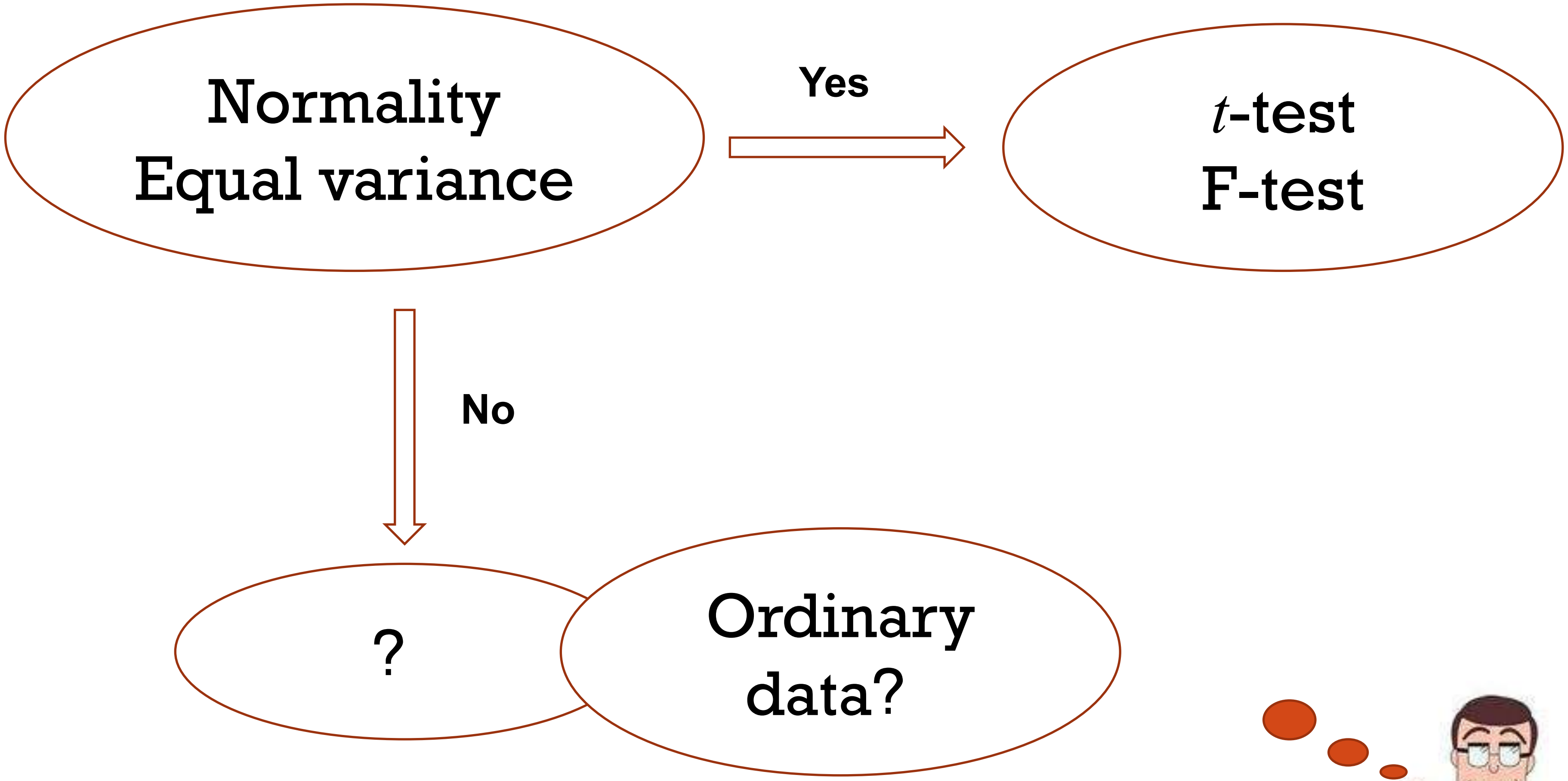
NONPARAMETRIC TESTS


Normality
Equal variance
Yes
t-test
F-test
No
?
Ordinary
data?

CONTENTS

1. Introduction: distribution-free tests

PARAMETRIC VS. NONPARAMETRIC TESTS

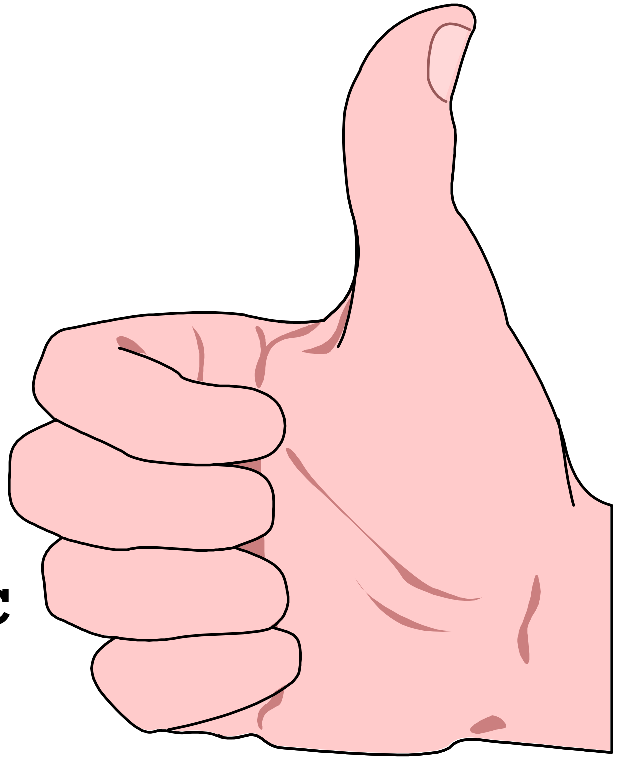
- **Parametric statistics** is a branch of statistics which assumes that sample data comes from a population that follows a probability distribution (i.e., normal distribution) based on a fixed set of parameters. Most well-known elementary statistical methods are parametric (z-test, t-test, F-test).
- **Nonparametric tests**, also called distribution-free hypothesis tests, generally have fewer required conditions. In particular, nonparametric tests do not require the population to follow a particular distribution, such as the normal distribution . Nonparametric tests replace the actual data values with either **signs** (positive or negative) or **ranks**. Do not deal with specific population parameters, such as the mean or standard deviation.

Nonparametric Tests -Conditions

- Unknown distribution; Skewed distribution; Uncertain data
- Variances are not equal
- Ordinal data -example: good-better-best
- Nominal data -example: male-female
- Outlier

Advantages of nonparametric tests

- Used with all scales
- Easier to compute
  - Developed originally before wide computer use
- Make fewer assumptions
- Need not involve population parameters
- Results **may be** as exact as parametric procedures



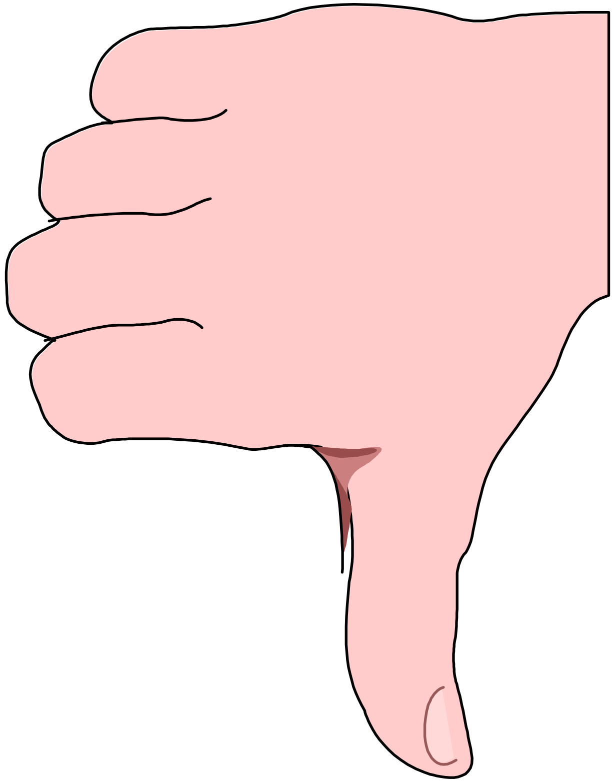
Disadvantages of Nonparametric Tests

- May waste information
  - If data permit using parametric procedures

Table 1 Efficiency of nonparametric tests compared with parametric tests

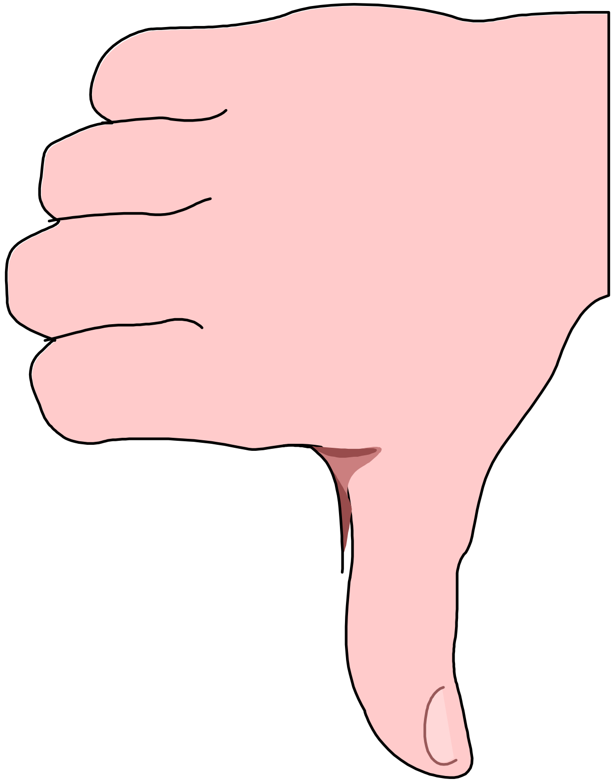
Section	Situation	Parametric test	Nonparametric test	Efficiency
14.2	Matched pairs (dependent samples)	*t* test or *Z* test	Sign test	0.63
14.3	Matched pairs (dependent samples)	*t* test or *Z* test	Wilcoxon signed rank test	0.95
14.4	Two independent samples	*t* test or *Z* test	Wilcoxon rank sum test	0.95
14.5	Several independent samples	Analysis of variance (*F* test)	Kruskal-Wallis test	0.95
14.6	Correlation	Linear correlation	Rank correlation test	0.91
14.7	Randomness	No parametric test	Runs test	—

- I
  l
- T



Disadvantages of Nonparametric Tests

- May waste information
  - If data permit using parametric procedures
  - Example: converting data from ratio to ordinal scale
- Difficult to compute by hand for large samples
- Tables not widely available

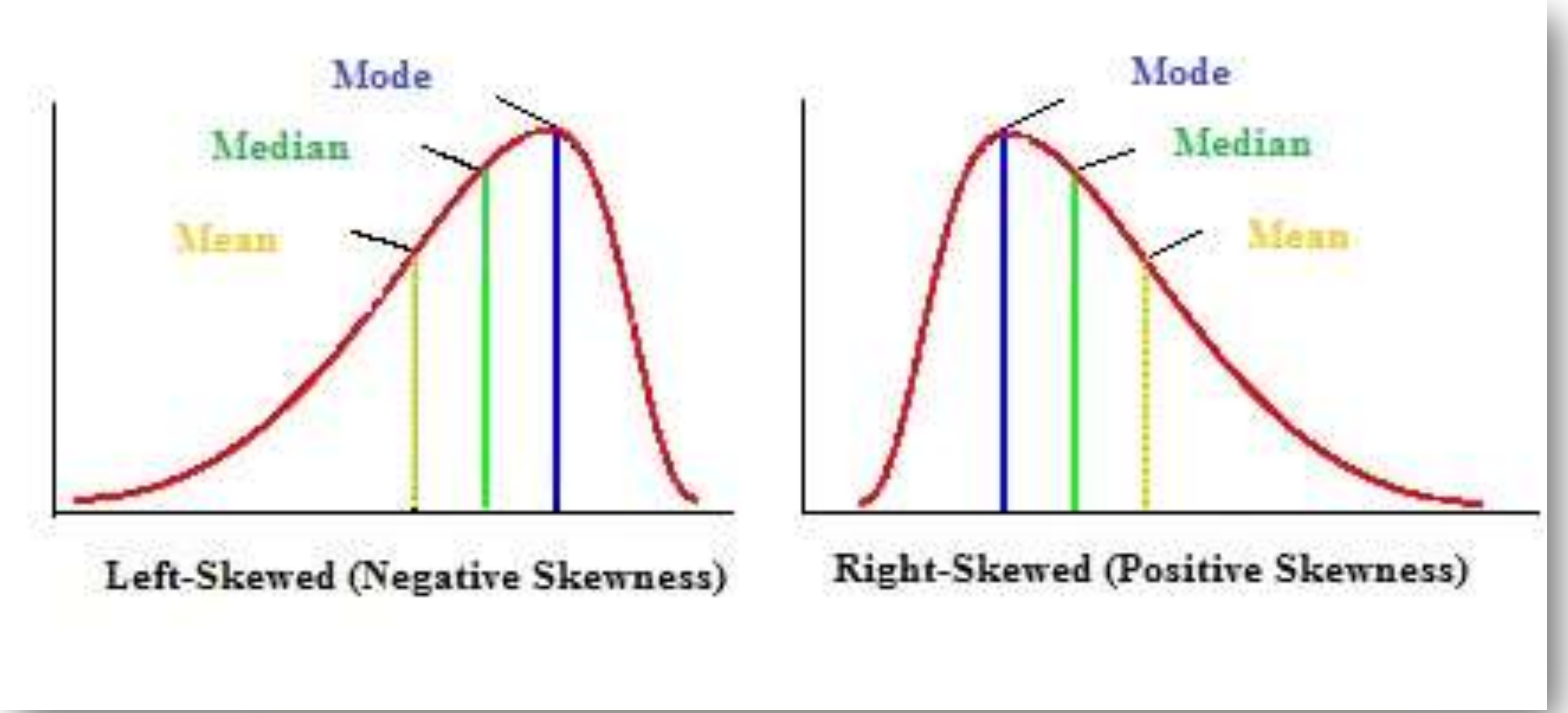


E.g.
1,2,3,4,5,7,13,22,38,45 - actual
1,2,3,4,5,6, 7, 8, 9,10 - rank

Frequently Used Nonparametric Tests

- Sign Test
- Wilcoxon Rank Sum Test
- Wilcoxon Signed Rank Test
- Kruskal Wallis *H*-Test
- Spearman's Rank Correlation Coefficient

Skewed data: Use median as a measure instead of mean



Median (n=Odd)	Median (n=Even)
$Median = x_{\frac{(n+1)}{2}}$	$Median = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$

Order statistic

Suppose a random sample of five rats yields the following weights (in grams):

$$x_1 = 602 \qquad x_2 = 781 \qquad x_3 = 709 \qquad x_4 = 742 \qquad x_5 = 633$$

What are the observed order statistics of this set of data?

The observed order statistics are:

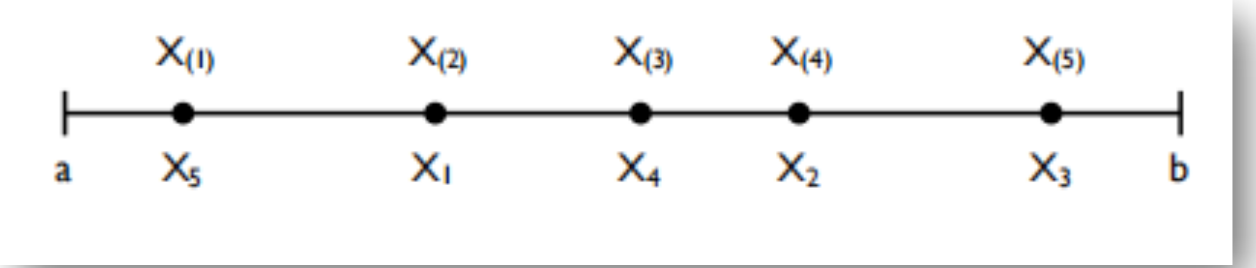
$$y_1 = 602 < y_2 = 633 < y_3 = 709 < y_4 = 742 < y_5 = 781$$

When **ranking** the **data**, **ties**, that is, two or more subjects having exactly the same value of a variable.

Order statistic

Let X_1, X_2, X_3, X_4, X_5 be random variables with a distribution F with a range of (a, b). We can relabel these X's such that their labels correspond to arranging them in increasing order so that

$$X_{(1)} \le X_{(2)} \le X_{(3)} \le X_{(4)} \le X_{(5)}$$



In the case where the distribution F is continuous we can make the stronger statement that

$$X_{(1)} < X_{(2)} < X_{(3)} < X_{(4)} < X_{(5)}$$

Order statistic

Let X_1, X_2, X_3, …… X_n iid be random variables, $X_{(K)}$ is the kth smallest X, usually called the kth order statistic

$X_{(1)}$ is therefore the smallest X and

$$X_{(1)} = \min(X1, \ldots.., Xn)$$

Similarly, $X_{(n)}$ is therefore the largest X and

$$X_{(n)} = \max(X1, \ldots.., Xn)$$

Iid: independent and identically distributed

Order statistic

For $X1, X2, \ldots, Xn$ iid continuous random variables with pdf f and cdf F the density of the maximum is

$$
\begin{aligned}
P(X_{(n)} \in [x, x+\epsilon]) &= P(\text{one of the } X\text{'s} \in [x, x+\epsilon] \text{ and all others} < x) \\
&= \sum_{i=1}^{n} P(X_i \in [x, x+\epsilon] \text{ and all others} < x) \\
&= nP(X_1 \in [x, x+\epsilon] \text{ and all others} < x) \\
&= nP(X_1 \in [x, x+\epsilon])P(\text{all others} < x) \\
&= nP(X_1 \in [x, x+\epsilon])P(X_2 < x) \cdots P(X_n < x) \\
&= nf(x)\epsilon F(x)^{n-1}
\end{aligned}
$$

$$\boxed{f_{(n)}(x) = nf(x)F(x)^{n-1}}$$

Order statistic

For $X1, X2, \ldots, Xn$ iid continuous random variables with pdf f and cdf F the density of the minimum is

$$\begin{aligned}
P(X_{(1)} \in [x, x+\epsilon]) &= P(\text{one of the } X\text{'s} \in [x, x+\epsilon] \text{ and all others} > x) \\
&= \sum_{i=1}^{n} P(X_i \in [x, x+\epsilon] \text{ and all others} > x) \\
&= nP(X_1 \in [x, x+\epsilon] \text{ and all others} > x) \\
&= nP(X_1 \in [x, x+\epsilon])P(\text{all others} > x) \\
&= nP(X_1 \in [x, x+\epsilon])P(X_2 > x) \cdots P(X_n > x) \\
&= nf(x)\epsilon(1 - F(x))^{n-1}
\end{aligned}$$

$$\boxed{f_{(1)}(x) = nf(x)(1 - F(x))^{n-1}}$$

Order statistic

For $X1, X2, \ldots, Xn$ *iid* continuous random variables with pdf f and cdf F the density of the kth order statistic is

$$
\begin{aligned}
P(X_{(k)} \in [x, x+\epsilon]) &= P(\text{one of the } X\text{'s} \in [x, x+\epsilon] \text{ and exactly } k-1 \text{ of the others} < x) \\
&= \sum_{i=1}^{n} P(X_i \in [x, x+\epsilon] \text{ and exactly } k-1 \text{ of the others} < x) \\
&= nP(X_1 \in [x, x+\epsilon] \text{ and exactly } k-1 \text{ of the others} < x) \\
&= nP(X_1 \in [x, x+\epsilon])P(\text{exactly } k-1 \text{ of the others} < x) \\
&= nP(X_1 \in [x, x+\epsilon]) \left(\binom{n-1}{k-1} P(X < x)^{k-1} P(X > x)^{n-k} \right)
\end{aligned}
$$

$$
\boxed{f_{(k)}(x) = nf(x)\binom{n-1}{k-1} F(x)^{k-1}(1-F(x))^{n-k}}
$$

Order statistic: Cumulative Distribution of the min and max

For $X1, X2, \ldots, Xn$ *iid* continuous random variables with pdf f and cdf F the Cumulative Distribution of the min and max is

$$
\begin{aligned}
F_{(1)}(x) &= P(X_{(1)} < x) = 1 - P(X_{(1)} > x) \\
&= 1 - P(X_1 > x, \ldots, X_n > x) = 1 - P(X_1 > x) \cdots P(X_n > x) \\
&= 1 - (1 - F(x))^n
\end{aligned}
$$

$$
\begin{aligned}
F_{(n)}(x) &= P(X_{(n)} < x) = 1 - P(X_{(n)} > x) \\
&= P(X_1 < x, \ldots, X_n < x) = P(X_1 < x) \cdots P(X_n < x) \\
&= F(x)^n
\end{aligned}
$$

2. Wilcoxon Signed-Rank Test: Paired comparisons

Wilcoxon Signed Rank Test

Designed by Frank Wilcoxon (1892-1965)

FRANK WILCOXON: THE MAN BEHIND THE TEST

- Born in Ireland, grew up in Catskills in New York
- Earned B.S. at Penn. Military Academy, master's at Rutgers, Ph.D. at Cornell, all in chemistry
- Worked as a research scientist at several laboratories
- Became interested in statistical methods after reading R.A. Fisher's *Statistical Methods for Research Workers*
- In response to Fisher's Student's T-tests, he developed non-parametric tests for paired and unpaired data sets

Source: http://www.wikipedia.org

Wilcoxon Signed Rank Test

- Tests probability distributions of two related populations
- Corresponds to t-test for dependent (paired) means
- Assumptions
  - Random samples
  - Both populations are continuous
- Can use normal approximation if $n > 25$

- Example 1
- Suppose we want to compare the effectiveness of two ointments (A, B) in reducing excessive redness in people who cannot otherwise be exposed to sunlight. Ointment A is randomly applied to either the left or right arm, and ointment B is applied to the corresponding area on the other arm. The person is then exposed to 1 hour of sunlight, and the two arms are compared for degrees of redness. Suppose instead the degree of burn can be quantified on a 10-point scale, with 10 being the worst burn and 1 being no burn at all.

(1) Arm A is not as red as arm B.
(2) Arm B is not as red as arm A.
(3) Both arms are equally red.

$d_i = x_i - y_i$

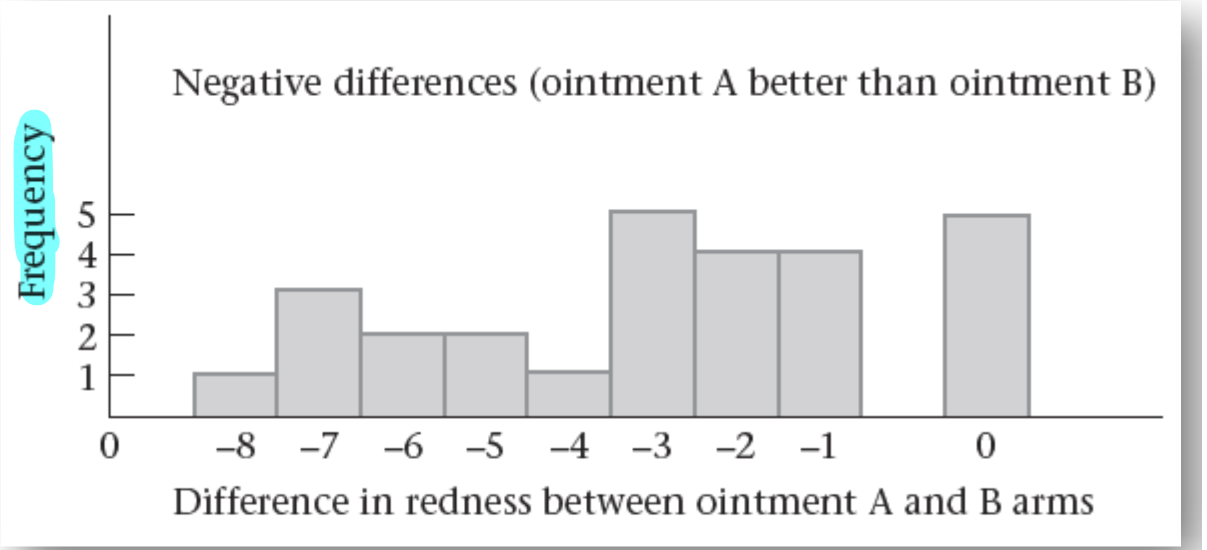
X_i=digress of burn for ointment A

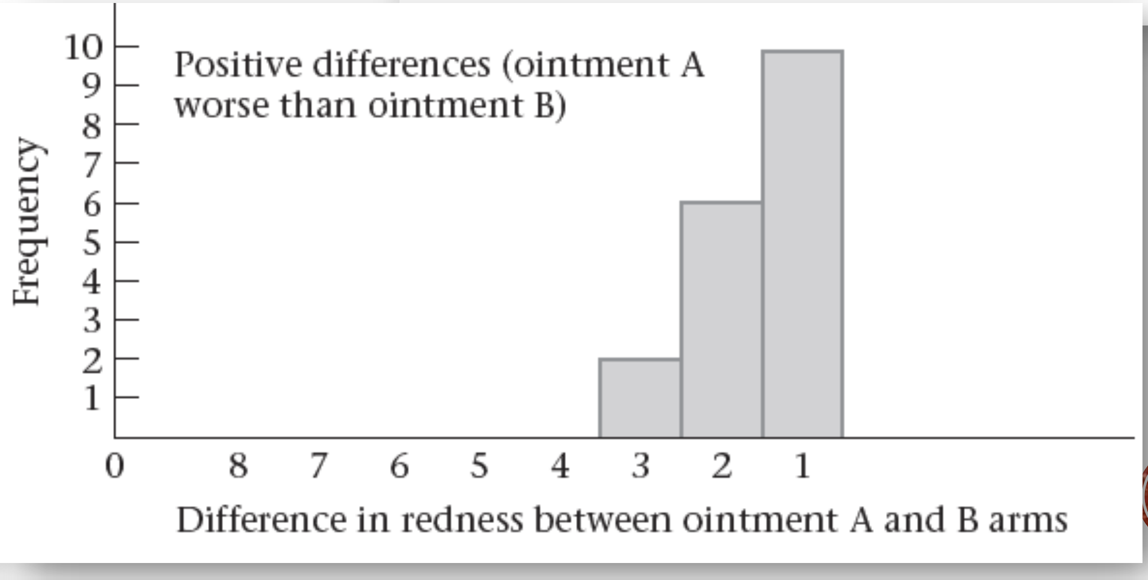
Y_i=digress of burn for ointment B

$H0: \Delta = 0$; $H1: \Delta \neq 0$

Where Δ =the median score difference between the ointment A and ointment B arms

	Negative		Positive	
$\lvert d_i \rvert$	d_i	f_i	d_i	f_i
10	−10	0	10	0
9	−9	0	9	0
8	−8	1	8	0
7	−7	3	7	0
6	−6	2	6	0
5	−5	2	5	0
4	−4	1	4	0
3	−3	5	3	2
2	−2	4	2	6
1	−1	4	1	10
		22		18





Wilcoxon Signed-Rank Test procedure

① Arrange the difference $d_i = X_{1i} - X_{2i}$ in order of $|d_i|$

② Discard differences with 0 value, and sample size reduce the number of zeros

③ Assign ranks, R_i, from lowest to highest; for the same $|d_i|$ with different signs, give them the averaged rank

④ Assign ranks the same signs as d_i

⑤ Sum '+' ranks (T_+) and '–' ranks (T_-)

- Test statistic is T_- or T_+ (one-tail test)
- Test statistic is T = smaller of T_- or T_+ (two-tail test)

$\lvert d_i \rvert$	Negative d_i	Negative f_i	Positive d_i	Positive f_i	Number of people with same absolute value	Range of ranks	Average rank
10	−10	0	10	0		—	—
9	−9	0	9	0	0	—	—
8	−8	1	8	0	1	40	40.0
7	−7	3	7	0	3	37–39	38.0
6	−6	2	6	0	2	35–36	35.5
5	−5	2	5	0	2	33–34	33.5
4	−4	1	4	0	1	32	32.0
3	−3	5	3	2	7	25–31	28.0
2	−2	4	2	6	10	15–24	19.5
1	−1	4	1	10	14	1–14	7.5
		22		18			
0	0	5					

?

Ties: no groups of differences with the same absolute value

$$T_+ = 10\times7.5 + 6\times19.5 + 2\times28.0 = 248$$

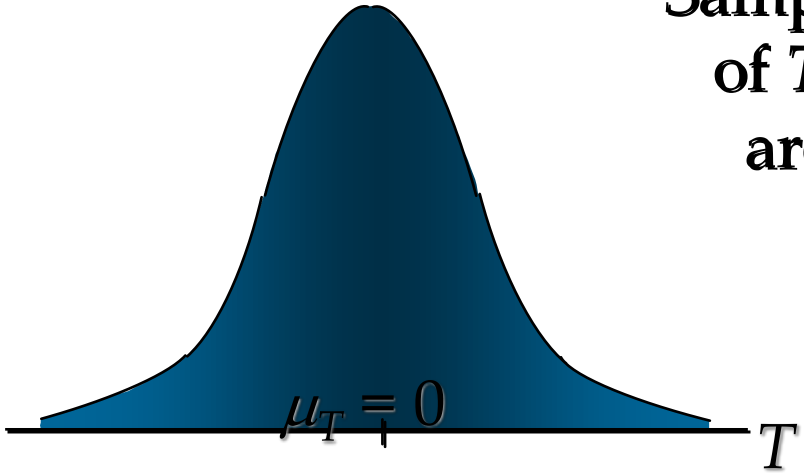
$$T_- = 4\times7.5 + 4\times19.5 + 5\times28.0 + 1\times32 + \cdots + 1\times40 = 572$$

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

RATIONAL

- If there is no difference between the two experimental conditions

1. there will be similar numbers of positive and negative difference scores
2. about half of the signs would be plus signs and about half would be minus signs



Sampling distribution of T if populations are identical

Under the null hypothesis, then the sum of the ranks, or rank sum (T), has the following properties

$$E(T)=\frac{n(n+1)}{4}, Var(T)=\frac{n(n+1)(2n+1)}{24}$$

where n is the number of nonzero differences

Proof can be found in self study - Variance of Wilcoxon Signed-Rank Statistic - Cross Validated (stackexchange.com)

If $n \geq 50$, and there are no ties, then a normal approximation can be used

$$z=\frac{T-E(T)}{\sigma_T}=\frac{|T-n(n+1)/4|}{\sqrt{n(n+1)(2n+1)/24}}$$

If n is large and there are ties, then the term 1/2 in the computation of T serves as a continuity correction in the same manner as for the sign test

$$z = \frac{|T - n(n+1)/4 - 0.5|}{\sqrt{\dfrac{n(n+1)(2n+1)}{24} - \dfrac{\sum_{i=1}^{g}(t_j^3 - t_j)}{48}}}$$

where t_i refers to the number of differences with the same absolute value in the *i*th tied group and *g* is the number of tied groups

An alternative variance formula for Var(T)

$$Var(T) = \frac{\sum_{j=1}^{n} r_j^2}{4}$$

where *rj* = rank of the absolute value of the *j*th observation

Reject if $Z_{\alpha/2} \leq Z_0$

- Solution :

$$E(T) = \frac{n(n+1)}{4} = \frac{40\times 41}{4} = 410$$

$$\mathrm{var}(T) = \frac{n(n+1)(2n+1)}{24} - \frac{\sum_{i=1}^{g}(t_j^3 - t_j)}{48}$$

$$= \frac{40\times 41\times 81}{24} - \frac{(14^3-14)+(10^3-10)+\ldots+(1^3-1)}{48} = 5449.75$$

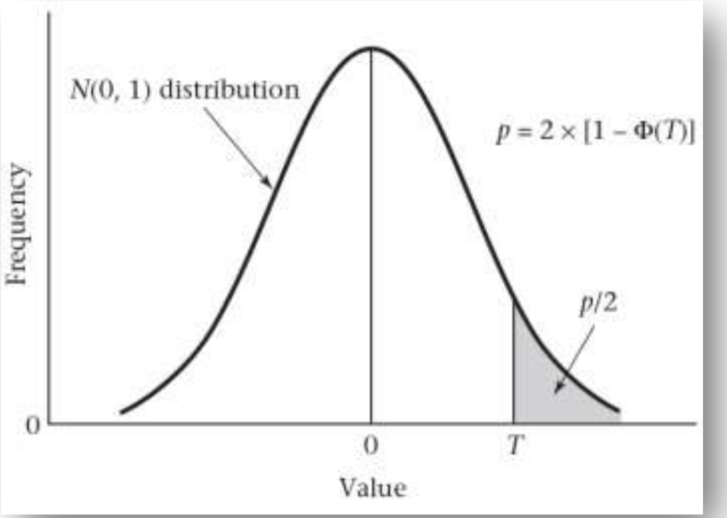
If the alternative variance formula is used, then

$$\mathrm{var}(T) = \frac{\sum_{j=1}^{n} r_j^2}{4} = \frac{14\times 7.5^2 + 10\times 19.5^2 + \ldots + 40^2}{4} = 5449.75$$

- **Solution:**

Thus $\sqrt{\text{var}(T)} = \sqrt{5449.75} = 73.82$

$$z = \frac{|T - n(n+1)/4| - 0.5}{sd(T) = \sqrt{\text{var}(T)}} = \frac{|248 - 410| - 0.5}{73.82} = 2.19$$



The *p*-value of the test is given by

$$P = 2 \times [1 - \Phi(2.19)] = 2 \times (1 - 0.9857) = 0.029$$

Conclusion: there is a significant difference between ointments

Single population inferences-SIGN TEST

- Tests one population median
- Corresponds to t-test for one mean
- Assumes population is continuous
- di=values-the known median
- Can use normal approximation if $n \geq 30$

3. Mann-Whitney U Test: Comparing two independent populations

Wilcoxon Rank Sum Test, also called as *Mann and Whitney U-test*

Wilcoxon RANK SUM TEST

- Tests two independent population probability distributions
- Corresponds to *t*-test for two independent means
- Can use normal approximation if $n_i \geq 10$

Wilcoxon Rank Sum Test

- The original data from two independent samples are transformed into their ranks. It tests whether the two population medians are equal or not.
- The two samples are temporarily combined, and the ranks of the combined data values are calculated. Then the smaller sum of the ranks is used to calculate the test statistic.

The null and alternative hypotheses:

H_0: The distributions of two populations are identical

H_1: The two population distributions are not identical

- Example 2: One researcher extracted information from medical records of patients with stroke, to explore the influence of admission NIHSS score on acute ischemic stroke's prognosis. Are the NIHSS scores different between the 10 patients with favorable outcome and 12 patients with poor prognosis?

NIH Stroke Scale Score	Stroke Severity
0	No stroke symptoms
1-4	Minor stroke
5-15	Moderate stroke
16-20	Moderate to severe stroke
21-42	Severe stroke

Table 3.1 the NIHSS scores of 10 AIS patients with good outcome and 12 with poor outcome

Favorable outcome			Poor outcome	
NIHSS score	Rank		NIHSS score	Rank
1	1		2	4.5
1	2		3	8
2	4.5		4	10
2	4.5		5	11
2	4.5		9	14
3	8		10	16
3	8		10	16
7	12		11	18
8	13		11	19
10	16		11	20
			12	21
			12	22
n_1=10	T_1=73.5		n_2=12	T_2=179.5

ANALYZE PROCEDURE

Preliminary Steps of the Test

1. Scores are ranked in ascending order, irrespective of which experimental group they come from (combine the data from the two groups);
2. Tied scores take the mean of the ranks they occupy;
3. Sum the ranks in the two groups;
4. T=T with smaller sample size.

RATIONAL

- Imagine two samples of scores drawn at random from the same population
- There should be a similar number of high and low ranked scores in each original group
  - if you sum the ranks in each group, the totals should be about the same
  - this is the null hypothesis

RATIONAL

- If however, the two samples are from different populations with different medians then most of the scores from one sample will be lower in the ranked list than most of the scores from the other sample
  - the sum of ranks in each group will differ

average rank in the combined sample is $(1 + n_1 + n_2)/2$

- 1. State the hypotheses

H_0: The two populations median are identical, M1=M2;

H_a: The two populations' median are not identical, M1≠M2；

α=0.05.

- 2. Find the Test Statistic

$T=T_{n1}=73.5$

- 3. Find the critical value

Here, n1=10， n_2-n_1 = 2

T critical= 76~154

- 4. Conclusion

Reject H_0. The two populations' median are not identical.

Large sample case

mean

If n_1 (smaller sample size)>10 or n_2 - n_1>*10*

$$z = \frac{\left|T - \boxed{n_1(N+1)/2}\right| - 0.5}{\boxed{\sqrt{n_1 n_2 (N+1)/12}}}$$

Standard deviation

Adjusted Zc

$$z_c = z / \sqrt{c}$$

$$C = 1 - \sum(t_j^3 - t_j)/(N^3 - N)$$

Rejection region:

$z > z_{1-\alpha/2}$

[probability - How can I find the expected value and variance of the Wilcoxon Sum of Ranks test? - Cross Validated (stackexchange.com)]

(2) the ***Mann and Whitney U-test*** → compare two ind populations

1.Compare x_i with y_i.

u_1= number of pairs $x_i > y_i$

u_2= number of pairs $x_i < y_i$

and $u_1 + u_2 = n_1 n_2$.

2.Reject H_0 if u_1 is large or u_2 is small.

Two **Test Statistics** are related as follows:

$$u_1 = w_1 - \frac{n_1(n_1+1)}{2}, u_2 = w_2 - \frac{n_2(n_2+1)}{2}$$

Advantage of the Mann-Whitney test:

Same distribution (whether u_1 or u_2) & Distribution range : $[0, n_1 n_2]$

P- value

$=P\{U \geq u_1\}=P\{U \leq u_2\}$

At significant level α, we reject H_0 if P-value ≤ α or $u_1 \geq u_{n_1, n_2, \alpha}$.

Denote: $u_{n_1, n_2, \alpha}$ - the upper α critical point.

For large n_1 and $n_{2,}$ the null distribution of U is Normal distributed. U =min(U1, U2)

$$E(U)=\frac{n_1 n_2}{2}, Var(U)=\frac{n_1 n_2 (N+1)}{12}$$

Z-test(Large Sample)

Test Statistic:

$$Z=\frac{u_1-\frac{n_1 n_2}{2}-\frac{1}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}=\frac{u_1-E(U)-\frac{1}{2}}{\sqrt{Var(U)}}$$

We **reject** H_0 at significant level α, if $z \geq z_\alpha$

Or

$$u_1 \geq \frac{n_1 n_2}{2}+\frac{1}{2}+z_\alpha \sqrt{\frac{n_1 n_2 (N+1)}{12}} \approx u_{n_1, n_2, \alpha}$$

Two-sided test,

Test Statistics: $u_{max} = \max(u_1 , u_2)$

or $u_{min} = \min(u_1 , u_2)$

P-value = $2P\{U \geq u_{max}\}=2P\{U \leq u_{min}\}$

Example 3:

Failure Times of Capacitors (Wilcoxon-Mann-Whitney Test)

18 capacitors, 8 under control group and 10 under stressed group

Perform the Wilcoxon-Mann-Whitney test to determine if thermal stress significantly reduces the time to failure of capacitors. α = 0.05.

Times to Failure for Two Capacitor Groups			
Control Group		Stressed Group	
5.2	17.1	1.1	7.2
8.5	17.9	2.3	9.1
9.8	23.7	3.2	15.2
12.3	29.8	6.3	18.3
		7.0	21.1

Ranks of Times to Failure			
Control Group		Stressed Group	
4	13	1	7
8	14	2	9
10	17	3	12
11	18	5	15
		6	16

$n_1 = 8$ $n_2 = 10$

The rank sums are

$w_1 = 4+8+10+11+13+14+17+18$
$= 95$

$w_2 = 1+2+3+5+6+7+9+12+15+16$
$= 76$

$$u_1 = w_1 - \frac{n_1(n_1+1)}{2} = 95 - \frac{(8)(9)}{2} = 59$$

$$u_2 = w_2 - \frac{n_2(n_2+1)}{2} = 76 - \frac{(10)(11)}{2} = 21$$

$$H_0 : F_1 = F_2 \text{ vs. } H_1 : F_1 < F_2$$

Let F_1 be c.d.f of the control group.

F_2 be c.d.f of the stressed group.

Check that $u_1 + u_2 = n_1 n_2 = 80$.

From Table A.11 P-Value=0.051

Large sample Z-test:

$$Z = \frac{u_1 - n_1 n_2 / 2 - 1/2}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}$$

$$= \frac{59 - (8)(10)/2 - 1/2}{\sqrt{\frac{(8)(10)(19)}{12}}} = 1.643$$

Conclusion:

yields the P-Value= 1- $\Phi(1.643)$

=0.0502

Table A.11 Upper-Tail Probabilities of the Null Distribution of the Wilcoxon-Mann-Whitney Statistic

n_1	n_2	w_1	u_1	$P(W \geq w_1)=P(U \geq u_1)$
8	8	84	48	0.052
	8	87	51	0.025
	8	90	54	0.010
	8	92	56	0.005
	9	89	53	0.057
	9	93	57	0.023
	9	96	60	0.010
	9	98	62	0.006
	10	95	59	0.051
	10	98	62	0.027
	10	102	66	0.010
	10	104	68	0.006

For large n_1 and n_2, the null distribution of U is Normal distributed.

$$E(U) = \frac{n_1 n_2}{2}, Var(U) = \frac{n_1 n_2 (N+1)}{12}$$

Z-test(Large Sample)

Test Statistic:

$$Z = \frac{u_1 - \frac{n_1 n_2}{2} - \frac{1}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}} = \frac{u_1 - E(U) - \frac{1}{2}}{\sqrt{Var(U)}}$$

We **reject** H_0 at significant level α, if $z \geq z_\alpha$

Or

$$u_1 \geq \frac{n_1 n_2}{2} + \frac{1}{2} + z_\alpha \sqrt{\frac{n_1 n_2 (N+1)}{12}} \approx u_{n_1, n_2, \alpha}$$

Two-sided test,

Test Statistics: $u_{max} = \max(u_1, u_2)$

or $u_{min} = \min(u_1, u_2)$

P-value = $2P\{U \geq u_{max}\} = 2P\{U \leq u_{min}\}$

EXAMPLE 4 : LARGE-SAMPLE MANN-WHITNEY U TEST

Score	Program	Rank	Score Rank Sum
85	1	20.0	20.0
87	1	21.0	41.0
92	1	27.0	68.0
98	1	30.0	98.0
90	1	26.0	124.0
88	1	23.0	147.0
75	1	17.0	164.0
72	1	13.5	177.5
60	1	6.5	184.0
93	1	28.0	212.0
88	1	23.0	235.0
89	1	25.0	260.0
96	1	29.0	289.0
73	1	15.0	304.0
62	1	8.5	312.5

Score	Program	Rank	Score Rank Sum
65	2	10.0	10.0
57	2	4.0	14.0
74	2	16.0	30.0
43	2	2.0	32.0
39	2	1.0	33.0
88	2	23.0	56.0
62	2	8.5	64.5
69	2	11.0	75.5
70	2	12.0	87.5
72	2	13.5	101.0
59	2	5.0	106.0
60	2	6.5	112.5
80	2	18.0	130.5
83	2	19.0	149.5
50	2	3.0	152.5

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = (15)(15) + \frac{(15)(15+1)}{2} - 312.5 = 32.5$$

$$E[U] = \frac{n_1 n_2}{2} = \frac{(15)(15)}{2} = 112.5$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(15)(15)(15+15+1)}{12}} = 24.109$$

$$z = \frac{U - E[U]}{\sigma_U} = \frac{32.5 - 112.5}{24.109} = -3.32$$

Since the test statistic is z = -3.32, the p-value ≈ 0.0005, and H_0 is rejected.

4. Kruskal-Wallis test: Completely Randomized Design

Kruskal-Wallis Test

KRUSKAL-WALLIS H-TEST

- Tests the equality of more than two (p) population probability distributions
- Corresponds to ANOVA for more than two means
- Used to analyze completely randomized experimental designs
- Uses χ^2 distribution with $p - 1$ df
  - if sample size $n_j \geq 5$

- Example 5: To study the effect of arginine on lymphocyte transformation function in mice after amputation, 21 mice were divided into 3 groups: control group A, amputation group B, amputation +arginine treatment group C. Then ^{3}H absorb value (CPM) was measured to show the spleen lymphocyte proliferation stimulated by heparin enzyme (HPA) . The measurement is shown in Table 4.1, try to analyze whether the spleen lymphocyte proliferation in these 3 groups are different. (The population variance are not equal).

ANALYZE PROCEDURE

- 1.State Hypotheses

H_0: All of *the population medians are all equal.*

H_a: Not all of *the population medians are equal.*

α=*0.05.*

Kruskal-wallis Test

Preliminary Steps of the Test

1. Temporarily combine the three samples and arrange them in increasing order.
2. Rank the data values from smallest to largest. Resolve ties using the mean rank.
3. Calculate the sum of the ranks for each group, $R1$, $R2$, and $R3$.
4. Calculate H (the test statistic).

Table 4.1 The spleen lymphocyte proliferation stimulated by heparin enzyme (HPA) （measured as ^{3}H *absorb value* ）

Group A (1)	Rank (2)	Group B (3)	Rank (4)	Group C (5)	Rank (6)
3012	11	2532	8	8138	15
9458	18	4682	12	2073	6
8419	16	2025	5	1867	4
9580	19	2268	7	885	2
13590	21	2775	9	6490	13
12787	20	2884	10	9003	17
6600	14	1717	3	690	1
R_i	$R_1 = 119$		$R_2 = 54$		$R_3 = 58$
n_i	7		7		7

Rational

Total variability of ranks (TV)

Variability between group (BV)

Variability within group (WV)

Group i=1 to k

id j=1 to ni	Group1	2	.....	k
	R_{11}			R_{k1}
			Rij	
	Sum=$R_{1.}$	$R_{2.}$		$R_{k.}$

$$\textit{average}\ \text{rank}=\frac{1+2+\cdots+n}{n}=\frac{n+1}{2}$$

$$\textit{The average of the total sum of squares of rank}(\text{aTV})=\frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(R_{ij}-\frac{n+1}{2})^2$$

$$=\frac{1}{n-1}\sum_{i=1}^{n_i}(i-\frac{n+1}{2})^2=\frac{1}{n-1}(\sum_{i=1}^{n_i}i^2-\frac{n(n+1)^2}{4})$$

$$=\frac{1}{n-1}(\frac{n(n+1)(2n+1)}{6}-\frac{n(n+1)^2}{4})=\frac{n(n+1)}{12}$$

$$\boxed{BV=\sum_{i=1}^{k}n_i(\frac{R_{i.}}{n_i}-\frac{n+1}{2})^2}$$

$$\boxed{KW=\frac{BV}{aTV}=\frac{12}{n(n+1)}\sum_{i=1}^{k}n_i(\frac{R_{i.}}{ni}-\frac{n+1}{2})^2=\frac{12}{n(n+1)}\sum_{i=1}^{k}n_i\frac{R_{i.}}{ni}-3(n+1)}$$

2. Calculate the test statistic

Squared total of each group

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1)$$

$$N = \sum N_i$$

R_i and n_i is the rank sum and sample size of the ith group, respectively.

$$H = \frac{12}{21(21+1)} \left(\frac{119^2}{7} + \frac{54^2}{7} + \frac{58^2}{7} \right) - 3(21+1) = 9.848$$

Adjusted H_c (if there are many observations have the same rank)

$$H_c = \frac{H}{c}$$

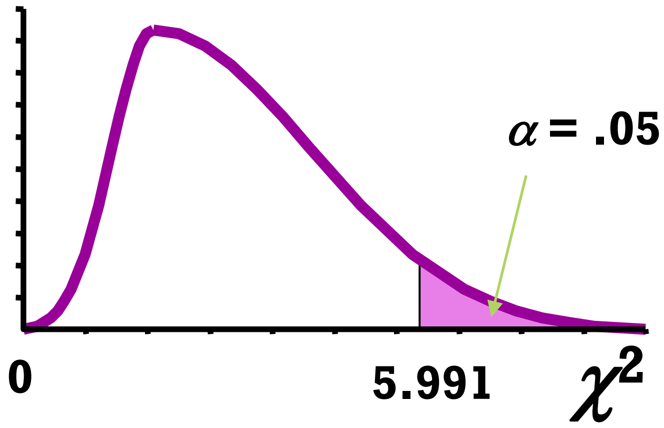
$$C = 1 - \sum (t_j^3 - t_j)/(N^3 - N)$$

3. Find the critical and state conclusion

① If group $k=3$， and each group has $n_i \le 5$， check *H table;*

② If each $n_i > 5$， H or H_c is approximately distributed as a χ^2 with $v=k-1$， check χ^2 table。

In this case, $n_i = 7>5$， then check χ^2 table， $\chi^2_{0.05,\ 2} = 5.99$， $H = 9.848 > \chi^2_{0.05,\ 2} = 5.99$， therefore， $P < 0.05$， reject H_0， not all of the population medians are identical.



Further analysis (pairwise comparisons of average ranks)

If the null hypothesis in the Kruskal-Wallis test is rejected, then we may wish, in addition, compare each pair of populations to determine which are different and which are the same.

The pairwise comparison test statistic:

$$D = \left| \bar{R}_i - \bar{R}_j \right|$$

where $\bar{R}_i$ is the mean of the ranks of the observations from population i.

The critical point for the paired comparisons:

$$C_{KW} = \sqrt{(\chi^2_{\alpha, k-1}) \left[\frac{n(n+1)}{12} \right] \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

Reject if $D > C_{KW}$

5. Friedman test: Randomized Block Design/repeated measures

Friedman test

FRIEDMAN TEST

The **Friedman test** is a nonparametric version of the randomized block design ANOVA or repeated measured ANOVA. Sometimes this design is referred to as a two-way ANOVA with one item per cell because it is possible to view the blocks as one factor and the treatment levels as the other factor. The test is based on ranks.

The Friedman hypothesis test:

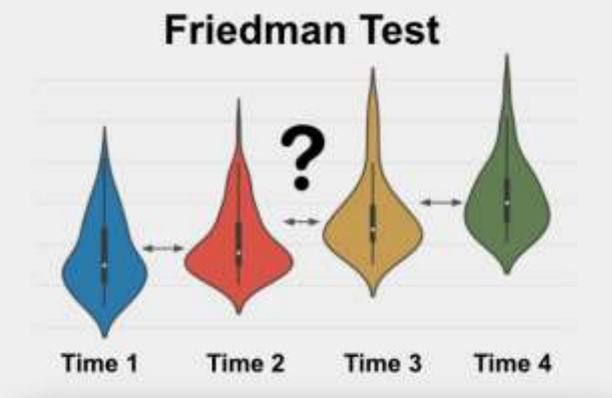
H_0: The distributions of the k treatment populations are identical

H_1: Not all k distribution are identical

The Friedman test statistic (large sample size case):

$$\chi^2 = \frac{12}{nk(k+1)}\sum_{j=1}^{k} R_j^2 - 3n(k+1)$$

The degrees of freedom for the chi-square distribution is (k – 1).



	Treatment		
Block	1	2	3
1	3.2	3.1	2.4
2	2.8	3.0	1.7
3	4.5	5.0	3.9
4	2.5	2.7	2.6
5	3.7	4.1	3.5
6	2.4	2.4	2.0

- Example 6: A university uses students' comprehensive scores to evaluate the teaching effect of courses. Now 10 medical students are randomly selected to evaluate the teaching effect of three basic medical courses, as shown in Table 5.1, and try to compare whether the teaching effect of these three basic medical courses is the same.

Table 5.1 Comparison of the comprehensive scores of 10 medical students on the teaching effects of three basic medical courses

student	Anatomy		Physiology		Histoembryology	
	score	rank	score	rank	score	rank
1	4.0	1.5	4.0	1.5	5.0	3
2	2.5	1	4.0	2.5	4.0	2.5
3	4.0	2	3.5	1	4.5	3
4	3.5	1	4.0	2	5.0	3
5	3.5	2	3.0	1	4.0	3
6	2.5	1	3.5	2.5	3.5	2.5
7	4.0	3	3.5	1.5	3.5	1.5
8	3.5	1.5	3.5	1.5	4.5	3
9	3.0	1	4.0	2.5	4.0	2.5
10	2.5	1	3.0	2	4.0	3
R_i		15		18		27

FRIEDMAN F_R-TEST FOR A RANDOMIZED BLOCK DESIGN

H_0: The probability distributions for the k treatments are identical

H_a: At least two of the probability distributions differ in location

LOGIC

- The basic idea of Friedman's rank sum test is that the observed values in each block group are ranked in ascending order; If the effects of all treatments are the same, the rank 1, 2,... k (k is the number of treatment groups) should appear in each treatment group (column) with equal probability, and the rank sum of each treatment group should be roughly equal, and it is unlikely to be significantly different. If the sample rank of each treatment group and R1, R2,... vary greatly, it is reasonable to doubt whether the overall distribution is the same among treatment groups.

ANALYZE PROCEDURE

1. Rank order the scores SEPARATELY FOR EACH SUBJECT'S DATA with the smallest score getting a value of 1. If there are ties (within the scores for a subject) each receives the average rank they would have received;
2. Compute the sum of the ranks for each condition;
3. Compute Friedman's M.

$$M = \sum_{i=1}^{k} \left(R_i - \bar{R} \right)^2$$

k is the number of groups or treatment levels

$$\bar{R} = \frac{15+18+27}{3} = 20$$

$$M = \sum_{i=1}^{k}\left(R_i - \bar{R}\right)^2 = (15-20)^2 + (18-20)^2 + (27-20)^2 = 78$$

Look at the M table, determine the critical value， $M_{0.05}$=62 (when k=3, b=10)

Reject H0

FOR LARGE SAMPLES (K>5 OR B>13)

- Approximate chi-square method

$$x^2 = \frac{12}{bk(k+1)} \sum_{i=1}^{k} R_i^2 - 3b(k+1)$$

Look at the x2 table with k-1 degrees of freedom

If there are too many ties

$$x_c^2 = \frac{x^2}{c} \qquad C = 1 - \frac{\sum (t_j^3 - t_j)}{bk(k^2-1)}$$

where R_i is the sum of the ranks for sample i.
b is the number of independent blocks
k is the number of groups or treatment levels

where t_j is the number of jth observation with the same rank in each block group

- For this example: b=10, k=3, R1=15, R2=18, R3=27

$$x^2 = \frac{12}{10\times 3\left(3+1\right)}\sum_{i=1}^{k}(15^2+18^2+27^2)-3\times 10(3+1)=7.8$$

$$C=1-\frac{\sum(t_j^3-t_j)}{bk(k^2-1)}=1-\frac{1}{10\times 3(3^2-1)}[(2^3-2)+(2^3-2)+....+(2^3-2)]=0.85$$

$$x_c^2=\frac{x^2}{c}=9.18,\ \text{where } x_{0.05,2}^2=5.99$$

Reject H0

6. Rank correlation

Spearman’s Rank Correlation & Kendall rank correlation

WHEN TO USE IT-SPEARMAN

- The Spearman's rank-order correlation is the nonparametric version of the Pearson product-moment correlation. Spearman's correlation coefficient, (ρ, also signified by r_s) measures the strength and direction of association between two ranked variables.

- Example 7. A project team conducted a survey on sleep and anxiety among sophomores in one high school in Suzhou, Jiangsu Province. In this project, the Pittsburgh Sleep Quality Index Scale (PSQI) and Student Anxiety Scale (SAS) were used to investigate the sleep and anxiety status. Please calculate the Spearman level correlation coefficient rs between sleep and anxiety.

Table 6.1 scores for sleep quality and anxiety among 10 students

id	Sleep quality score x	Sleep quality rank P_i	Anxiety score y	Anxiety rank Q_i	Rank difference $d=P_i - Q_i$	d^2	P_i^2	Q_i^2	$P_i \ Q_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	4	6.0	5	7.0	-1.0	1.00	36.00	49.00	42.00
2	8	9.5	8	10.0	-0.5	0.25	90.25	100.00	95.00
3	5	7.0	6	8.0	-1.0	1.00	49.00	64.00	56.00
4	8	9.5	4	5.5	4.	16.00	90.25	30.25	52.25
5	3	4.0	4	5.5	-1.5	2.25	16.00	30.25	22.00
6	2	1.5	3	3.5	-2.0	4.00	2.25	12.25	5.25
7	3	4.0	3	3.5	0.5	0.25	16.00	12.25	14.00
8	7	8.0	7	9.0	-1.0	1.00	64.00	81.00	72.00
9	3	4.0	0	1.0	3.0	9.00	16.00	1.00	4.00
10	2	1.5	1	2.0	-0.5	0.25	2.25	4.00	3.00
Total	45	55.0	41	55.0	-	35.00	382.00	384.00	365.50

SPEARMAN’S RANK CORRELATION PROCEDURE

1. Assign ranks for sleep and separately;
2. when you have two identical values in the data (called a "tie"), you need to take the average of the ranks that they would have otherwise occupied.
3. Calculate differences, d_i , between each pair of ranks
4. Square differences, d_i^2, between ranks
5. Sum squared differences for each variable
6. Use shortcut approximation formula

- Similar to Pearson correlation, spearman rank-coefficient can be calculated as

The formula for when there are no tied ranks is:

$$r_s = \frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} = \frac{\sum(P_i - \overline{P})(Q_i - \overline{Q})}{\sqrt{\sum(P_i - \overline{P})^2 \sum(Q_i - \overline{Q})^2}} \quad \text{or } r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

The formula to use when there are tied ranks is:

$$r_s' = \frac{[(n^3-n)/6] - (T_x + T_y) - \sum d^2}{\sqrt{[(n^3-n)/6] - 2T_x}\sqrt{[(n^3-n)/6] - 2T_y}}$$

T_x (or T_y) $= \sum(t^3 - t)/12$ *Where* t is the number of x (or y) with the same value

where d_i = difference in paired ranks and n = number of cases.

$$r_s = \frac{\sum (P_i - \bar{P})(Q_i - \bar{Q})}{\sqrt{\sum (P_i - \bar{P})^2 \sum (Q_i - \bar{Q})^2}}$$

$$= \frac{365.5 - 55 \times 55 / 10}{\sqrt{(382 - 55^2 / 10)(384 - 55^2 / 10)}} = \frac{63}{\sqrt{79.5 \times 81.5}} = 0.783$$

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 35.00}{10 \times (100 - 1)} = 0.788$$

$$r_s^{'} = \frac{[(n^3 - n)/6] - (T_x + T_y) - \sum d^2}{\sqrt{[(n^3 - n)/6] - 2T_x} \sqrt{[(n^3 - n)/6] - 2T_y}} = \frac{\left[\left(10^3\text{-}10\right)/6\right]\text{-}\left(3+1\right)\text{-}35.00}{\sqrt{\left[\left(10^3\text{-}10\right)/6\right]\text{-}2\times 3}\sqrt{\left[\left(10^3\text{-}10\right)/6\right]\text{-}2\times 1}} = 0.783$$

RATIONAL

- if there were a perfect positive correlation between the two variables, then the ranks for each person on each variable would be the same and $rs = 1$. The less perfect the correlation, the closer to zero rs would be.
- Or, when n is a fixed value, if x and y are perfectly and positively correlated, then Σd_i^2 would have a minimal value of 0; if x and y are perfectly and negatively correlated, then Σd_i^2 would have a maximum value of

$$\sum d_i^2 = \sum (P_i - Q_i)^2 = \sum [(n+1-i)-i]^2 = n(n^2-1)/3$$

Hypothesis test

Two-Tailed Test

H_0: $\rho = 0$

H_a: $\rho \neq 0$

t Test for Spearman Rank Correlation

(1) Compute the test statistic

$$t_s = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}$$

which under the null hypothesis of no correlation follows a t distribution with $n-2$ degrees of freedom.

(2) For a two-sided level α test,

if $\quad t_s > t_{n-2,1-\alpha/2} \quad$ or $\quad t_s < t_{n-2,\alpha/2} = -t_{n-2,1-\alpha/2}$

then reject H_0; otherwise, accept H_0.

(3) The exact p-value is given by

$$p = 2 \times (\text{area to the left of } t_s \text{ under a } t_{n-2} \text{ distribution}) \qquad \text{if } t_s < 0$$
$$p = 2 \times (\text{area to the right of } t_s \text{ under a } t_{n-2} \text{ distribution}) \qquad \text{if } t_s \geq 0$$

(4) $\boxed{\text{This test is valid only if } n \geq 10.}$

KEY IDEAS

Distribution-Free Tests

Do not rely on assumptions about the probability distribution of the sampled population

Nonparametrics

Distribution-free tests that are based on **rank statistics**

One-sample nonparametric test for the population median – **sign test**

Nonparametric test for *two independent samples* – **Wilcoxon rank sum test**

Nonparametric test for *matched pairs* – **Wilcoxon signed rank test**

Nonparametric test for a completely randomized design – **Kruskal-Wallis test**

Nonparametric test for a randomized block /repeated mesurment design – **Friedman test**

Nonparametric test for rank correlation – **Spearman's test**