Statistics and Estimation for Computer Science

İstanbul Teknik Üniversitesi

Mustafa Kamasak, PhD

Version: 2022.3.15

Analysis of Variance (ANOVA)

Pairwise Comparison of Populations

- How can we compare means of K population?
- Compare population combinations by pairs using two-sample hypothesis testing
- $C(K,2)$ hypothesis testing are required
- Let significance level (type I error) to be α for all tests
- Prability of type-I error is inflated

$$P(\text{at least one type I error}) = 1 - \underbrace{P(\text{no type-I error})}_{(1-\alpha)^{C(K,2)}}$$

- Let $K = 5$ an $\alpha = 0.05$

$$P(\text{at least one type I error}) = 1 - (1-\alpha)^{10} = 0.4$$

Pairwise Comparison of Populations

- To prevent inflated significance level (type-I error) use Bonferroni correction
- If M tests will be performed on the same data use corrected significance level
$$\alpha_c = \frac{\alpha}{M}$$
- For $K = 5$ an $\alpha = 0.05$, use $\alpha_c = 0.05/10 = 0.005$
$$P(\text{at least one type I error}) = 1 - (1 - \alpha_c)^{10} \approx 0.05$$
- Even with Bonferroni correction, it is never a good idea to divide data
- Using all data to test a hypothesis will increase power of analysis

Analysis of Variance (ANOVA)

- Means of multiple populations (≥ 3) can be compared using the analysis of variances.
- Process was introduced by Sir Ronald Fisher.
- Assuming that there are K samples
  - H_0: $\mu_1 = \mu_2 = \cdots = \mu_K$
  - H_1: $\mu_i \neq \mu_j$ for at least one i, j ($i \neq j$)
- One-way or one-factor ANOVA investigates the mean of samples that vary with a single factor. For example, blood sugar level with respect to BMI factor.
- Two-way or two-factor ANOVA investigates the mean of samples that vary with a two factors. For example, blood sugar level with respect to BMI and age factors.
- More factors can also be investigated.

ANOVA Assumptions

Following criteria should be satisfied to use ANOVA:

- Each population should have normal distribution $\rightarrow$ Normality
- Variance of the populations should be identical $\rightarrow$ Homoscedaticity
- Samples should be independent $\rightarrow$ Independence

Definitions

Population	1	2	...	K
Pop. mean	μ_1	μ_2	...	μ_K
Pop. var.	σ^2	σ^2	...	σ^2
Sample size	n_1	n_2	...	n_K
Sample mean	$\overline{x}_1$	$\overline{x}_2$	...	$\overline{x}_K$
Sample var.	s_1^2	s_2^2	...	s_K^2

- Let x_{ki} denote the i^{th} observation in group k.

$$\overline{x}_k = \sum_{i=1}^{n_k} x_{ki}$$

$$s_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_{ki} - \overline{x}_k)^2$$

Definitions

- N: Total number of observations in all samples

$$N = \sum_{k=1}^{K} n_k$$

- T: Sum of all observed values

$$T = \sum_{k=1}^{K} \sum_{i=1}^{n_k} x_{ki} = \sum_{k=1}^{K} n_k \overline{x}_k$$

- $\overline{\overline{x}}$: Grand mean

$$\overline{\overline{x}} = \frac{T}{N}$$

Analysis of Variance – ANOVA

- ANOVA analyses the variance in the data (as its name suggests) and sources of variances.
- Total variance in the data: SST – sum of squares total

$$
\begin{aligned}
SST &= \sum_{k=1}^{K}\sum_{i=1}^{n_k}(x_{ki}-\overline{\overline{x}})^2 \\
&= \sum_{k=1}^{K}\sum_{i=1}^{n_k}(x_{ki}-\overline{x}_k+\overline{x}_k-\overline{\overline{x}})^2 \\
&= \sum_{k=1}^{K}\sum_{i=1}^{n_k}(x_{ki}-\overline{x}_k)^2+(\overline{x}_k-\overline{\overline{x}})^2+\underbrace{2(x_{ki}-\overline{x}_k)(\overline{x}_k-\overline{\overline{x}})}_{A}
\end{aligned}
$$

ANOVA

$$
\begin{aligned}
A &= \sum_{k=1}^{K} \sum_{i=1}^{n_k} 2(x_{ki} - \overline{x}_k)(\overline{x}_k - \overline{\overline{x}}) \\
&= 2 \sum_{k=1}^{K} \sum_{i=1}^{n_k} (x_{ki}\overline{x}_k - x_{ki}\overline{\overline{x}} - \overline{x}_k^2 + \overline{x}_k\overline{\overline{x}}) \\
&= 2 \sum_{k=1}^{K} n_k\overline{x}_k^2 - n_k\overline{x}_k\overline{\overline{x}} - n_k\overline{x}_k^2 + n_k\overline{x}_k\overline{\overline{x}} \\
&= 0
\end{aligned}
$$

where $\sum_{i=1}^{n_k} x_{ki} = n_k\overline{x}_k$.

Analysis of Variance – ANOVA

- Total variance in the data: SST – sum of squares total

$$SST = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(x_{ki} - \overline{\overline{x}})^2 = \underbrace{\sum_{k=1}^{K}\sum_{i=1}^{n_k}(x_{ki} - \overline{x}_k)^2}_{SSE} + \underbrace{\sum_{k=1}^{K}\sum_{i=1}^{n_k}(\overline{x}_k - \overline{\overline{x}})^2}_{SSTr}$$

- Total variance in the data has two sources of variance
  - Variation **between** samples/groups/treatments (intrasample): *SSTr or SSG* - sum of squares in treatment (group)

$$SSTr = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(\overline{x}_k - \overline{\overline{x}})^2 = \sum_{k=1}^{K} n_k(\overline{x}_k - \overline{\overline{x}})^2$$

  - Variation **within** samples/groups (intersample): *SSE* - sum of squares error

$$SSE = \sum_{k}\sum_{i}^{n_k}(x_{ki} - \overline{x}_k)^2 = \sum_{k=1}^{K}(n_k - 1)s_k^2$$

Analysis of Variance

- Degrees of freedoms for
  - SST: N-1 (due to $\overline{\overline{x}}$)
  - SSTr: K-1 (due to $\overline{\overline{x}}$)
  - SSE: N-K (due to $\{\overline{x}_1, \cdots, \overline{x}_K\}$)

Intuition of ANOVA

- If all populations have equal variance (σ^2), then SSE is [9]

$$SSE = \sum_k \sum_i^{n_k} (x_{ki} - \overline{x}_k)^2 \approx \sum_{k=1}^{K} (n_k - 1)s^2 = (N - K)s^2$$

$$\hat{\sigma}^2 = \frac{SSE}{N - K}$$

- Assuming all samples have the same mean μ (if H_0 is correct), $\overline{x}$ has the following distribution

$$\overline{x} \sim (\mu, \sigma^2/N)$$

- SSTR is then

$$SSTr = \sum_{k=1}^{K} n_k(\overline{x}_k - \overline{\overline{x}})^2 \approx N \sum_{k=1}^{K} (\overline{x}_k - \overline{\overline{x}})^2 = (K - 1)\underbrace{Ns^2_{\overline{x}}}_{\hat{\sigma}^2}$$

$$\hat{\sigma}^2 = \frac{SSTR}{K - 1}$$

[9] ANOVA assumes equal variance for all populations

Analysis of Variance

- Mean between-sample variance:

$$MSTr = \frac{SSTr}{dof} = \frac{SSTr}{K-1}$$

- Mean within-sample variance:

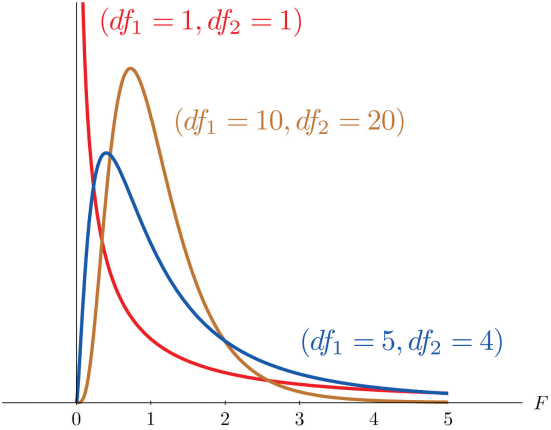
$$MSE = \frac{SSE}{dof} = \frac{SSE}{N-K}$$

- MSE is **always** an estimator of population variance, MSTr is an estimator of population variance **if $\mathbf{H_0}$ is correct**
- Define F-statistics as follows:

$$F = \frac{MSTr}{MSE}$$

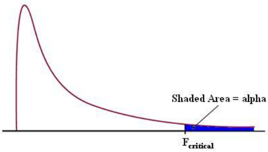
- If H_0 is correct, F-statistics has F-distribution with dof1 = K-1 and dof2 = N-K.

F Distribution



Steps of ANOVA

- Compute F statistic from data
- Decide significance level (α)
- Find critical F-value from F distribution table



F Distribution: Critical Values of F (5% significance level)

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	14
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.36
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86

- If $F > F_{critical}$ reject H_0 (group means are different), otherwise do not reject H_0.

ANOVA Exercise

- Consider 3 brands of batteries
- Using the same home appliance, the duration of batteries are measured (in minutes)
- With each brand 5 measurements are performed

Brand	Duration	Mean
A	220, 251, 226, 246, 260	240.6
B	244, 235, 232, 242, 225	235.6
C	252, 272, 250, 238, 256	253.6

- Test the hypothesis that claims the mean duration of all brands are equal using significance level of 0.05

ANOVA Exercise

- Grand mean $\bar{\bar{x}} \approx 243.27$

- Within group variation SSE

Brand	(duration-group mean)^2	Total
A	424.36, 108.16, 213.16, 29.16, 376.36	1151.20
B	70.56, 0.36, 12.96, 40.96, 112.36	237.12
C	2.56, 338.56, 12.96, 243.36, 5.76	603.2
Total	41151.20+237.12+603.20	1191.52

- dof of SSE is 15-3=12
- MSE=SSE/12=165.96

- Between group variation SSTr

  $SSTr = 5\times((240.6-243.27)^2+(235.6-243.27)^2+(253.6-243.27)^2) = 86$

- dof of SSTr is 3-1=2
- MSTr = SSTr/2=431.68

ANOVA Exercise

- F statistics is
$$f_{obs} = \frac{MSTr}{MSE} = \frac{432.68}{165.96} = 2.61$$
- Check out F table with dof1=2 and dof2=12 for critical F value

***F* Distribution: Critical Values of *F* (5% significance level)**

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.36	246.46	247.32	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.44	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33

- From this table $F_{critical} = 3.89$
- As $f_{obs} < F_{critical}$, retain H_0

Interactive ANOVA Visualization

Checkout

`https://demonstrations.wolfram.com/VisualANOVA/`

Two-Factor ANOVA

- Sometimes, there are multiple factors that may effect population means.
- To analyze the effect of two factors, two-factor (two-way) ANOVA is performed.
- Assume, factor A and factor B may effect population means

		Factor B				
		1	...	C	total	avg.
Factor A	1	$\{x_{11,1}, x_{11,2}, \dots\}$	...	$\{x_{1C,1}, x_{1C,2}, \dots\}$	x_{1*}	$\overline{x}_{1*}$
	2	$\{x_{21,1}, x_{21,2}, \dots\}$	...	$\{x_{2C,1}, x_{2C,2}, \dots\}$	x_{2*}	$\overline{x}_{2*}$
	R	$\{x_{R1,1}, x_{R1,2}, \dots\}$	...	$\{x_{RC,1}, x_{RC,2}, \dots\}$	x_{R*}	$\overline{x}_{R*}$
	total	x_{*1}	...	x_{*C}	T	
	avg.	$\overline{x}_{*1}$	...	$\overline{x}_{*C}$		$\overline{\overline{x}}$

Two-way ANOVA – Definitions

- R is the number of levels in factor A, C is the number of levels in factor B.
- $x_{rc,i}$ denotes observation i in row r and column c.
- n_{rc} is the number of observations row r and column c.
- This observation deviate from grand mean $\overline{\overline{x}}$

$$x_{rc,i} = \overline{\overline{x}} + \tau_r + \beta_c + (\tau\beta)_{rc} + e_{rc}$$

  - τ_r is the effect (deviation) of row level r
  - β_c is the effect (deviation) of column level c
  - $(\tau\beta)_{rc}$ is the effect of interaction of row level r and column level c
  - e_{rc} is within group error. $e_{rc} \sim \mathcal{N}(0, \sigma^2)$
- Total sum of deviations are always zero.

$$\sum_{r=1}^{R} \tau_r = \sum_{r=1}^{R} (\tau\beta)_{rc} = \sum_{c=1}^{C} \beta_c = \sum_{c=1}^{C} (\tau\beta)_{rc} = 0$$

Two-way ANOVA – Definitions

- Cell average

$$\overline{x}_{rc} = \sum_{i=1}^{n_{rc}} x_{rc,i}$$

- Row sums are

$$x_{r*} = \sum_{c=1}^{C} \sum_{i=1}^{n_{rc}} x_{rc,i}$$

- Row averages are

$$\overline{x}_{r*} = \frac{x_{r*}}{\sum_{c=1}^{C} n_{rc}}$$

- Number of row observations

$$n_{r*} = \sum_{c=1}^{C} n_{rc}$$

Two-way ANOVA – Definitions

- Number of column observations

$$n_{*c} = \sum_{r=1}^{R} n_{rc}$$

- Column sums are

$$x_{*c} = \sum_{r=1}^{R} \sum_{i=1}^{n_{rc}} x_{rc,i}$$

- Column averages are

$$\overline{x}_{*c} = \frac{x_{*c}}{\sum_{r=1}^{R} n_{rc}}$$

- Grand mean is

$$\overline{\overline{x}} = \frac{\sum_{r=1}^{R} x_{r*}}{\sum_{r=1}^{R} \sum_{c=1}^{C} n_{rc}} = \frac{\sum_{c=1}^{C} x_{*c}}{\sum_{r=1}^{R} \sum_{c=1}^{C} n_{rc}}$$

Two-way ANOVA – Hypothesis

- 3 hypothesis will be tested
  - Effect of factor A (rows)
    H_0: $\tau_r = 0$ for all r
    H_1: $\tau_r \neq 0$ for at least one r
  - Effect of factor B (columns)
    H_0: $\beta_c = 0$ for all c
    H_1: $\beta_c \neq 0$ for at least one c
  - Interaction between factor A and factor B
    H_0: $(\tau\beta)_{rc} = 0$ for all r, c
    H_1: $(\tau\beta)_{rc} \neq 0$ for at least one r or c

Two-way ANOVA – Sources of Variance

- Analyze variance in data and sources of variance.
  - SST - sum of squares total
  - SSR - sum of squares in rows
  - SSC - sum of squares in columns
  - SSRC - sum of squares in rows & columns

$$SST = SSR + SSC + SSRC$$

Two-way ANOVA – Sources of Variance

$$SST = \sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{n_{rc}}(x_{rc,i} - \overline{\overline{x}})^2$$

$$SSR = n_{r*}\sum_{r=1}^{R}(\overline{x}_{r*} - \overline{\overline{x}})^2$$

$$SSC = n_{*c}\sum_{c=1}^{C}(\overline{x}_{*c} - \overline{\overline{x}})^2$$

$$SSRC = \sum_{r=1}^{R}\sum_{c=1}^{C} n_{rc}(\overline{x}_{rc} - \overline{x}_{r*} - \overline{x}_{*c} + \overline{\overline{x}})^2$$

$$SSE = \sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{n_{rc}}(x_{rc,i} - \overline{x}_{rc})^2$$

Two-way ANOVA – Sources of Variance

- degrees of freedom for each source of variation is as follows:
  - SST: dof = N - 1 (due to $\overline{\overline{x}}$)
  - SSR: dof = R - 1 (due to $\overline{\overline{x}}$)
  - SSC: dof = C - 1 (due to $\overline{\overline{x}}$)
  - SSRC: dof = (R-1)(C-1) (due to $\overline{x}_{r*}$ and $\overline{x}_{*c}$)
  - SSE: dof = N-RC (due to $\overline{x}_{rc}$)

Two-way ANOVA – Mean of Variance

- Mean sum of square for each source of variance is divided by its dof

$$MST = \frac{SST}{N-1}$$

$$MSR = \frac{SSR}{R-1}$$

$$MSC = \frac{SSC}{C-1}$$

$$MSRC = \frac{SSRC}{(R-1)(C-1)}$$

$$MSE = \frac{SSE}{N-RC}$$

Two-way ANOVA – Test Statistics

- To check effect of rows use

$$F_r = \frac{MSR}{MSE}$$

- To check effect of columns use

$$F_c = \frac{MSC}{MSE}$$

- To check interaction between rows and columns

$$F_{rc} = \frac{MSRC}{MSE}$$

Two-way ANOVA – Hypothesis

- Determine significance level for 3 hypothesis, $\alpha_1, \alpha_2, \alpha_3$
  - Effect of factor A (rows)
  - Effect of factor B (columns)
  - Interaction between factor A and factor B
- Find critical F-value from F distribution table $\rightarrow$ Different tables due to different dof1, dof2, and α values.
- If $F_r > F_{r,critical}$ reject H_0 (row effect), otherwise do not reject H_0.

Nonparametric Methods

Nonparametric Methods

- Sometimes parametric methods cannot be used
  - Distribution is unknown (infact hypothesis about distribution need to be tested)
  - Assumptions of parametric methods do not hold
  - Skewed distribution and small sample size
  - Outliers that cannot be removed
  - Data is ordinal (ordered but not scaled)
- In these cases, nonparametric test methods are used
- Ranks of observations are used instead of observation values
- Nonparametric tests typically have lower test power (than parametric tests) as they have fewer (or no) assumptions

Sign Test

- Sign test can be used to test:
  - To test about median value of a population. $\rightarrow$ One-sample sign test
  - To test if two populations have identical medians when observations are paired. $\rightarrow$ Two-sample sign test
- It is a "non-parametric" or "distribution free" test (no assumptions about data distribution)
- It only requires continuous distribution for data.
- It canbe used for numeric and ordinal data.
- It is easy to use.
- It has low power due to least amount of assumptions.

One Sample Sign Test

- Let η denote the median of a population.

$$P(X < \eta) = P(X > \eta) = 0.5$$

- Hypothesis are:

$$\begin{aligned} \mathrm{H}_0 &: \eta = m \\ \mathrm{H}_1 &: \eta \neq m \end{aligned}$$

- Equivalently:

$$\begin{aligned} \mathrm{H}_0 &: p = 1/2 \\ \mathrm{H}_1 &: p \neq 1/2 \end{aligned}$$

where $p = P(X < m)$

One Sample Sign Test

- Let T $=$ number of observations below m (median of H_0).
- Alternatively, T1 $=$ number of observations above m.
- Values equal to m (that is theoretically impossible, but practically possible due to observation errors etc.) are ignored.

- In order to obtain T or T1
  - Subtract m from each observation x_i: $d_i = x_i - m$
  - Insert sign "$-$" if $d_i < 0$, "$+$" if $d_i > 0$, and 0 if $d_i = 0$
  - T is the number of "-", and T1 is the number of "$+$"
  - Sign of differences are counted (hence the name "sign test")

One Sample Sign Test

- If T (or T1) $\approx$ N/2, H_0 cannot be rejected.
- If T (or T1) is large/small, H_0 can be rejected. Large/small?
- T is an rv, and T $\sim$ Binomial(N,1/2) if H_0 is correct.
- p-value is

$$p = 2\min(P(T \leq T_{observed}), P(T \geq T_{observed}))$$

- Note: p-value obtained using T and T1 will be equal due to min operator!
- Note: Multiplication with 2 is due to the two-tailed nature of the test.

One Sample Sign Test – Example

- Assume completion time (in min) of an exam by 10 students are as follows:

  ```
  18.58 21.11 31.41 19.13 29.75
  19.30 21.23 27.22 19.26 22.28
  ```

- Instructor's hypothesis about the median of exam completion time is 20 min.
- Subtract 20 from each observation

$$\text{Signs} = \{-, +, +, -, +, -, +, +, -, +\}$$

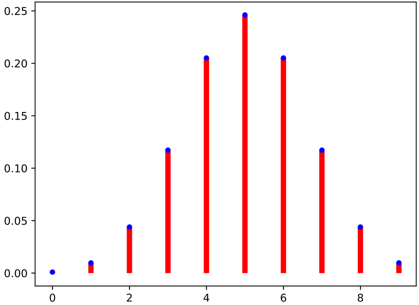
- $T_{observed} = |\{18.58, 19.13, 19.30, 19.26\}| = 4$
- T $\sim$ Binomial(10,1/2)
- If H_0 is correct, then $E(T) = Np = 5$ and $\sigma_T^2 = Np(1-p) = 2.5$

One Sample Sign Test – Example

- Remember binomial distribution:

$$P(T_{observed} = k) = \binom{N}{k} p^k (1-p)^{N-k}$$

for $k \in \{0, .., N\}$.



- In this case $P(T \leq 4) < P(T \geq 4)$, hence $p = 2P(T \leq 4)$

One Sample Sign Test – Example

```
import numpy as np
from scipy.stats import binom

n=10
p=.5

x = np.arange(0,11)
binom_pdf = binom.pmf(x, n, p)
print(2*sum(binom_pdf[:5]))
```

- p = 0.75
- H_0 cannot be rejected with significance level of $\alpha = 0.05$.

One Sample Sign Test Table Lookup

- Alternatively, check critical value from following table
- Critical value for two-sided $\alpha = 0.05$ and $N = 10$ is 1 (or 9)
- If $T_{obs} = 1$ or 9, H_0 is rejected
- Let $X \sim$Binom(10,0.5), then

$$P(X = 0) + P(X = 1) \approx 0.01 \text{ and } p \approx 0.02$$
$$P(X = 0) + P(X = 1) + P(X = 2) \approx 0.05 \text{ and } p \approx 0.11$$

- As $T_{obs} = 4$ in our example, H_0 should be retained

TABLE • VIII Critical Values for the Sign Test

r_α^*

n \ α	0.10 0.05	0.05 0.025	0.01 0.005	Two-sided tests One-sided tests	n \ α	0.10 0.05	0.05 0.025	0.01 0.005	Two-sided tests One-sided tests
5	0				23	7	6	4	
6	0	0			24	7	6	5	
7	0	0			25	7	7	5	
8	1	0	0		26	8	7	6	
9	1	1	0		27	8	7	6	
10	1	1	0		28	9	8	6	
11	2	1	0		29	9	8	7	
12	2	2	1		30	10	9	7	
13	3	2	1		31	10	9	7	
14	3	2	1		32	10	9	8	
15	3	3	2		33	11	10	8	
16	4	3	2		34	11	10	9	
17	4	4	2		35	12	11	9	
18	5	4	3		36	12	11	9	
19	5	4	3		37	13	12	10	
20	5	5	3		38	13	12	10	
21	6	5	4		39	13	12	11	

Normal Approximation of Binomial Distribution

- When $Np > 5$ and $N(1-p) > 5$, we can approximate binomial distribution with normal distribution.
- $T \sim \mathcal{N}(Np, Np(1-p))$
- When a discrete distribution is approximated by a continuous distribution, approximation can be improved by *continuity correction*.
- For example, when $X \sim$binomial(N,p) is approximated with $Y \sim \mathcal{N}(Np, Np(1-p))$,

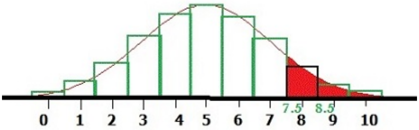
$$P(X \geq k) \approx P(Y \geq k - 0.5)$$

and

$$P(X \leq k) \approx P(Y \leq k + 0.5)$$

Normal Approximation of Binomial Distribution

- For example, with N=10 and p=0.5: X ~binomial(10,0.5) and $Y \sim \mathcal{N}(5, 2.5)$
  - $P(X \geq 8)$ is better approximated with $P(Y \geq 7.5)$ compared to $P(Y \geq 8)$
  - $P(X \leq 8)$ is better approximated with $P(Y \leq 7.5)$ compared to $P(Y \leq 8)$



https://www.statisticshowto.datasciencecentral.com/what-is-the-continuity-correction-factor/

Normal Approximation of Binomial Distribution

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm,binom

n=50
p=.5

t = [24.00, 22.31, 27.59, 19.73, 19.62, 23.51, 15.58, 28.98, 24.33, 19.58, \
     18.00, 12.99, 20.66, 28.97, 23.37, 18.14, 14.33, 27.39, 28.30, \
     21.82, 9.65, 23.97, 24.25, 21.19, 22.33, 18.68, 32.55, 20.68,  \
     24.88, 23.39, 20.0, 19.72, 20.77, 16.37, 23.80, 41.28, 35.08,  \
     24.39, 20.88, 26.60, 17.35, 20.70, 19.20, 20.05, 27.10, 18.01,  \
     12.40, 21.36, 20.0, 21.07]

# find t_obs
t1 = np.sum(np.less(t,20*np.ones(50)) )
t2 = np.sum(np.greater(t,20*np.ones(50)) )

# we want t1<t2
if t1>t2:
  t1,t2 = t2,t1
# ... 1 of 2 ...
```

Normal Approximation of Binomial Distribution

```python
# ... 2 of 2 ...

# find more extreme observation
t_obs  = t1 if abs(t1-n*p) > abs(t2-n*p) else t2
print(t1,t2,t_obs)

# exact p-value
x = np.arange(n)
binom_pdf = binom.pmf(x, n, p)

if t_obs==t1:
    exact_p = 2*sum(binom_pdf[:t_obs+1])
    t_cor = t_obs + 0.5
elif t_obs==t2:
    exact_p = 2*sum(binom_pdf[t_obs:])
    t_cor = t_obs - 0.5
print('exact p-value: ', exact_p)


# approximate p-value using normal approximation
z = abs(t_cor-n*p)/np.sqrt(n*p*(1-p))
approx_p = 2*(1 - norm.cdf(z))
print('approximate p-value', approx_p)
```

Normal Approximation of Binomial Distribution

- Output:

```
16 32 16
exact p-value:  0.01534667783263009
approximate p-value 0.01620954140922537
```

- H_0 (median exam completion time $= 20$ min) can be rejected with significance level of $\alpha = 0.05$.

Single Tailed Sign Test

- Sometimes it is needed to test

$$\mathrm{H}_0 : \eta = m \text{ (implicitly } \eta \leq m)$$
$$\mathrm{H}_1 : \eta > m$$

  or

$$\mathrm{H}_0 : \eta = m \text{ (implicitly } \eta \geq m)$$
$$\mathrm{H}_1 : \eta < m$$

- To test these hypothesis, single tailed sign test is used.

Single Tailed Sign Test

- In order to test

$$H_0 : \eta = m \text{ (implicitly } \eta \leq m)$$
$$H_1 : \eta > m$$

- Alternatively,

$$H_0 : p = 1/2 \text{ (implicitly } p \geq 1/2)$$
$$H_1 : p < 1/2$$

  where $p = P(X < m)$.
- Use T=# of observations **below** m.
- p-value is

$$p = P(T \leq T_{observed})$$

- Note that the multiplication by 2 is removed.

Single Tailed Sign Test

- In order to test

$$\mathrm{H}_0 : \eta = m \text{ (implicitly } \eta \geq m)$$
$$\mathrm{H}_1 : \eta < m$$

- Use T=# of observations **above** m.
- p-value is

$$p = P(T \geq T_{observed})$$

- Note that the multiplication is removed.

Two Sample Sign Test

- If two samples have paired observations, two-sample sign test can be used to test:

$$\begin{aligned} \mathsf{H}_0 &: \eta_1 = \eta_2 \\ \mathsf{H}_1 &: \eta_1 \neq \eta_2 \end{aligned}$$

alternatively

$$\begin{aligned} \mathsf{H}_0 &: \eta_1 - \eta_2 = 0 \\ \mathsf{H}_1 &: \eta_1 - \eta_2 \neq 0 \end{aligned}$$

where η_1 and η_2 are the medians of samples with paired observations.

Two Sample Sign Test – Example

- Assume that you want to analyze if feature selection really improves accuracy.
- Let:
  X: # of correctly classified observations **without** feature selection
  Y: # of correctly classified observations **with** feature selection
- Let η_X denote the median of X and η_Y denote the median of Y.
- Test

$$\mathrm{H}_0 : \eta_X = \eta_Y$$

Two Sample Sign Test – Example

- Number of correctly classified observations with and without feature selection.

Classifier	no feat. sel. (X)	feat. sel. (Y)	difference	sign
C1	314	350	-36	-
C2	365	365	0	NA
C3	465	415	50	+
...	...	...	...	...
C20	342	349	-7	-

- Assume, 14 negatives (feature sel worked well) and 3 positives (no feature sel worked well) are obtained.
- 3 ties (equal values) are ignored.

Two Sample Sign Test – Example

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm,binom

n=20
p=.5

t1 = 14 # negatives
t2 = 3 # positives
if t1>t2:
  t1,t2 = t2,t1

# find more extreme observation
t_obs  = t1 if abs(t1-n*p) > abs(t2-n*p) else t2

# exact p-value
x = np.arange(n)
binom_pdf = binom.pmf(x, n, p)

if t_obs==t1:
    exact_p = 2*sum(binom_pdf[:t_obs+1])
    t_cor = t_obs + 0.5
elif t_obs==t2:
    exact_p = 2*sum(binom_pdf[t_obs:])
    t_cor = t_obs - 0.5
print('exact p-value: ', exact_p)

# approximate p-value using normal approximation
z = abs(t_cor-n*p)/np.sqrt(n*p*(1-p))
approx_p = 2*(1 - norm.cdf(z))
print('approximate p-value', approx_p)
```

Two Sample Sign Test – Example

Code Output:

```
exact p-value:  0.0025768280029296844
approximate p-value 0.0036504344044419046
```

- Reject H_0 for significance level of $\alpha = 0.05$.
- There is statistically significant difference between the median number of correctly classified observations with and without feature selection.

Disadvantages of Sign Test

- **Advantage** of sign test: No assumption about distribution
- **Disadvantage** of sign test: It only considers the sign of difference
- Does not consider or weight the amount of difference.
- If observations come from a symmetric distr around zero, amounts can be used
- Difference amounts are sensitive to outliers and it can dominate the observed statistic
- Ranks of the differences can be used for inclusion of difference values $\implies$ Wilcoxon signed rank test

Wilcoxon Signed Rank Test

- Developed in 1945 by Frank Wilcoxon.
- Makes two assumption about about the underlying distribution of the data:
  - Distribution is continuous
  - Distribution is symmetric around zero
- Wilcoxon signed rank test can be used:
  - To test hypothesis about median value of a population $\rightarrow$ One-sample signed rank test
  - To test if two populations have same median when observations are paired $\rightarrow$ Two-sample signed rank test

One Sample Wilcoxon Signed Rank Test

- Let η denote the median of a population such that

$$P(X < \eta) = P(X > \eta) = 0.5$$

- Hypothesis are:

(two-tailed)	(one-tailed)	(one-tailed)
$H_0 : \eta = m$	$H_0 : \eta \leq m$	$H_0 : \eta \geq m$
$H_1 : \eta \neq m$	$H_1 : \eta > m$	$H_1 : \eta < m$

One Sample Wilcoxon Signed Rank Test

- Take difference of observations from m.

$$d_i = x_i - m$$

- Sort absolute differences $|d_i|$ in ascending order.
- Rank absolute differences from 1 to N (sample size).
- Define

$$u_i = \begin{cases} i & \text{if } d_i < 0 \\ 0 & \text{if } d_i > 0 \end{cases}$$

- If $|d_i|$ are same for multiple observations, divide the sum of their total rank – later.
- Add the ranks of negative and positive differences.

$$W^- = \sum_{i=1}^{n} u_i \qquad\qquad W^+ = \sum_{i=1}^{n} i - u_i$$

- For validation:

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

One Sample Wilcoxon Signed Rank Test – Example

- Assume that you hypothesize the median of a course midterm as 60.

$$\begin{aligned} H_0 &: \eta = 60 \\ H_1 &: \eta \neq 60 \end{aligned}$$

- You asked 10 of your friends about their grades:
  G = [35, 87, 50, 55, 67, 75, 80, 62, 43, 49]

Grade	$d_i = g_i - \eta$
35	-25
87	27
50	-10
...	...
63	2
43	-17
49	-11

One Sample Wilcoxon Signed Rank Test – Example

Grade	$d_i = g_i - \eta$	W^-	W^+
63	2		1
55	-5	2	
67	7		3
50	-10	4	
49	-11	5	
75	15		6
43	-17	7	
80	20		8
35	-25	9	
87	27		10
	Total	27	28

One Sample Wilcoxon Signed Rank Test

- What happens when there are observations with $d_i = 0$?
  - Observations are ignored.
  - Sample size is updated by reducing the number of these observations.
- What happens when there are observations with equal $|d_i|$?
  - Their ranks are added and divided equally between the observations.

One Sample Wilcoxon Signed Rank Test – Example

Grade	$d_i = g_i - \eta$	W^-	W^+
63	2		1
55	-5	2	
67	7		3
50	-10	4	
49	-11	5	
45	**-15**	**6.5**	
75	**15**		**6.5**
80	20		8
60	**0**	-	-
87	27		9
	Total	17.5	27.5

- Sample size is updated due to ignored observation $n \leftarrow 9$

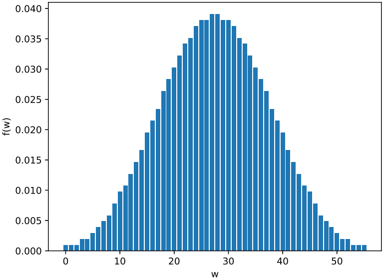
One Sample Wilcoxon Signed Rank Test

- You can use W^+ or W^-.
- W^+ and $W^- \in [0, \frac{n(n+1)}{2}]$
- Lets use $W = \min(W^-, W^+)$.
- $W \approx \frac{n(n+1)}{4}$, if H_0 is retained. Otherwise it will be either too small or too large.
- What is the distribution of W? Let $n = 3$:

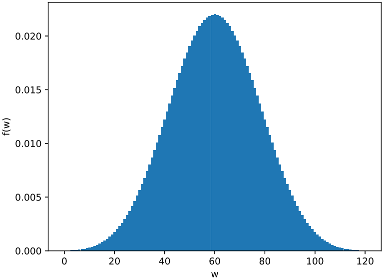
Rank	1	2	3	W^-	W^+	$W = \min(W^-, W^+)$
Sign	$-$	$-$	$-$	6	0	0
	$+$	$-$	$-$	5	1	1
	$-$	$+$	$-$	4	2	2
	$+$	$+$	$-$	3	3	3
	$-$	$-$	$+$	3	3	3
	$+$	$-$	$+$	2	4	2
	$-$	$+$	$+$	1	5	1
	$+$	$+$	$+$	0	6	0

One Sample Wilcoxon Signed Rank Test

- Exact pdf of W



$n = 10$



$n = 15$

One Sample Wilcoxon Signed Rank Test – W Table

- Critical values of W can be found in table.
- If $W_{obs} \leq W_{critical}$ reject H_0

TABLE • IX Critical Values for the Wilcoxon Signed-Rank Test

w^*_α

n^* \ α	0.10 0.05	0.05 0.025	0.02 0.01	0.01 0.005	Two-sided tests One-sided tests
4					
5	0				
6	2	0			
7	3	2	0		
8	5	3	1	0	
9	8	5	3	1	
10	10	8	5	3	
11	13	10	7	5	
12	17	13	9	7	
13	21	17	12	9	
14	25	21	15	12	
15	30	25	19	15	
16	35	29	23	19	
17	41	34	27	23	
18	47	40	32	27	
19	53	46	37	32	
20	60	52	43	37	
21	67	58	49	42	
22	75	65	55	48	
23	83	73	62	54	
24	91	81	69	61	
25	100	89	76	68	

One Sample Wilcoxon Signed Rank Test – Example

- In the revised example, $W^- = 17.5$ and $W^+ = 27.5$. Hence $W = min(W^-, W^+) = 17.5$.
- For $n = 9$ and $\alpha = 0.05$ (two-tailed), $W_{critical} = 5$ from this table.
- As $W \nless W_{critical}$, do not reject H_0 .

One Sample Wilcoxon Signed Rank Test

- As N increases,

$$W \sim \mathcal{N}\left(\frac{N(N+1)}{4}, \frac{N(N+1)(2N+1)}{24}\right)$$

- Probability of each difference is equally likely to be positive or negative (remember symmetry around zero assumption!):

$$P(sgn(d_i) = +1) = P(sgn(d_i) = +1) = 0.5$$

- Remember $W = \sum_{i=1}^{n} u_i$ where

$$u_i = \begin{cases} i & \text{if } d_i < 0 \\ 0 & \text{if } d_i > 0 \end{cases}$$

- Hence

$$E(W) = E\left(\sum_{i=1}^{n} u_i\right) = \sum_{i=1}^{n} E(u_i)$$

One Sample Wilcoxon Signed Rank Test

$$E(u_i) = 0 \times P(sgn(d_i) = +1) + i \times P(sgn(d_i) = -1) = \frac{i}{2}$$

$$E(W) = \sum_{i=1}^{n} E(u_i) = \sum_{i=1}^{n} \frac{i}{2} = \frac{n(n+1)}{4}$$

- Furthermore,

$$\sigma_W^2 = Var\left(\sum_{i=1}^{n} Var(u_i) u_i\right) = \sum_{i=1}^{n} Var(u_i)$$

as u_i are independent.

$$Var(u_i) = E(u_i^2) - E^2(u_i) = \frac{i^2}{2} - \frac{i^2}{4} = \frac{i^2}{4}$$

$$E(u_i^2) = 0^2 \times \frac{1}{2} + i^2 \times \frac{1}{2} = \frac{i^2}{2}$$

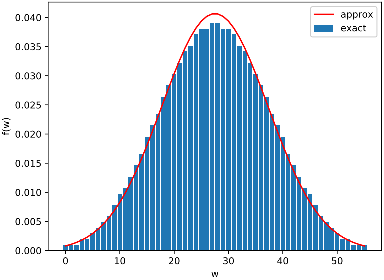
One Sample Wilcoxon Signed Rank Test

$$\sigma_W^2 = \sum_{i=1}^{n} \frac{i^2}{4} = \frac{n(n+1)(2n+1)}{24}$$

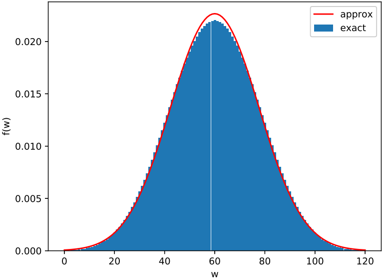
- For $n = 10$,
$$W \sim \mathcal{N}(27.5, 96.25)$$
- For $n = 15$,
$$W \sim \mathcal{N}(60, 310)$$

One Sample Wilcoxon Signed Rank Test

- Exact and approximate pdf of W



$n = 10$



$n = 15$

One Sample Wilcoxon Signed Rank Test – Example

- In revised example, $n = 9$,

$$W \sim \mathcal{N}(22.5, 71.25)$$

- For $\alpha = 0.05$ (two-tailed),

$$W_{critical} = 22.5 - 1.96\sqrt{71.25} \approx 5.96$$

(remember exact $W_{critical} = 5$)

- As $W > W_{critical}$, do not reject H_0 .

One Sample Wilcoxon Signed Rank Test – p-value

- Define $W_{min} = \min(W^-, W^+)$ and $W_{max} = \max(W^-, W^+)$.
- As $W^-, W^+ = n(n+1)/2$, W_{min} and W_{max} are equal distance from the mean value $(n(n+1)/4)$:

$$P(W \leq W_{min}) = P(W \geq W_{max}))$$

- p-value can be found as:

$$p = 2P(W \leq W_{min}) = 2P(W \geq W_{max})$$

- Continuity correction should be applied if normal approximation will be used.

One Sample Wilcoxon Signed Rank Test – Python Implementation

```
# Wilcoxon signed rank test
# one sample
# scipy.stats.wilcoxon(x, y=None, zero_method='wilcox',
    correction=False)
from scipy.stats import wilcoxon
import numpy as np

# sample
x = np.array([63, 55, 67, 50, 49, 45, 60, 75, 80, 87])

# hypothesized median
m = 60

d = x - m

# wilcox will ignore zero differences
w,p = wilcoxon(d,zero_method='wilcox',correction=False)
print('W stat: %d p-value %.3f'%(w,p))
```

One Sample Wilcoxon Signed Rank Test – Python Implementation

- Exact p-value 0.5703125
- Approximate p-value: $W \sim \mathcal{N}(22.5, 71.25)$

$$z = \frac{17.5 - 22.5}{\sqrt{71.25}} = 0.592$$

- From z-table

$$p = 2(1 - 0.7224) = 0.552$$

- Apply continuity correction

$$z = \frac{18 - 22.5}{\sqrt{71.25}} = 0.533$$

- From z-table

$$p = 2(1 - 0.7019) = 0.596$$

One Sample One Tailed Wilcoxon Signed Rank Test

- Test
$$\begin{aligned} \mathrm{H}_0 &: \eta \leq m \\ \mathrm{H}_1 &: \eta > m \end{aligned}$$
- Test
$$\begin{aligned} \mathrm{H}_0 &: \eta \geq m \\ \mathrm{H}_1 &: \eta < m \end{aligned}$$
- p-value is
$$p = P(W \leq W_{min}) = P(W \geq W_{max})$$
- Note that multiplication with 2 is removed.

Two Paired Sample Wilcoxon Signed Rank Test

- Wilcoxon signed rank test can also be used to compare the median of two populations
- Consider the following hypothesis
  $H_0 : \eta_1 = \eta_2$
  $H_1 : \eta_1 \neq \eta_2$
  equivalently
  $H_0 : \eta_1 - \eta_2 = 0$
  $H_1 : \eta_1 - \eta_2 \neq 0$
  where η_1 and η_2 are the medians of two populations
- Let x_i and y_i be the paired observations of sample 1 and sample 2

$$d_i = x_i - y_i$$

- Compute W^- and W^+, and $W = \min(W^-, W^+)$
- Use exact, approx distribution or table look-up to find critical value for a given significance level

Wilcoxon Rank Sum Test

- Equivalent to Mann-Whitney U test
- Consider comparing distributions of two independent populations
  H_0 : Population 1 and 2 has same distribution
  H_1 : Population 1 and 2 has different distribution
- Merge samples ($N = N_1 + N_2$) and find ranks in increasing order
- Let r_{1i} is the rank of observation i in sample 1
- Compute following statistic

$$W = \sum_{i=1}^{N_1} r_{1i}$$

  where

$$\frac{N(N+1)}{2} = \sum_{i=1}^{N_1} r_{1i} + \sum_{i=1}^{N_2} r_{2i}$$

- If H_0 is correct

$$\sum_{i=1}^{N_1} r_{1i} \approx \sum_{i=1}^{N_2} r_{2i} \approx \frac{N(N+1)}{4}$$

Wilcoxon Rank Sum Test

- If $N_1, N_2 > 8$ then normal approximation is possible for W

$$W \sim \mathcal{N}(\mu_W, \sigma_W^2)$$

where

$$\mu_W = \frac{N_1(N_1 + N_2 + 1)}{2}$$
$$\sigma_W^2 = \frac{N_1 N_2(N_1 + N_2 + 1)}{12}$$

Wilcoxon Rank Sum Test

TABLE • X Critical Values for the Wilcoxon Rank-Sum Test

$w_{0.05}$

n_2 \ n_1^*	4	5	6	7	8	9	10	11	12	13	14	15
4	10											
5	11	17										
6	12	18	26									
7	13	20	27	36								
8	14	21	29	38	49							
9	15	22	31	40	51	63						
10	15	23	32	42	53	65	78					
11	16	24	34	44	55	68	81	96				
12	17	26	35	46	58	71	85	99	115			
13	18	27	37	48	60	73	88	103	119	137		
14	19	28	38	50	63	76	91	106	123	141	160	
15	20	29	40	52	65	79	94	110	127	145	164	185
16	21	31	42	54	67	82	97	114	131	150	169	
17	21	32	43	56	70	84	100	117	135	154		
18	22	33	45	58	72	87	103	121	139			
19	23	34	46	60	74	90	107	124				
20	24	35	48	62	77	93	110					
21	25	37	50	64	79	95						
22	26	38	51	66	82							
23	27	39	53	68								
24	28	40	55									
25	28	42										
26	29											
27												
28												

Runs Test

- Also called Wald–Wolfowitz runs test
- Checks randomness of binary form data
  H_0 : Data is random
  H_1 : Data is not random
- Consider following data

$$x = [\underbrace{+,+,+,+}_{run\ 1},\underbrace{-,-,-}_{run\ 2},\underbrace{+,+,+}_{run\ 3},\underbrace{-,-}_{run\ 4},\underbrace{+,+,+,+,+,+}_{run\ 5},\underbrace{-,-,-,-}_{run\ 6}]$$

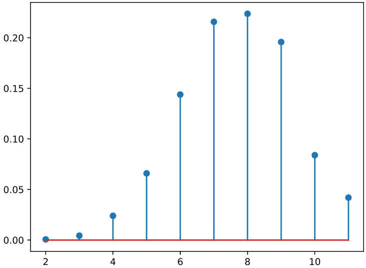
- Data has 3 '$+$' and 3 '$-$' runs
- Let R be the number of runs in the data
- If R is too small or too large reject H_0
- Let N_1 be number of negative observations, N_2 is the number of positive observations
- For the given data $N_1 = 9$ and $N_2 = 13$, and $N = N_1 + N_2$

Runs Test

- For a sample of size N, let $N_1 \neq 0$ and $N_2 \neq 0$, then

$$R_{min} = 2$$
$$R_{max} = 2\min(N_1, N_2) + 1$$

- If N_1 or N_2 is zero $\implies R = 1$
- For $N = 10$



Runs Test - Exact PDF

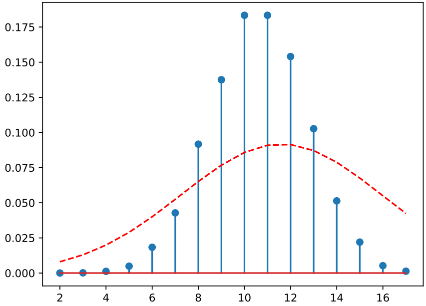
```python
import numpy as np
import matplotlib.pyplot as plt

N = 15
n1 = 5
n2 = N-n1

def runs(data):
    # count number of transitions and add 1
    r = 1
    for i in range(1,len(data)):
        if data[i] != data[i-1]:
            r += 1
    return r

t = np.arange(2,2*min(n1,n2)+2)
run_pdf = np.zeros(len(t))
for i in range(2**N):
    x = [int(xi) for xi in bin(i)[2:].zfill(N)]
    if sum(x) != n1:
        continue
    r = runs(x)
    print(x,r)
    run_pdf[r-2] += 1 # run in [1,N] not in [0, N-1]

# normalize for pdf
run_pdf /= sum(run_pdf)

plt.stem(t,run_pdf)
plt.show()
```

Runs Test

- When N is large (> 20), pdf of R can be approximated by normal distr
$$R \sim \mathcal{N}(\mu_R, \sigma_R^2)$$
where
$$\mu_R = \frac{2N_1 N_2}{N} + 1$$
$$\sigma_R^2 = \frac{2N_1 N_2(2N_1 N_2 - N)}{N^2(N-1)}$$
- Note both μ_R and σ_R^2 are symmetric with respect to N_1 and N_2

Runs Test - Normal Approximation

- For N=20 and $N_1 = 8$
- Under approximation in the middle is due to density assigned to impossible R (ie. $R > 17$)
- Better approximation with increasing N_1 and N_2



Runs Test

- If data is not binary, subtract its median and use signs

$$
\begin{gathered}
- \text{ if } x_i - \eta < 0 \\
+ \text{ if } x_i - \eta > 0 \\
? \text{ if } x_i - \eta = 0
\end{gathered}
$$

- If data is random, signs will also be random
- What to do with equal values (ties)
  - Count as $-$
  - Count as $+$
  - Count as previous sign (Ignored)
- If signs are different on two sides of a tie, R will not be effected $\implies$ non-critical tie
- If signs are the same on two sides of a tie, R will not incremented by 2 (per each tie) $\implies$ critical tie
- Count ties as $-$ compute R_- and count ties as $+$ and compute R_+
- If there is big difference between R_- and R_+ $\implies$ Consider ignoring

Runs Test

- Let data be

$$x = [3, 4, 3, 2, 3, 5, 7]$$

- $\eta = 3$
- Count $+$

$$x - \eta = [t, +, t, -, t, +, +]$$
$$sgn(x - \eta) = [+, +, +, -, +, +, +]$$

- $R_+ = 3$

- Count $-$

$$x - \eta = [t, +, t, -, t, +, +]$$
$$sgn(x - \eta) = [-, +, -, -, -, +, +]$$

- $R_- = 4$

Runs Test

- Let data be
$$x = [3, 4, 3, 2, 3, 5, 7]$$
- $\eta = 3$
- Ignore
$$x - \eta = [t, +, t, -, t, +, +]$$
$$sgn(x - \eta) = [+, +, +, -, -, +, +]$$
- $R_i = 3$

Kruskal Wallis Test

- Consider comparison of distributions for $T \geq 3$ populations
  H_0 : All populations have the same distr
  H_1 : At least one population have different distr
- Let sample t (from population t) has size n_t
- Merge all samples ($N = \sum n_t$), sort in increasing order
- Let r_{ti} be the rank of observation i in treatment t
- Average rank of total data is,

$$\bar{\bar{r}} = \frac{1}{N} \sum_t \sum_i r_{ti} = \frac{N+1}{2}$$

- When H_0 is correct, ranks will be distributed evenly between treatments

$$\bar{r}_t = \frac{1}{n_t} \sum_i^{n_t} r_{ti} \approx \frac{N+1}{2}$$

when H_0 is correct

Kruskal-Wallis Test

- Under H_0 , it is expected $\overline{r_t}$ will be close to $\overline{\overline{r}} = (N+1)/2$
- Define Kruskal-Wallis statistics (if there are no ties)

$$K = \frac{12}{N(N+1)} \sum_{t}^{T} n_t(\overline{r}_t - \overline{\overline{r}})^2$$
$$= \frac{12}{N(N+1)} \Big(\sum_{t}^{T} \sum_{i} \overline{r_t} - 3(N+1) \Big)$$

and

$$K \sim \chi^2_{T-1}$$

Kruskal-Wallis Test

- When there are ties, ranks are averaged to ties
- Define Kruskal-Wallis statistics

$$K = \frac{1}{S^2}\Big(\sum_{t}^{K} \overline{r_t} - \frac{N(N+1)^2}{4}\Big)$$

where S^2 is the variance of ranks that is defined as

$$S^2 = \frac{1}{N-1}\Big(\sum_{t}^{T}\sum_{i}^{n_t} r_{ti} - \frac{N(N+1)^2}{4}\Big)$$

- If there are not many ties, both methods of computation will be similar

Kruskal-Wallis Test

- As K is χ^2 distributed, to test H_0 with significance level of α critical value $\chi^2_{\alpha, T-1}$ is calculated
- This is always one-sided test as small values of K is not an evidence against H_0

Comparison of Methods - Big Picture

Parametric (mean)	**Nonparametric (median)**
1 sample t test	Sign test, Wilcoxon signed rank test
2 sample independent t test	Mann-Whitney test
2 sample paired t test	Wilcoxon signed rank test
One way ANOVA	Kruskal-Wallis test

Parametric vs Nonparametric Tests

	Parametric	Nonparametric
Sample size	large	small
Assumption	normal distr	none
Hypothesis	distribution parameter	distribution
Outliers	less robust	more robust
Power	more test power	less test power
Data type	numeric	numeric or ordinal