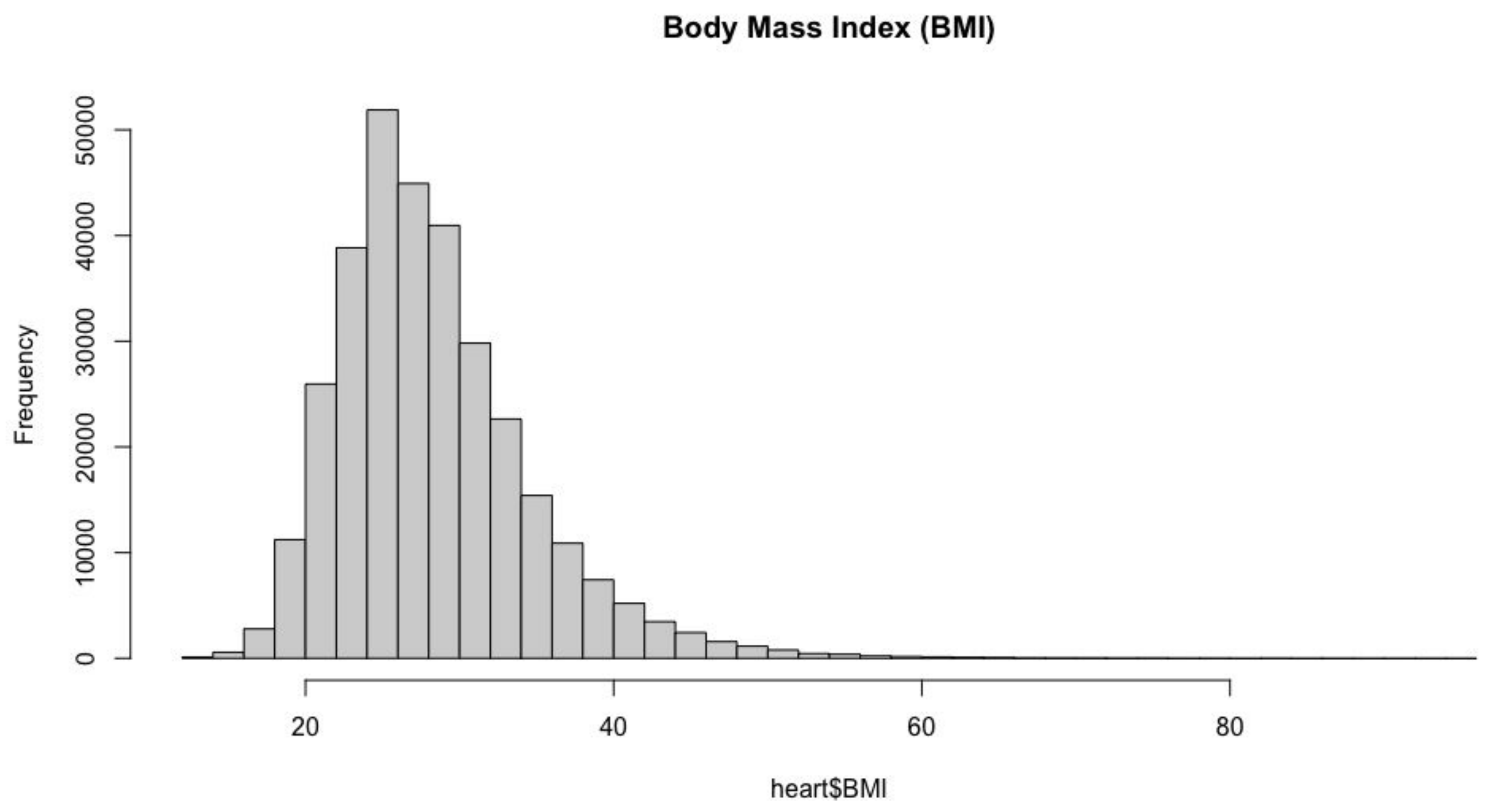
Nonparametric Tests

Charles Hwang

Background

- Nonparametric statistics is an entire subsection of statistics
  - STAT 351/488 - Nonparametric Statistical Methods[1] offered at Loyola every spring semester
- Nonparametric tests can be used when assumptions of traditional tests are violated
  - Most real-world data are not normally distributed


Body Mass Index (BMI)
Frequency
0
10000
20000
30000
40000
50000
20
40
60
80
heart$BMI

Concepts

- Median (*m*)
  - Not unduly influenced by outliers or distribution
- Ordinal data (Likert scale, rating, education level, etc.)
- Order statistics (rank)

Advantages

- Tests make fewer assumptions
    - Tests are often as or more robust than traditional tests
- Methods can be easier to use and interpret

Disadvantages

- Some tests have less power (larger *n* needed)
- Generally less commonly used and recognized
- Some tests don't have test statistics (only *p*-values)
  - Example: Fisher's exact test

(Incomplete) List of Methods[2]

- Analysis of similarities
- Anderson–Darling test
- Bootstrap methods
- Cochran's Q
- Cohen's κ
- Friedman two-way ANOVA by ranks
- Kaplan–Meier
- Kendall's τ
- Kendall's W
- Kolmogorov–Smirnov test
- Kruskal–Wallis one-way ANOVA by ranks
- Kuiper's test
- Log-rank test
- Mann–Whitney U (Wilcoxon rank sum) test
- McNemar's test
- Median test
- Pitman's permutation test
- Rank products
- Siegel–Tukey test
- Sign test
- Spearman's ρ
- Squared ranks test
- Tukey–Duckworth test
- Wald–Wolfowitz runs test
- Wilcoxon signed-rank test

(Incomplete) List of Methods[2]

- Analysis of similarities: tests similarities between groups against similarities within groups
- Anderson–Darling test: tests whether a sample is drawn from a given distribution
- Bootstrap methods: estimates the accuracy/sampling distribution of a statistic
- Cochran's Q: tests whether k treatments in randomized block designs with 0/1 outcomes have identical effects
- Cohen's κ: measures inter-rater agreement for categorical items
- Friedman two-way ANOVA by ranks: tests whether k treatments in randomized block designs have identical effects
- Kaplan–Meier: estimates the survival function from lifetime data, modeling censoring
- Kendall's τ: measures statistical dependence between two variables
- Kendall's W: a measure between 0 and 1 of inter-rater agreement
- Kolmogorov–Smirnov test: tests whether a sample is drawn from a given distribution, or whether two samples are drawn from the same distribution
- Kruskal–Wallis one-way ANOVA by ranks: tests whether > 2 independent samples are drawn from the same distribution
- Kuiper's test: tests whether a sample is drawn from a given distribution, sensitive to cyclic variations such as day of the week
- Log-rank test: compares survival distributions of two right-skewed, censored samples
- Mann–Whitney U (Wilcoxon rank sum) test: tests whether two samples are drawn from the same distribution, as compared to a given alternative hypothesis
- McNemar's test: tests whether, in 2×2 contingency tables with a dichotomous trait and matched pairs of subjects, row and column marginal frequencies are equal
- Median test: tests whether two samples are drawn from distributions with equal medians
- Pitman's permutation test: a statistical significance test that yields exact p values by examining all possible rearrangements of labels
- Rank products: detects differentially expressed genes in replicated microarray experiments
- Siegel–Tukey test: tests for differences in scale between two groups
- Sign test: tests whether matched pair samples are drawn from distributions with equal medians
- Spearman's ρ: measures statistical dependence between two variables using a monotonic function
- Squared ranks test: tests equality of variances in two or more samples
- Tukey–Duckworth test: tests equality of two distributions by using ranks
- Wald–Wolfowitz runs test: tests whether the elements of a sequence are mutually independent/random
- Wilcoxon signed-rank test: tests whether matched pair samples are drawn from populations with different mean ranks

(Incomplete) List of Methods

- Analysis of similarities
- Anderson–Darling test
- Bootstrap methods
- Cochran's *Q*
- Cohen's *κ*
- Friedman two-way ANOVA by ranks
- Kaplan–Meier
- **Kendall's *τ***
- Kendall's *W*
- **Kolmogorov–Smirnov test**
- Kruskal–Wallis one-way ANOVA by ranks
- Kuiper's test
- Log-rank test
- **Mann–Whitney *U* (Wilcoxon rank sum) test**
- **McNemar's test**
- Median test
- Pitman's permutation test
- Rank products
- Siegel–Tukey test
- **Sign test**
- **Spearman's *ρ***
- Squared ranks test
- Tukey–Duckworth test
- Wald–Wolfowitz runs test
- **Wilcoxon signed-rank test**

Kendall's τ

- Take all ${}_nC_2 = n(n\text{-}1) / 2$ pairs of points
- A pair is concordant if $x_i < x_j$ and $y_i < y_j$
- A pair is discordant if $x_i < x_j$ and $y_i > y_j$
- τ = (concordant - discordant) / ${}_nC_2$
- Ties (if $x_i = x_j$ or $y_i = y_j$)
  - τ_A (ignore)
  - τ_B (RStudio)
  - τ_C (Stuart–Kendall)

Kendall's τ

$$n_0 = n(n-1)/2$$
$$n_1 = \sum_i t_i(t_i - 1)/2$$
$$n_2 = \sum_j u_j(u_j - 1)/2$$
$$n_c = \text{Number of concordant pairs}$$
$$n_d = \text{Number of discordant pairs}$$
$$t_i = \text{Number of tied values in the } i^{\text{th}} \text{ group of ties for the first quantity}$$
$$u_j = \text{Number of tied values in the } j^{\text{th}} \text{ group of ties for the second quantity}$$
$$r = \text{Number of rows}$$
$$c = \text{Number of columns}$$
$$m = \min(r, c)$$

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}$$

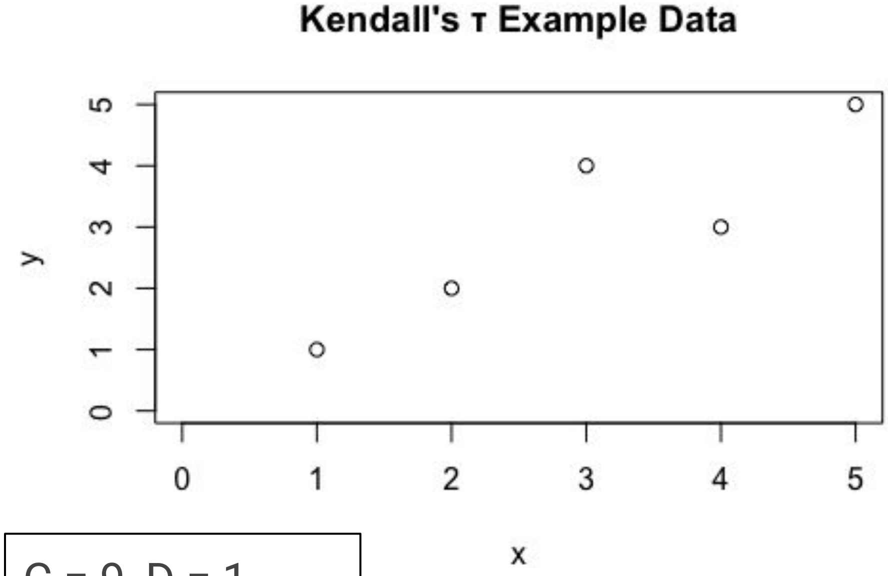
$$\tau_C = \frac{2(n_c - n_d)}{n^2 \frac{(m-1)}{m}}$$

Kendall's τ Example

```
cor(x,y,method="kendall")
```

j

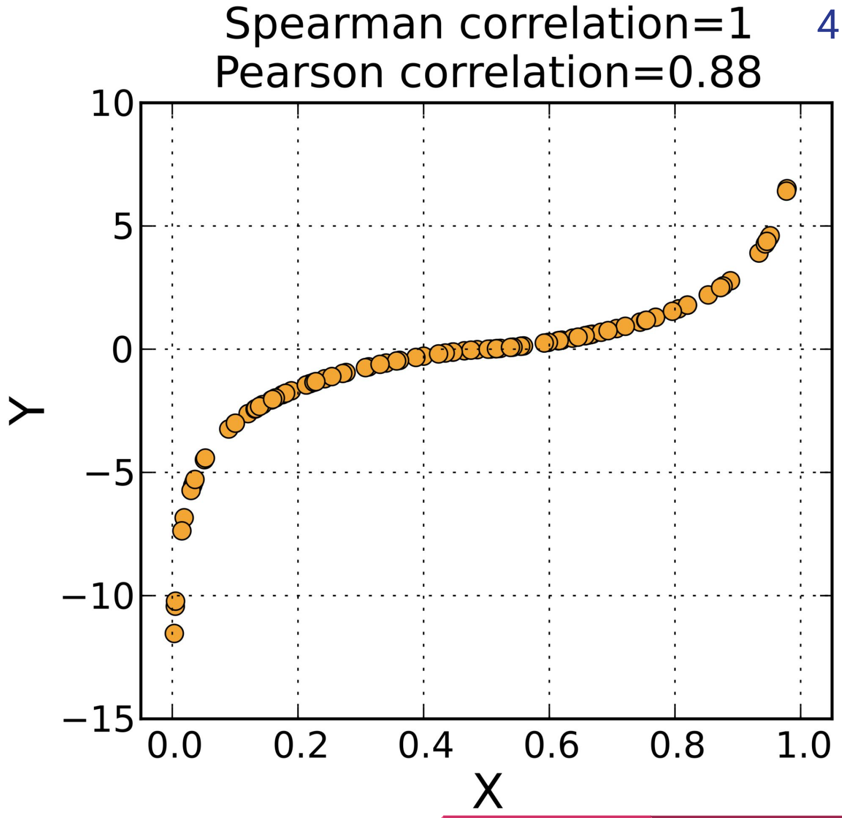
		(1,1)	(2,2)	(3,4)	(4,3)	(5,5)
	(1,1)		C	C	C	C
	(2,2)			C	C	C
i	(3,4)				D	C
	(4,3)					C
	(5,5)					



C = 9, D = 1
$\tau = (9 - 1) / 10 =$
$\tau = 0.8$

Spearman's *ρ*

- Calculates *r* on ranks
    - Ties: fractional (mean) ranks
- Not unduly affected by outliers or influential points
- Useful for data resembling a monotonic function (exponential, logistic, logarithmic, etc.)
- `cor(x,y,method="spearman")`



Sign test (paired or one-sample)

- Points where $x_i > y_i$ are ⊕
- Points where $x_i = y_i$ are ∅
- Points where $x_i < y_i$ are ⊖
- One-sample: replace y_i with η (estimated population median)
- $H_0: m_x - m_y = d$; $H_A: m_x - m_y \neq d$ or $H_0: m = \eta$; $H_A: m \neq \eta$
- $p = 2 * Binomcdf(\oplus; n - \varnothing, 0.5)$
- Disadvantage: Low power

Sign test (paired, H_0: $m_x - m_y = 0$)

x	0	1	2	3	3	4	5	6	7	7	8	9	10
y	1	2	4	3	4	5	6	7	8	8	9	9	9
Sgn	⊖	⊖	⊖	∅	⊖	⊖	⊖	⊖	⊖	⊖	⊖	∅	⊕

$p = 2 * Binomcdf(\oplus; n - \varnothing, 0.5) = 2 * Binomcdf(1; 13 - 2, 0.5) = 0.01171875$

```
binom.test(1,11,p=0.5,alternative="two.sided")
```

Sign test (one-sample, H_0: $m = 1$)

x	0	1	2	3	3	4	5	6	7	7	8	9	10
Sgn	⊖	∅	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕

$p = 2 * Binomcdf(\oplus; n - \varnothing, 0.5) \rightarrow 1 - 2 * Binomcdf(11; 13 - 1, 0.5) = 0.006347656$

```
binom.test(11,12,p=0.5,alternative="two.sided")
```

Wilcoxon signed-rank test

- Similar to sign test
- Points are sorted by absolute difference and ranked (ties ignored)
- $W^- + W^+ = W = n(n+1) / 2$

$$t = \frac{k + \frac{1}{2} - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{6}}}. \qquad \sigma^2 = \frac{n(n+1)(2n+1) - z(z+1)(2z+1) - c/2}{6},$$

Wilcoxon signed-rank test (paired, H_0: $m_x = m_y$)

x	100	101	102	103	103	104	105	106	107	107	108	109	110
y	101	100	99	98	97	96	95	94	93	93	92	92	92
Diff.	-1	1	3	5	6	8	10	12	14	14	16	17	18
R_i	1.5	1.5	3	4	5	6	7	8	9.5	9.5	11	12	13

$W^- = 1.5$

$W^+ = 1.5 + 3 + 4 + 5 + 6 + 7 + 8 + 9.5 + 9.5 + 11 + 12 + 13 = 89.5$

$W = 89.5$, $p = 0.002351027$

```
wilcox.test(x,y,paired=TRUE)
```

Wilcoxon signed-rank test (one-sample, H_0: m = 0)

x	0	1	2	3	3	4	5	6	7	7	8	9	10
Diff.	0	1	2	3	3	4	5	6	7	7	8	9	10
R_i	–	1	2	3.5	3.5	5	6	7	8.5	8.5	10	11	12

$W^- = 0$

$W^+ = 1 + 2 + 3.5 + 3.5 + 5 + 6 + 7 + 8.5 + 8.5 + 10 + 11 + 12 = 78$

$W = 78$, $p = 0.002506842$

```
wilcox.test(x,alternative="two.sided")
```

Mann–Whitney *U* (Wilcoxon rank sum) test

- Similar family of tests for non-paired two-sample
- H_0: $m_x - m_y = d$; H_A: $m_x - m_y \neq d$
- $U_1 = \Sigma_{i=1}^{n1} R_1 - n_2(n_2+1)/2$ $U_2 = \Sigma_{i=1}^{n2} R_2 - n_1(n_1+1)/2$

$$z = \frac{U - m_U}{\sigma_U}, \qquad m_U = \frac{n_1 n_2}{2}, \sigma_{\text{ties}} = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \frac{n_1 n_2 \sum_{k=1}^{K} ({t_k}^3 - t_k)}{12n(n-1)}},$$

Mann–Whitney *U* (Wilcoxon rank sum) test

x, y	0	1	2	3	3	4	5	6	7	7	8	9	10
R_x	1	2	3	4.5		6		8					
R_y					4.5		7		9.5	9.5	11	12	13

$U_1 = 1 + 2 + 3 + 4.5 + 6 + 8 - n_2(n_2+1)/2 = 24.5 - 6*7/2 = 3.5$

$U_2 = 4.5 + 7 + 9.5 + 9.5 + 11 + 12 + 13 - n_1(n_1+1)/2 = 66.5 - 5*6/2 = 51.5$

$W = 3.5, p = 0.01488064$

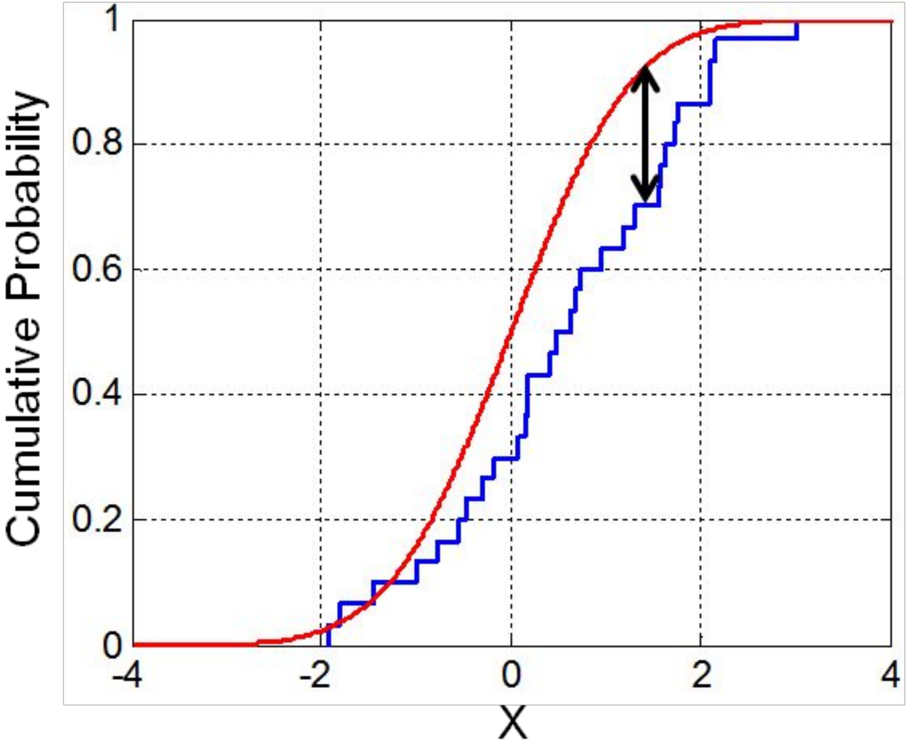
```
wilcox.test(x,y,alternative="two.sided")
```

Kolmogorov–Smirnov test

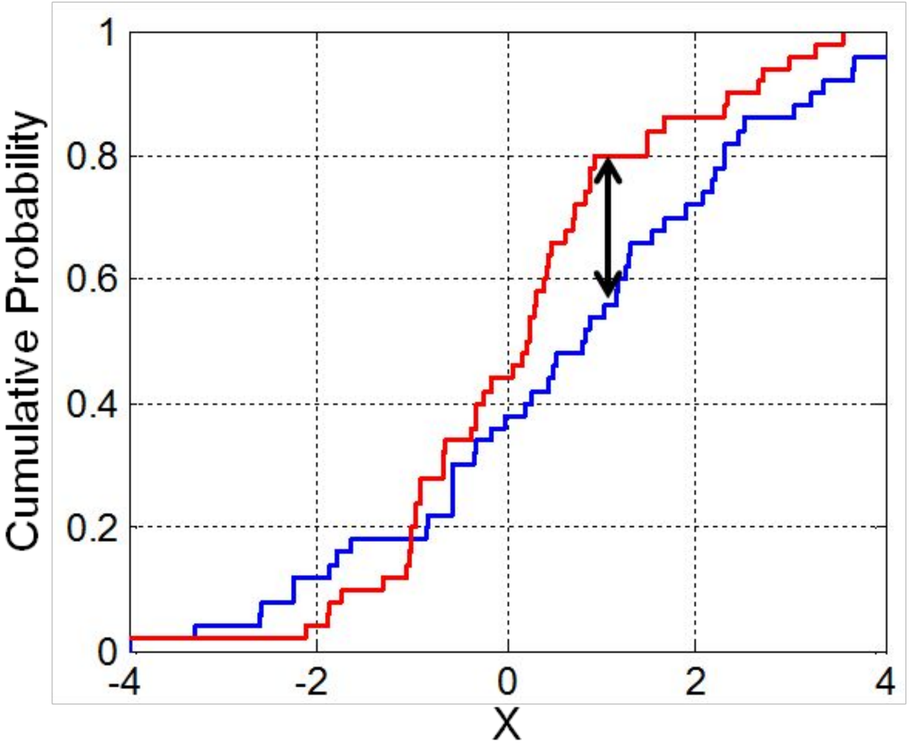
- One-sample (Kolmogorov) or two-sample (Smirnov)
- H_0: The data are from a certain/the same distribution
- H_A: The data are not from a certain/the same distribution
- Disadvantage: less power
  - Normal distribution: Shapiro–Wilk test
  - Anderson–Darling test gives more weight to tails

Kolmogorov–Smirnov test[7]

`ks.test(x,y)`



$$D_n = \sup_x |F_n(x) - F(x)|$$



$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|,$$

McNemar's (MACK-neh-MARS) test

- Tests for marginal homogeneity (paired data)
- $H_0: b = c;$ $\quad H_A: b \neq c$
- $\chi^2_{(1)} = (b - c)^2 / (b + c)$
- `mcnemar.test(x)`

2×2	Y	N	Tot
Y	*a*	*b*	*a+b*
N	*c*	*d*	*c+d*
Tot	*a+c*	*b+d*	*n*

Bonus: Fisher's exact test

- H_0: The two variables are independent
- H_A: The two variables are not independent
- `fisher.test(x,alternative="two.sided")`

2×2			Tot
	a	*b*	*a+b*
	c	*d*	*c+d*
Tot	*a+c*	*b+d*	*n*

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{\binom{a+b}{b}\binom{c+d}{d}}{\binom{n}{b+d}} = \frac{(a+b)!\;(c+d)!\;(a+c)!\;(b+d)!}{a!\;\;b!\;\;c!\;\;d!\;\;n!}$$

Conclusion

- Nonparametric statistics is an entire subsection of statistics
- Nonparametric tests can be used when assumptions of traditional tests are violated
    - Most real-world data are not normally distributed

References

1. https://www.luc.edu/media/lucedu/math/docs/Stat-Grid.pdf
    a. Hollander, Myles, Douglas A. Wolfe and Eric Chicken. *Nonparametric Statistical Methods*. 3rd ed., Wiley-Blackwell, 2014.
2. https://en.wikipedia.org/wiki/Nonparametric_statistics#Methods
3. https://en.wikipedia.org/wiki/Kendall_rank_correlation_coefficient#Accounting_for_ties
4. https://commons.wikimedia.org/wiki/File:Spearman_fig1.svg
5. https://en.wikipedia.org/wiki/Wilcoxon_signed-rank_test#Computing_the_null_distribution
6. https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test
7. https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test
    a. https://commons.wikimedia.org/wiki/File:KS_Example.png
    b. https://commons.wikimedia.org/wiki/File:KS2_Example.png
8. https://en.wikipedia.org/wiki/Fisher%27s_exact_test#Example

Questions?