SSSA
Small Sample Statistical Analysis

Day 2
Non-parametric Methods I

Dominik Duell (University of Essex)

September 27, 2016

1. Intro to non-parametric tests
2. Success probability: binomial test
3. Tests of differences between groups – paired from one-sample and independent from two-sample
    - Fisher sign test and Wilcoxon rank sign test
    - Comparing success probabilities: Fisher's exact test
    - Median test and Wilcoxon/Mann-Whitney rank sum test
    - Kolmogorov-Smirnov equal distribution test
    - Kruskal-Wallis test
4. Alternatives to correlation coefficients – bivariate relationships
    - Spearman's ρ
    - Kendall's τ
    - Goodman and Kruskal's gamma and Somers' D

Introduction

What are non-parametric tests

Tests that

- make fewer assumptions about the population – **distribution-free** tests or infinite-dimensional statistical models
- are more robust to outliers and heterogeneity of variance
- are robust even if you cannot reasonably characterize the distribution of the underlying population
- are applicable for interval and ordinal data – some for nominal data
- have test statistics that are distributed normally when N is large

When are non-parametric tests advantageous

- when assumptions of parametric tests/estimators are not met
  Example: t-test statistic does not have a t-distribution if underlying population not normal or sample size to small
-

Disadvantages of non-parametric tests

- No estimates of variance
- Mostly no confidence intervals
- Need more observations to draw a conclusion with same certainty i.e., less powerfull as parametric alternative when assumptions for parametric tests are met – differences are small, though and parametric alternatives perform vastly worse when assumptions are not met

Binomial test: Basics

- Observing outcomes of n independent repeated Bernoulli trials, what is the probability of success p?
- Assumptions:
  1. Dichotomous data: outcomes can be classified as either success or failure
  2. p remains constant for each trial
  3. n trials are independent
- $H_0 : p = 0$
- Other tests/statistics below relate to the basic binomial test of significance
- Under assumptions 1-3, it is a distribution-free test of H_0 because the probability distribution of B is determined without further assumptions on the distribution of the underlying population

Binomial test: Procedure

- To $H_0 : p = 0$ set the desired level of significance α and set B to number of observed successes
- Reject H_0 if $B \geq b_{\alpha_{lower}}$ and $B \leq c_\alpha$
- where b_α is the upper α_1 percentile point and b_{α_2} is the lower percentile point with $\alpha = \alpha_1 + \alpha_2$

Binomial test: Example

- Say $n = 8$ and we test $H_0 : p = .4$ vs $p > .4$
- From the table of the binomial distribution for $n = 8$ and $p = .4$ we get

b	0	1	2	3	4	5	6	
$Prob_{.4}(B \geq b)$	1	.9832	.8936	.6846	. 4059	.1737	.0498	.0

- Suppose we want $\alpha < .05$, b_αs that satisfy $Prob_{.4}(B \geq b_\alpha) = \alpha$ are 6, 7, 8 – for the upper-tail test, reject $H_0 : p = .4$ if 6 or more successes are observed

One- or two-sample difference tests

- investigate treatment effects on ...
  - observations from one sample (paired or **dependent** data)
  - observations from two samples (**independent** data)
- Wilcoxon sign rank and Wilcoxon/Mann-Whitney rank sum tests in detail, lots of other tests

Dependent samples

Sign test: Basics

- Alternative to paired t-test which assumes normality and equal variance across groups in underlying data
- Information taken from signs in difference between paired observations
- Assumptions:
  - paired observations $(X_1^1, X_1^2), \ldots, (X_N^1, X_N^2)$ are random sample and iid
  - paired observations are dependent
  - paired differences come from same continuous distribution
  - Use when direction of difference between two measurements on same unit can be determined

Sign test: Procedure

- Compute difference $D_i = X_i^1 - X_i^2$ between N pairs of matched observations
- Say, θ is median of distribution of D_i
- $H_0 : \theta = 0$ vs $H_A : \theta > 0$ or distribution of differences has median 0
- Test statistic D^+ is number of positive differences
- What is distribution of D^+? Think of $\theta_i = 1 if D^+ > 0$ and 0 otherwise as Bernoulli random variable $\rightarrow$ Distribution is binomial
- Under H_0 number of positive and negative differences should be equal or $H_0 : D^+ \sim binomial(N, 1/2)$
- Say number of positive D_i is n^+, then $B/2^N$ where $B = \binom{N}{n^+}$ gives the probability of getting exactly as many positive D_i
- To get obtain a p-Value, sum all binomial coefficients that are small than B and divided by 2^N

Sign test: Example

- Consider the `income`-variable in `gssData.dta`
- Question: Did income increase from '08 to '12?
- Note, it is an ordinal measured variable, taking a difference may not make sense – assume for this example that it makes sense

```
     +--------------------------+
     | income08   income12   D_i |
     |--------------------------|
  1. |        3         19   -16 |
  2. |        4          9    -5 |
  3. |       17         20    -3 |
  4. |       15         17    -2 |
  5. |       14         16    -2 |
     |--------------------------|
  6. |       20         21    -1 |
  7. |       20         21    -1 |
  8. |       22         22     0 |
  9. |       16         13     3 |
 10. |       21         17     4 |
     |--------------------------|
 11. |       19         15     4 |
 12. |       25         19     6 |
 13. |       14          6     8 |
 14. |       21         12     9 |
 15. |       11          1    10 |
     +--------------------------+
```

Sign test: Example

- Consider the `income`-variable in `gssData.dta`
- Note, it is an ordinal measured variable, taking a difference may not make sense – assume for this example that it makes sense looking at the categories not the income label
- Differences: -10 -9 -8 -6 -4 -4 -3 0 **1 1 2 2 3 5 16**
- How likely is it to observe 7 positive D_i when H_0 if $p = .5$ is true
- Binomial with $N = 15$, $p = .5$, and $x = 7$:$Prob(X \leq 7) = 0.50$

Sign test: Example

p=		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
$n=14$	$x=0$	0.8687	0.7536	0.6528	0.5647	0.4877	0.4205	0.3620	0.3112	0.2670	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
	1	0.9916	0.9690	0.9355	0.8941	0.8470	0.7963	0.7436	0.6900	0.6368	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
	2	0.9997	0.9975	0.9923	0.9833	0.9699	0.9522	0.9302	0.9042	0.8745	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
	3	1.0000	0.9999	0.9994	0.9981	0.9958	0.9920	0.9864	0.9786	0.9685	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
	4	1.0000	1.0000	1.0000	0.9998	0.9996	0.9990	0.9980	0.9965	0.9941	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
	5	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n=15$	$x=0$	0.8601	0.7386	0.6333	0.5421	0.4633	0.3953	0.3367	0.2863	0.2430	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.9904	0.9647	0.9270	0.8809	0.8290	0.7738	0.7168	0.6597	0.6035	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9996	0.9970	0.9906	0.9797	0.9638	0.9429	0.9171	0.8870	0.8531	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	1.0000	0.9998	0.9992	0.9976	0.9945	0.9896	0.9825	0.9727	0.9601	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986	0.9972	0.9950	0.9918	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Sign test: Example

```
Sign test

        sign |    observed    expected
-------------+------------------------
    positive |           7           7
    negative |           7           7
        zero |           1           1
-------------+------------------------
         all |          15          15

One-sided tests:
  Ho: median of income08 - income12 = 0 vs.
  Ha: median of income08 - income12 > 0
      Pr(#positive >= 7) =
         Binomial(n = 14, x >= 7, p = 0.5) =  0.6047

  Ho: median of income08 - income12 = 0 vs.
  Ha: median of income08 - income12 < 0
      Pr(#negative >= 7) =
         Binomial(n = 14, x >= 7, p = 0.5) =  0.6047

Two-sided test:
  Ho: median of income08 - income12 = 0 vs.
  Ha: median of income08 - income12 != 0
      Pr(#positive >= 7 or #negative >= 7) =
         min(1, 2*Binomial(n = 14, x >= 7, p = 0.5)) =  1.0000
```

Sign test: Example

$p=$		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
$n=14$	$x=0$	0.8687	0.7536	0.6528	0.5647	0.4877	0.4205	0.3620	0.3112	0.2670	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
	1	0.9916	0.9690	0.9355	0.8941	0.8470	0.7963	0.7436	0.6900	0.6368	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
	2	0.9997	0.9975	0.9923	0.9833	0.9699	0.9522	0.9302	0.9042	0.8745	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
	3	1.0000	0.9999	0.9994	0.9981	0.9958	0.9920	0.9864	0.9786	0.9685	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
	4	1.0000	1.0000	1.0000	0.9998	0.9996	0.9990	0.9980	0.9965	0.9941	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
	5	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n=15$	$x=0$	0.8601	0.7386	0.6333	0.5421	0.4633	0.3953	0.3367	0.2863	0.2430	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.9904	0.9647	0.9270	0.8809	0.8290	0.7738	0.7168	0.6597	0.6035	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9996	0.9970	0.9906	0.9797	0.9638	0.9429	0.9171	0.8870	0.8531	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	1.0000	0.9998	0.9992	0.9976	0.9945	0.9896	0.9825	0.9727	0.9601	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986	0.9972	0.9950	0.9918	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Sign test: Small sample issues

- Only needs sign of difference, not difference itself
- Less efficient than Wilcoxon sign-rank test (uses only sign not ordering) but rather robust to outliers
- How to deal with $D_i = 0$: Usually ignored but reduces effective sample size
- works for interval data but pay attention to ties - which correction for tied values?
- Generally, within-subject design may require fewer subjects

Wilcoxon sign-rank test: Basics

- Alternative to paired t-test which assumes normality and equal variance across groups in underlying data
- Information taken from signs in difference between paired observations (pre- vs post-treatment)
- When actual difference pre- vs post-treatment is greater than 0, tendency to larger proportion of positive differences

Wilcoxon sign-rank test: Basics

- Assumptions:
  - paired observations $(X_1^1, X_1^2), \dots, (X_N^1, X_N^2)$ are random sample and iid i.e., differences are mutually independent (while paired observations are dependent)
  - paired differences come from a continuous distribution
- Test of null hypothesis of zero shift in location (no treatment effect), $H_0 : \theta = 0$ – null hypothesis states that each of the distributions for the differences is symmetrically distributed about 0
- Use when direction of difference and magnitude between two measurements on same unit can be determined

Wilcoxon sign-rank test: Procedure

- Compute difference $D_i = X_i^1 - X_i^2$ between N pairs of matched observations
- Order absolute values of differences from smallest to largest
- Let S_1^2 denote the rank of $D_1, \ldots, S_N$ denote the rank of D_N in the joint ordering
- Assign average rank to ties
- Wilcoxon signed rank statistic, W^+, is sum of positive signed ranks
- Under $H_0 : \theta = 0$, W^+ is distributed according to the distribution derived by Wilcoxon (1954)
- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$

Wilcoxon sign-rank test: Distribution of W

- Based on permutations of all possible rankings
- Btw, related to Mann-Whitney U
- Where is the distribution coming from?

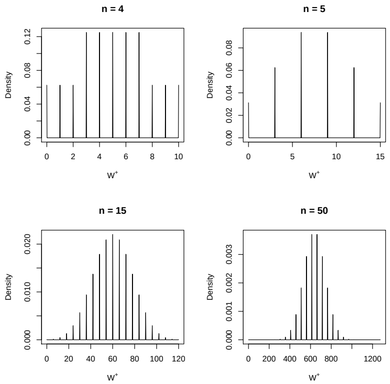
Wilcoxon sign-rank test: Example

	income08	income12	D_i	absD_i	rank
1.	22	22	0	0	1
2.	20	21	-1	1	2.5
3.	20	21	-1	1	2.5
4.	14	16	-2	2	4.5
5.	15	17	-2	2	4.5
6.	16	13	3	3	6.5
7.	17	20	-3	3	6.5
8.	21	17	4	4	8.5
9.	19	15	4	4	8.5
10.	4	9	-5	5	10
11.	25	19	6	6	11
12.	14	6	8	8	12
13.	21	12	9	9	13
14.	11	1	10	10	14
15.	3	19	-16	16	15

Wilcoxon sign-rank test: Distribution of W^+

- Sum of ranks of positive differences: 74.5
- (Sum of ranks of negative difference: 45.5)
- Lowest possible rank? 0 – no difference is positive
- Highest possible rank? $N(N+1)/2 = 15(16)/2 = 120$ – all differences are positive

Wilcoxon sign-rank test: Distribution of W^+



- What is the smallest significance level at which these data lead to rejection of H_0?
- For our example, we have $N = 14$ and $W^+ = 74.5$

Wilcoxon sign-rank test: Critical values of W^+ for $N = 14$

.117	$n = 14$	66	.213		93
.102		67	.196		94
.088		68	.179		95
.076		69	.163		96
.065		70	.148		97
.055		71	.134		98
.046		72	.121		99
.039		73	.108		100
.032		74	.097		101
.026		75	.086		102
.021		76	.077		103
.017		77	.068		104
.013		78	.059		105
.010		79	.052		106
.008		80	.045		107
.006		81	.039		108
.005		82	.034		109
.003		83	.029		110
.002		84	.025		111
.002		85	.021		112
.001		86	.018		113
.001		87	.015		114
<.0005		88	.012	$n = 16$	93
.207		89	.010		94
.188		90	.008		95
.170		91	.007		96
.153		92	.005		97
.137		93	.004		98
.122		94	.003		99
.108		95	.003		100
.095		96	.002		101
.084		97	.002		102
.073		98	.001		103
.064		99	.001		104
.055		100	.001		105
.047		101	<.0005		106

Source: Hollander/Wolfe, p.576

Wilcoxon sign-rank test: Example

- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$
- For $\alpha = .052$, reject if $W^+ \geq 80$ or $W^+ \leq \frac{15(16)}{2} - 80 = 120 - 80 = 40$ – so we do not reject
- What is the large sample approximation:
  - Standardize W^+: Under the null, $E(W^+) = \frac{n(n+1)}{4}$ and $var(W^+) = \frac{n(n+1)(2n+1)}{24}$
  - $W^* = \frac{W^+ - E(W^+)}{\sqrt{var(W^+)}}$
  - With $n \to \infty$, $W^+ \sim N(0,1)$
  - Adjustments for ties are needed

Wilcoxon sign-rank test: Example

```
Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |        7        73.5        59.5
    negative |        7        45.5        59.5
        zero |        1           1           1
-------------+---------------------------------
         all |       15         120         120

unadjusted variance      310.00
adjustment for ties       -0.50
adjustment for zeros      -0.25
                     ----------
adjusted variance        309.25

Ho: income08 = income12
             z =   0.796
    Prob > |z| =   0.4260
```

signrankex income08 = income12

```
Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |        7        66.5        52.5
    negative |        7        38.5        52.5
        zero |        1
-------------+---------------------------------
         all |       15         120         120

Ho: income08 = income12
              S =  14.000
    Prob >= |S| =   0.3976
```

Wilcoxon sign-rank test: Small sample issues

- Wilcoxon sign-rank test incorporates more information than sign-test but also needs more information
- How to deal with $D_i = 0$: Usually ignored but reduces effective sample size
- works for interval data but pay attention to ties - which correction for tied values?
- Generally, within-subject design may require fewer subjects

CI for Wilcoxon sign rank test: Basics

- Related to Wilcoxon sign-rank statistic and Hodges-Lehman location estimator
- Hodges-Lehman estimator for real treatment effect θ is $\hat{\theta} = median\{\frac{D_i+D_j}{2}, i \leq j = 1, \ldots, n\}$
- Then, $O^1 \leq \ldots O^M$ are the ordered values of the average differences with $M = \frac{n(n+1)}{2}$

CI for Wilcoxon sign rank test: Procedure

- Obtain upper $(\alpha/2)$th percentile point $w_{\alpha/2}$ of the null distribution of W^+
- $C_\alpha = \frac{n(n+1)}{2} + 1 - w_{\alpha/2}$
- CI for two-sided test of $H_0 : \theta = 0$ (zero location shift):
  - $\theta_{lb} = O^{C_\alpha}$
  - $\theta_{ub} = O^{M+1-C}$

Independent samples

Fishers exact test: Basics

- Say we observe the outcomes $O_{11}\,(O_{21})$ of $n_1\,(n_2)$ independent repeated Bernoulli trials with real success probability $p_1\,(p_2)$ in a sample from population 1 (2)

	Successes	Failures	Totals
Sample 1	O_{11}	O_{12}	$n_{1\cdot}$
Sample 2	O_{11}	O_{12}	$n_{2\cdot}$
Totals	$n_{\cdot 1}$	$n_{\cdot 2}$	n

- Assumption
  - trials from sample 1 are independent of those from sample 2
- $H_0 : p_1 = p_2 = p$

Fishers exact test: Procedure

- $Prob(O_{11} = x | n_{1\cdot}, n_{2\cdot}, n_{1\cdot 1}, n_{\cdot 2}) = \frac{n_{\cdot 1}! n_{\cdot 2}! n_{1\cdot}! n_{1\cdot}!}{n! x! O_{12}! O_{21}! O_{22}!}$
- Fisher's exact test rejects $H_0 : p_1 = p_2$ if $O_{11} \geq q_\alpha$

  where q_α is chosen from the conditional distribution described above so that $Prob(O_{11} = x | n_{1\cdot}, n_{2\cdot}, n_{1\cdot 1}, n_{\cdot 2}) = \alpha$ where α is our desired level of significance

Fishers exact test: Example

```
           |        varBi
       cat |         0          1 |     Total
-----------+----------------------+----------
         0 |         5          4 |         9
         1 |         4          2 |         6
-----------+----------------------+----------
     Total |         9          6 |        15
```

- $H_0 : p_1 > p_2$
- What are the probabilities of the tables that would give us a value as larger as or larger than the observed value of $O_{11} = 5$?

Fishers exact test: Example

3	6	4	5	5	4	6	3	7	2	8	1	9	0
6	0	5	1	4	2	3	3	2	4	1	5	0	6
.017		.151		.378		.336		.108		.011		.000	

- $H_0 : p_1 < p_2$
- What are the probabilities of the tables that would give us a value as small as or small than the observed value of $O_{11} = 5$? – it is $.017 + .151 + .378 = .546$

Fishers exact test: Small sample issues

- Appropriate when expected frequency in any of the cells is below 5 – otherwise χ^2-test
- Also small sample test of independence

Fishers exact test: Example

```
.                 tab cat varBi, exact;

           |         varBi
       cat |         0          1 |     Total
-----------+----------------------+----------
         0 |         5          4 |         9
         1 |         4          2 |         6
-----------+----------------------+----------
     Total |         9          6 |        15

           Fisher's exact =                 1.000
   1-sided Fisher's exact =                 0.545
```

Median test: Basics

- Assumptions:
  - observations $X_1^1, \ldots, X_n^1$ ($X_2^1, \ldots, X_m^2$) from population 1 (2) are random samples and iid
  - independence between the two samples
- Hypothesis: medians of the two populations are the same

Median test: Procedure

- Take grand median of combined sample of $N = n + m$ observations
- Classify each observation as below or above the grand median, drop those equal to the median
- Fill 2x2 contingency table
- Perform Fisher's exact test
- Easily extendable to k samples

Median test: Example

	$cat == 0$	$cat == 1$	
Below grand median	5	2	7
Above grand median	3	4	7
	8	6	14

Median test: Example

```
. median var, by(cat) exact medianties (drop)

Median test

   Greater |
  than the |         cat
    median |         0          1 |     Total
-----------+----------------------+----------
        no |         5          2 |         7
       yes |         3          4 |         7
-----------+----------------------+----------
     Total |         8          6 |        14

          Pearson chi2(1) =   1.1667   Pr = 0.280
           Fisher's exact =                 0.592
   1-sided Fisher's exact =                 0.296

   Continuity corrected:
          Pearson chi2(1) =   0.2917   Pr = 0.589
```

Median test: Small sample issues

- Valid only for interval and ordinal data
- With skewed distributions, median is a robust statistic!
- Measures how many observations are below/above median in group and not by how much do observations differ – less powerful test than parametric alternative

Alternative parametric t-test

Dangerous beast if hunting for low p-values in this case

```
. ttest var, by(cat)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |       9    6.777778    1.769948    5.309844     2.69627    10.85929
       1 |       6        10.5    1.979057     4.84768    5.412672    15.58733
---------+--------------------------------------------------------------------
combined |      15    8.266667    1.367886    5.297798    5.332844    11.20049
---------+--------------------------------------------------------------------
    diff |           -3.722222    2.707443               -9.571297    2.126853
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -1.3748
Ho: diff = 0                                     degrees of freedom =       13

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0962         Pr(|T| > |t|) = 0.1924          Pr(T > t) = 0.9038
```

- Disregard? – skewed distribution, unknown variance

Mann-Whitney/Wilcoxon: Basics

- Basic hypothesis, no treatment effect or sample from one population
- If treatment effect positive, values from one sample tend to be larger than values from other sample $\rightarrow$ ranks of values in one sample larger than in the other
- Mann Whitney U / Wilcoxon W statistic provide base for similar test

Mann-Whitney/Wilcoxon: Basics

- Assumptions:
  - observations $X_1^1, \ldots, X_n^1$ ($X_2^1, \ldots, X_m^2$) from population 1 (2) are random samples and iid
  - independence between the two samples
  - continous outcome variable
- Hypothesis: $H_0 : F(x) = G(x) \forall x$ where F (G) is the distribution function corresponding to population 1 (2) – comparison of distributions!
- Alternatively: $H_0 : E(X^1) - E(X^2) = 0$ – test of shift in location only when underlying distribution of similar shape – check out Fligner-Policello

Mann-Whitney/Wilcoxon: Procedure

- Order combined sample of $N = n + m$ observations from smallest to largest
- Let S_1^2 denote the rank of $X_1^2, \ldots, S_m$ denote the rank of X_m^2 in the joint ordering
- Assign ties average rank
- Sum of ranks assigned to X^2-values is $W = \sum_j = 1^N S_j$
- Two-sided test: Reject H_0 if $W \geq w_{\alpha/2}$
- Get distribution of W from table – generated from all combinations of rank-orderings
- Mann Whitney U:
  - For each pair of X_i^1 and X_j^2 observe which is smaller and score one for that pair if X_i is smaller
  - Sum of scores is U
  - Without ties, $W = U + n(n+1)/2$

Mann-Whitney/Wilcoxon: Example

- Consider the variable `var` in the `fakeData.dta`
- We ask, is there a difference in distribution of `var` across the groups defined by `cat` – between-subject treatment effect

```
. ranksum var, by(cat)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

         cat |      obs    rank sum    expected
-------------+---------------------------------
           0 |        9        59.5          72
           1 |        6        60.5          48
-------------+---------------------------------
    combined |       15         120         120

unadjusted variance       72.00
adjustment for ties       -0.51
                     ----------
adjusted variance         71.49

Ho: var(cat==0) = var(cat==1)
             z =  -1.478
    Prob > |z| =   0.1393
```

Mann-Whitney/Wilcoxon: Example

- Is normal approximation appropriate let's look at exact probabilities

```
. ranksumex var, by(cat)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

         cat |      obs    rank sum    expected
-------------+---------------------------------
           0 |        9        59.5          72
           1 |        6        60.5          48
-------------+---------------------------------
    combined |       15         120         120

Exact statistics
Ho: var(cat==0) = var(cat==1)
    Prob <=      35.5 = 0.0769
    Prob >=      60.5 = 0.0771
    Two-sided p-value = 0.1540
```

- Note: exact distribution too conservative with many ties.

Mann-Whitney/Wilcoxon: Example

- And we do have many ties

```
     +------------------+
     | rank   var   cat |
     |------------------|
  1. |    1     1     0 |
  2. |  2.5     2     0 |
  3. |  2.5     2     0 |
  4. |  4.5     3     1 |
  5. |  4.5     3     0 |
     |------------------|
  6. |    6     5     0 |
  7. |    7     6     1 |
  8. |    8     9     0 |
  9. |    9    11     0 |
 10. |   10    12     1 |
     |------------------|
 11. | 11.5    13     1 |
 12. | 11.5    13     0 |
 13. |   13    14     1 |
 14. | 14.5    15     1 |
 15. | 14.5    15     0 |
     +------------------+
```

Mann-Whitney/Wilcoxon: Small sample issues

- Valid for any distribution of the sample – exact test only valid if few or no ties between the groups!
- Much less sensitive to outliers than two-sample t-test
- Wilcoxon only little less likely to detect location shift than t-test
- For joint sample sizes larger than 20, use normal approximation
- rank sum test only a test of equality in medians/means if distributions are of same shape but differ in location
- works for interval or ordinal data but pay attention to ties

Alternative parametric t-test

- Again, it's a dangerous beast if hunting for low p-values in this case

```
. ttest var, by(cat)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |       9    6.777778    1.769948    5.309844     2.69627    10.85929
       1 |       6        10.5    1.979057     4.84768    5.412672    15.58733
---------+--------------------------------------------------------------------
combined |      15    8.266667    1.367886    5.297798    5.332844    11.20049
---------+--------------------------------------------------------------------
    diff |           -3.722222    2.707443               -9.571297    2.126853
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -1.3748
Ho: diff = 0                                     degrees of freedom =       13

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0962         Pr(|T| > |t|) = 0.1924          Pr(T > t) = 0.9038
```

- Disregard? – skewed distribution, unknown variance

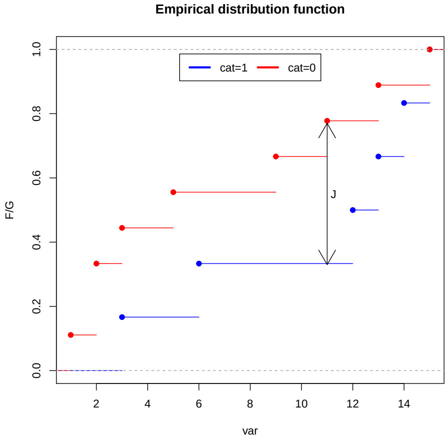
Kolmogorov-Smirnov: Basics

- Test of equality of distributions, not a directional test
- Assumptions:
  - observations $X_1^1, \ldots, X_n^1$ ($X_2^1, \ldots, X_m^2$) from population 1 (2) are random samples and iid
  - independence between the two samples
  - continous outcome variables
- Test of $H_0 : F(x) = G(x) \, \forall \, x$ vs $H_A : F(x) \neq G(x)$ for at least one x
- Transfers values of observations into a step function, makes it a distribution-free test

Kolmogorov-Smirnov: Procedure

- For the two samples X_1 and X_2 order the combined $N = n + m$ values denoted $Z_1, \ldots, Z_N$
- obtain the empirical distribution functions
- For every $i = 1, \ldots, N$, let
  $F(Z_i) = \frac{\#\text{ of sample } X_1\text{s} \leq Z_i}{m}$
  $G(Z_i) = \frac{\#\text{ of sample } X_1\text{s} \leq Z_i}{n}$
- This is the fraction of sample observations less than or equal to the value behind Z_i
- Then, $J = \frac{mn}{d} max_{i=1,\ldots,N}\{F(Z_i) - G(Z_i\}$ is the test statistic where d is the greatest common divisor of m and n
- Reject H_0 if $J \geq j_\alpha$

Kolmogorov-Smirnov: Example



Adjusted, largest distance between empirical distribution functions is J statistic

Kolmogorov-Smirnov: Example

	var	F	G	D_FG	maxD_FG	M	N	J
1.	1	.1111111	0	.1111111	.4444444	9	6	8
2.	2	.3333333	0	.3333333	.4444444	9	6	8
3.	2	.3333333	0	.3333333	.4444444	9	6	8
4.	3	.4444444	.1666667	.2777778	.4444444	9	6	8
5.	3	.4444444	.1666667	.2777778	.4444444	9	6	8
6.	5	.5555556	.1666667	.3888889	.4444444	9	6	8
7.	6	.5555556	.3333333	.2222222	.4444444	9	6	8
8.	9	.6666667	.3333333	.3333333	.4444444	9	6	8
9.	11	.7777778	.3333333	.4444444	.4444444	9	6	8
10.	12	.7777778	.5	.2777778	.4444444	9	6	8
11.	13	.8888889	.6666667	.2222222	.4444444	9	6	8
12.	13	.8888889	.6666667	.2222222	.4444444	9	6	8
13.	14	.8888889	.8333333	.0555556	.4444444	9	6	8
14.	15	1	1	0	.4444444	9	6	8
15.	15	1	1	0	.4444444	9	6	8

- We cannot reject H_0 at standard levels of significance
- Ties! For exact probabilities, each step should have been $1/15 = .067$

Kolmogorov-Smirnov: Example

31	.0559
35	.0280
36	.0140
40	.0060
45	.0010

$m = 6, n = 9$

x	$P_0\{J \geq x\}$
10	.1758
11	.0947
12	.0611
13	.0280
14	.0140
15	.0060
16	.0028
18	.0004

$m = 7, n = 9$

x	$P_0\{J \geq x\}$
35	.1267
36	.0979
38	.0787

Source: Hollander/Wolfe (1999), p.608

Kolmogorov-Smirnov: Small sample issues

- Ties require adjustment to how exact p-values are computed. Could derive the conditional null distribution by considering the $\binom{N}{\#\text{ of ties}}$ possible ways how our observations could be assigned – not implemented in R, Stata
- Exact values appropriate, Smirnovs (1933) approximations not good for samples smaller than 50
- Do not do the one-sample test for normality with Kolmogorov-Smirnov – even best (also non-parametric) alternative, Shapiro-Wilk test, has not enough power to reject normality

Kruskal-Wallis: Basics

- Test of the location of k populations
- Parametric alternative is the one-way ANOVA – builds on a measure of group differences but in ranks
- Extension of the Mann-Whitney U to more than two groups
- Population may be defined by confounding variables – moving into multi-variate analysis

Kruskal-Wallis: Basics

- Assumptions:
  - observations $X_1^1, \ldots, X_n^1$ $(X_2^1, \ldots, X_m^2)$ from population 1 (2) are random samples and iid
  - independence between the two samples
  - continous outcome variables
  - distribution of outcome variable has similar shape across groups
- Under these assumptions and H_0 the vector of ranks has a uniform distribution over the set of all $N!$ permutations of the vectors of integers $(1, 2, \ldots, N)$
- $H_0 : \theta_1 = \ldots = \theta_k$– Kruskal-Wallis tests against H_a of at least two treatment effects are not equal
- Applicable to ordinal and continuous scales

Kruskal-Wallis: Procedure

- Combine N observations from k samples and rank all X
- Let r_{ij} be the rank of observation X_{ij} then

  $R_j = \sum_{i=1} n_j$ and $R_{.j} = \frac{R_j}{n_j}$ for treatment j
- Then,

  $$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \left(R_{.j} - \frac{N+1}{2}\right)^2$$

  where n_j is the number of observations in treatment j – add appropriate correction of ties
- H is a constant $\times$ a weighted sum of squared differences between the observed average rank and the expected value under the null within a group
- Reject H_0 if $H \geq h_\alpha$

Kruskal-Wallis: Example

```
     +-------------------+
     | cat3   var   rank |
     |-------------------|
  1. |    3     1      1 |
  2. |    2     2    2.5 |
  3. |    3     2    2.5 |
  4. |    1     3    4.5 |
  5. |    1     3    4.5 |
     |-------------------|
  6. |    1     5      6 |
  7. |    2     6      7 |
  8. |    2     9      8 |
  9. |    2    11      9 |
 10. |    2    12     10 |
     |-------------------|
 11. |    3    13   11.5 |
 12. |    3    13   11.5 |
 13. |    2    14     13 |
 14. |    1    15   14.5 |
 15. |    3    15   14.5 |
     +-------------------+
```

Kruskal-Wallis: Example

```
. kwallis var, by(cat3);

Kruskal-Wallis equality-of-populations rank test

  +-----------------------+
  | cat3 | Obs | Rank Sum |
  |------+-----+----------|
  |    1 |   4 |    29.50 |
  |    2 |   6 |    49.50 |
  |    3 |   5 |    41.00 |
  +-----------------------+

chi-squared =     0.107 with 2 d.f.
probability =     0.9480

chi-squared with ties =     0.108 with 2 d.f.
probability =     0.9476
```

- $H = \frac{12}{15(16)} 4(29.5/4 - 8)^2 + 6(49.5/6 - 8)^2 + 5(41/5 - 8)^2 = .106875$

Kruskal-Wallis: Example

```
. set seed 010101;

. permute var h = r(chi2), reps(10000) nowarn nodots: kwallis var, by(cat3);

Monte Carlo permutation results                    Number of obs   =         15

      command:  kwallis var, by(cat3)
            h:  r(chi2)
  permute var:  var

------------------------------------------------------------------------------
T            |     T(obs)       c       n   p=c/n   SE(p) [95% Conf. Interval]
-------------+----------------------------------------------------------------
           h |   .1068756    9551   10000  0.9551  0.0021  .9508566   .9590757
------------------------------------------------------------------------------
Note:  confidence interval is with respect to p=c/n.
Note:  c = #{|T| >= |T(obs)|}
```

Kruskal-Wallis: Small sample issues

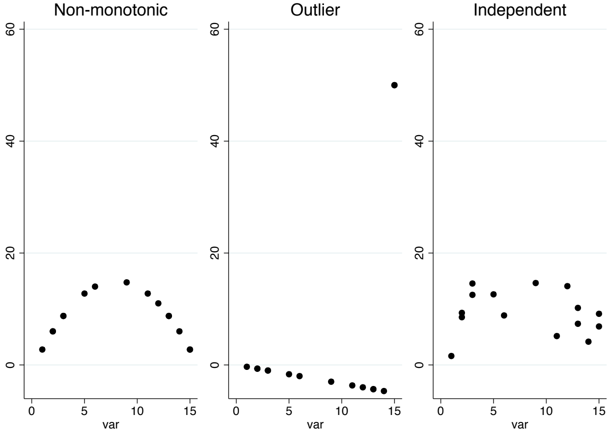
- When n grows larger, it the distribution of H approaches a χ^2-distribution
- Adjustments for ties necessary
- Sample size need to allow able to derive permutation distribution
- Not a test of location unless distributions of k groups similar

Alternatives to correlation coefficients

Spearman's ρ: Basics

- Pearsons correlation for variables converted to ranks
- Remember: $\rho X, Y = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$
- Spearman's ρ tells us something about the proportion of variability accounted for but computed from ranks
- Assumptions:
  - X and Y random sample and iid
  - ordinal or interval data
  - Monotonic relationship between the two variables
- H_0 : The variables do not have a rank-order relationship

Spearman's ρ: Example



Spearman's ρ: Example

- Let's look at pairwise Pearson's correlations coefficients

```
.        pwcorr var varX, sig
             |      var varNon~n varOut~r   varInd
-------------+------------------------------------
         var |   1.0000
             |
             |
   varNonMon |  -0.0657   1.0000
             |   0.8161
             |
  varOutlier |   0.4443  -0.5894   1.0000
             |   0.0971   0.0208
             |
      varInd |  -0.1208   0.5183  -0.1290   1.0000
             |   0.6680   0.0478   0.6467
```

Spearman's ρ: Example

```
. spearman var varX, stats(rho p)
             |      var varNon~n varOut~r   varInd
-------------+------------------------------------
         var |   1.0000
             |
             |
   varNonMon |  -0.0820   1.0000
             |   0.7713
             |
  varOutlier |  -0.2986  -0.5743   1.0000
             |   0.2797   0.0251
             |
      varInd |  -0.1380   0.5251  -0.0771   1.0000
             |   0.6238   0.0445   0.7849
```

Spearman's ρ: Small sample issues

- Relationship needs to be only monotonic not linear (or normal) as for Pearsons correlation coefficient
- Rather robust to outliers (thanks to ranks again)
- Are transformations an option to satisfy monotonicity? It is a rank measure, what transformations would that be?
- With $N > 30$, Pearson's r and Spearman's ρ are sufficiently equivalent – critical value with $p = 0.05$ for Pearson's with 28 df is .361, for Spearman's with N=30 is .363

Spearman's ρ: Small sample issues

- Report Spearman's ρ with a proper summary statistic of the data (e.g., median and IQR)
- Check for number of ties
- Don't use correlation coefficients for data with limit range (e.g. Likert scale)
- (Don't forget adjustments to p-value if testing multiple hypothesis)

Kendall's τ: Basics

- Define n as number of observations, any pair of ranks (x_i, y_i) and (x_j, y_j) of one variable pair as concordant if $(x_i - x_j)(y_i - yj) > 0$ and discordant otherwise – where $C(D)$ is number of concordant (discordant pairs),
- $\tau_a = \frac{C-D}{n(n-1)/2}$
- $\tau_b = \frac{C-D}{\sqrt{n(n-1)/2-U}\sqrt{n(n-1)/2-V}}$
  with U and V being the sum of the number of tied values in all tied sets in variable X and Y, respectively
- It's a probability: difference between the probability that two variables are in the same order in the observed data versus the probability that the two variables are in different orders

Kendall's τ: Example

```
.       ktau var varX, stats(rho obs p)

+-----------------+
|  Key            |
|-----------------|
|   tau_a         |
|   tau_b         |
|   Sig. level    |
+-----------------+

             |        var   varNon~n   varOut~r     varInd
-------------+--------------------------------------------
         var |     0.9619
             |     1.0000
             |
             |
   varNonMon |    -0.0571     0.8762
             |    -0.0622     1.0000
             |     0.7996
             |
  varOutlier |    -0.4667    -0.4000     0.9619
             |    -0.4851    -0.4357     1.0000
             |     0.0165     0.0373
             |
      varInd |    -0.0857     0.4000    -0.0667     1.0000
             |    -0.0874     0.4273    -0.0680     1.0000
             |     0.6907     0.0382     0.7654
             |
```

Kendall's τ: Small sample issues

- τ approaches a normal distribution more rapidly ($N \geq 10$)) than Spearman's ρ (Gilpin 1993), with continuos variables even for ($N \geq 8$), Kendall/Gibbons 1990)
- Said to be more accurate with smaller samples because less sensitive to discrepancies in data
- Give vastly different exact p-values for various sample sizes and data values: $-1 <= 3 * \tau - 2\rho <= 1$ (Siegel/Castellan 1988)

More tests

- Goodman and Kruskal's gamma:
  - $\gamma = \frac{C-D}{C+D}$
  - Distribution of G has high variability and is skewed for small to moderate sample sizes, convergence to ideal distribution in the asymptotic case is slow (Gans/Robertson 1981)
- Somers' D:
  - Define T_Y as number of pairs with equal y but unequal x
  - $D_{YX} = \frac{C-D}{C+D+T_Y}$
  - improvement in predicting X attributed to knowing an observation's value Y
  - Note, `ranksum` and `signrank` both test $D_{YX} = 0$
  - Asymptotic approximations work when smaller of the two samples has $N \geq 8$

References

Sign tests, rank sum tests

- Wilcoxon (1945): Individual comparisons by ranking methods. Biometrics Bulletin 1, pp. 80–83
- Sen (blog): Stata and R ranksum test p-values diff
- Harris and Hardin (2013): Exact Wilcoxon signed-rank and Wilcoxon Mann–Whitney ranksum tests,Stata Journal 13(2), pp.337-43
- Ronan Conroy (2012): What hypotheses do “nonparametric two-group tests actually test”, Stata Journal 12(2), pp.182-90
- Bellera, Julien, and Hanley (2010): Normal Approximations to the Distributions of the Wilcoxon Statistics: Accurate to What N? Graphical Insights, Journal ofStatistics Education 18(2), pp. 1-17

Alternatives to correlation coefficients

- Gilpin (1993): Table for conversion of Kendall's Tau to Spearman's Rho within the context measures of magnitude of effect for meta-analysis, Educational and Psychological Measurement 53(1), pp. 87-92.
- Gans and Robertson (1981): Distributions of Goodman and Kruskal's Gamma and Spearman's ρ in 2×2 Tables for Small and Moderate Sample Sizes, Journal of the American Statistical Association 76(376), pp.942-6
- Cuzick (1985): A Wilcoxon-type test for trend, Statistics in Medicine 4(4), pp. 543-7
- Roger Newson's Stata packages website (including `somersd`)

Texts and further references

- Hollander, Wolfe, Chicken (2014): Nonparametric Statistical Methods, 3rd Edition
- NSM3: Functions and Datasets to Accompany Hollander, Wolfe, and Chicken - Nonparametric Statistical Methods, Third Edition
- Siegel and Castellan (1988): Nonparametric statistics for the behavioral sciences, McGraw-Hill

- Razali and Wah: Power comparisons of Shapiro-Wilk, Kolmogrov-Smirnov, Lilliefors and Anderson-Darling tests