NonParametric Methods

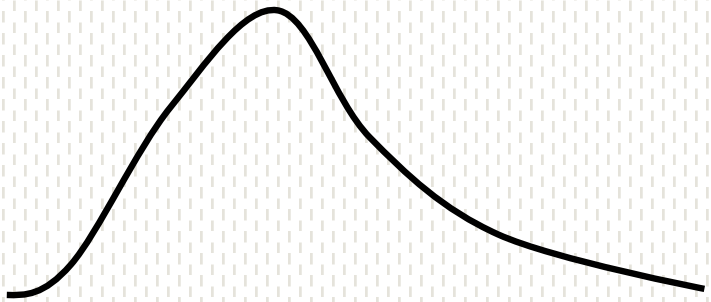
- Introduction
- Ranks & Median
- Paired Wilcoxon Signed Rank
- Mann-Whitney test (or Wilcoxon Rank Sum test)
- Others….

Parametric vs. non-parametric (distribution free) tests

- Non-parametric tests:
  - No normality requirement
  - Do require that the underlying distributions being compared have the same basic shape
  - Ranks are less sensitive to outliers and to measurement error
- If the underlying distributions are approximately normal, then the parametric tests are more powerful

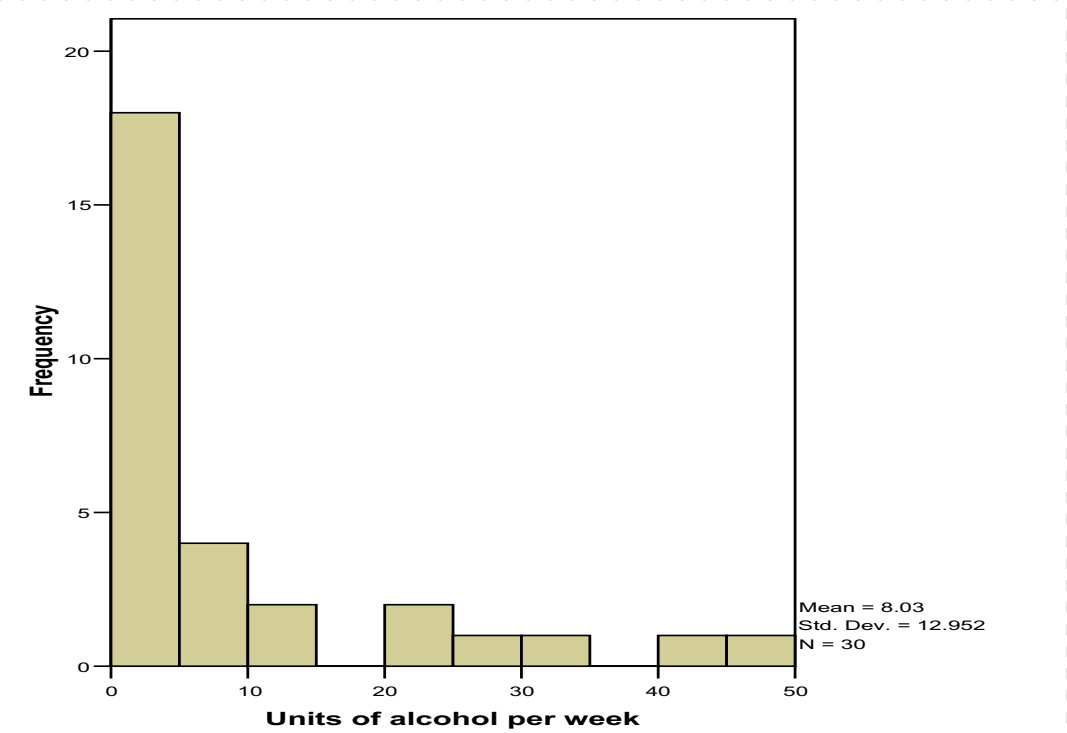
What are non-parametric tests?

- ‘Parametric’ tests involve estimating parameters such as the mean, and assume that distribution of sample means are ‘normally’ distributed
- Often data does not follow a Normal distribution e.g.; number of cigarettes smoked, health costs, etc.
- Positively skewed distributions



What are nonparametric tests?

A positively skewed distribution



What are nonparametric tests?

- ‘Non-parametric’ tests were developed for these situations where fewer assumptions have to be made
- Sometimes called Distribution-free tests
- NP tests STILL have assumptions but are less stringent and less powerful
- NP tests can be applied to Normal data but parametric tests have greater power IF the assumptions are met

Ranks

- Practical differences between parametric and NP are that NP methods use the ranks of values rather than the actual values
- e.g.

  1,2,3,4,5,7,13,22,38,45 - actual

  1,2,3,4,5,6, 7, 8, 9,10 - rank

Median

- The median is the value above and below which 50% of the data lie.
- If the data is ranked in order, it is the middle value
- In symmetric distributions the mean and median are the same
- In skewed distributions, median more appropriate

Student's t-test

- T-test used to test whether the mean of a sample is sig different from a hypothesised sample mean
- T-test relies on the sample being drawn from a normally distributed population
- If sample not Normal then use the Wilcoxon Signed Rank Test as an alternative

Wilcoxon tests

- Frank Wilcoxon was Chemist at American Cyanamid who wanted to develop a test similar to t-test but without requirement of Normal distribution
- Presented paper in 1945
- Wilcoxon Signed Rank Ξ non-parametric paired t-test
- Wilcoxon Rank Sum Ξ non-parametric independent t-test

Wilcoxon Signed Rank Test

- NP test relating to the median as measure of central tendency
- The ranks of the absolute differences between the data and the hypothesised median calculated
- The ranks for the negative and the positive differences are then summed separately (W_- and W_+ resp.)
- The minimum of these is the test statistic, W

Wilcoxon Signed Rank Test Normal Approximation

- As the number of ranks (n) becomes larger, the distribution of W becomes approximately Normal
- Generally, if n>20
- Mean W=n(n+1)/4
- Variance W=n(n+1)(2n+1)/24
- Z=(W-mean W)/SD(W)

Wilcoxon Signed Rank Test Assumptions

- Population should be approximately symmetrical but need not be Normal
- Results must be classified as either being greater than or less than the median ie exclude results=median
- Can be used for small or large samples
- Similar to the paired t-test
- In Stata $signrank\ x_1 = x_2$

Paired samples t-test

- Disadvantage: Assumes data are a random sample from a population which is Normally distributed

- Advantage: Uses all detail of the available data, and if the data are normally distributed it is the most powerful test

The Wilcoxon Signed Rank Test for Paired Comparisons

- Disadvantage: Throws away the real values and only the sign (+ or -) of any change is analysed

- Advantage: Easy to carry out and data can be analysed from any distribution or population

The Wilcoxon Signed Rank Test for Paired Comparisons

- Similar calculation to the Wilcoxon Signed Rank test, only the differences in the paired results are ranked
- A group of 10 patients with chronic anxiety receive sessions of cognitive therapy. Quality of Life scores are measured before and after therapy.

$W_- = 2$
$W_+ = 7$
1 tied

QoL Score				
x1	x2	Diff	Rank	-/+
6	9	3	5.5	+
5	12	7	10	+
3	9	6	9	+
4	9	5	8	+
2	3	1	4	+
1	1	0	3	tied
3	2	-1	2	-
8	12	4	7	+
6	9	3	5.5	+
12	10	-2	1	-

Stata output: Wilcoxon vs. t-test

Wilcoxon signed-rank test

sign	obs	sum ranks	expected
positive	2	6.5	27
negative	7	47.5	27
zero	1	1	1
all	10	55	55

unadjusted variance	96.25
adjustment for ties	-0.25
adjustment for zeros	-0.25
adjusted variance	95.75

Ho: var1 = var2
z = -2.095
Prob > |z| = 0.0362

. ttest var1= var2

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
var1	10	5	1.021981	3.231787	2.688119	7.311881
var2	10	7.6	1.284091	4.060651	4.695185	10.50481
diff	10	-2.6	.9568467	3.025815	-4.764538	-.4354624

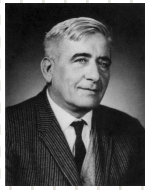
mean(diff) = mean(var1 - var2) t = -2.7173
Ho: mean(diff) = 0 degrees of freedom = 9

Ha: mean(diff) < 0 Ha: mean(diff) != 0 Ha: mean(diff) > 0
Pr(T < t) = 0.0119 Pr(|T| > |t|) = 0.0237 Pr(T > t) = 0.9881

Mann-Whitney test Ξ Wilcoxon Rank Sum

- Henry Mann, professor of mathematics @ Ohio State & Don Whitney was his student
- Used when we want to compare two unrelated or INDEPENDENT groups
- For parametric data you would use the unpaired (independent) samples t-test
- The assumptions of the t-test were:
  1. The distribution of the measure in each group is approx Normally distributed
  2. The variances are similar

HB Mann



Wilcoxon (Mann-Whitney) Rank Sum Example

- The following data shows the number of alcohol units per week collected in a survey:
- Men (n=13): 0,0,1,5,10,30,45,5,5,1,0,0,0
- Women (n=14): 0,0,0,0,1,5,4,1,0,0,3,20,0,0
- Is the amount greater in men compared to women?
- STATA: ranksum Alcohol, by(Gender)

Stata output: Wilcoxon vs. t-test

```
. ranksum Alcohol, by(Gender)

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

Gender	obs	rank sum	expected
0	12	180.5	162
1	14	170.5	189
combined	26	351	351

```
unadjusted variance      378.00
adjustment for ties      -49.62
                         ------
adjusted variance        328.38

Ho: Alcohol(Gender==0) = Alcohol(Gender==1)
             z =   1.021
    Prob > |z| =   0.3073
```

```
. ttest Alcohol, by(Gender)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	12	4.75	2.471428	8.561276	-.6895755	10.18958
1	14	2.428571	1.42472	5.330814	-.6493492	5.506492
combined	26	3.5	1.364664	6.958448	.689422	6.310578
diff		2.321429	2.753409		-3.361328	8.004185

```
    diff = mean(0) - mean(1)                                  t =   0.8431
Ho: diff = 0                     degrees of freedom =       24

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.7963         Pr(|T| > |t|) = 0.4075          Pr(T > t) = 0.2037
```

Other Non-Parametric Tests

- Sign Test
- Median Test
- Chi-squared test – next week!!
- Spearman's Rank Correlation Coefficient – in 2 weeks when we tackle regression!