1/19/2010 36-402/608 ADA-II H. Seltman
Handout #4: Alternate Test Approaches

1. Context

   (a) Classical testing and estimation: based on null sampling distributions of statistics under a particular model

   (b) Bayesian statistics: combine a likelihood for parameters given a model and some data with a prior distribution for the parameters to get a posterior distribution of the parameters

   (c) "Modern" nonparametric statistics: nonparametric really means an infinite number of parameters. This approach includes nonparametric regression (e.g., loess curves, splines), density estimation, and non-parametric classification.

   (d) ⇒ "Classic" nonparametric statistics: so-called distribution free (no Normality assumption) alternatives to t-test, paired t-test, and ANOVA.

   (e) ⇒ Permutation (randomization) tests: general tests based on alternate possible randomizations

   (f) Bootstrap and other resampling methods: general tests based on using the observed error distribution to represent the true error distribution

2. Alternative to the two-independent-samples t-test: **Wilcoxon rank-sum / Mann Whitney U test**

   Concept: If an outcome follows any distribution shape, and a randomized two-level treatment has zero effect (e.g., active treatment works exactly the same as control), then the pattern of ranks of the outcomes are similar for both groups.

   Comments: Using ranks makes this resistant to outliers and other forms of non-normality. It also makes it a test of equality of medians. It is *not* able to correctly decide if the medians are equal if the two groups have different variances.

Example:

```
Cholesterol in controls: 160, 190, 200, 203, 210, 212, 214
Cholesterol with drug X: 165, 180, 193, 203, 217
```

Treatment	Cholesterol	Rank	Adjusted Rank
C	160	1	1
X	165	2	2
X	180	3	3
C	190	4	4
X	193	5	5
C	200	6	6
C	203	7	7.5
X	203	8	7.5
C	210	9	9
C	212	10	10
C	214	11	11
X	217	12	12

Method:

(a) Assure that each group has at least, say, 5 values, and that there are not too many ties

(b) Calculate the sum of the ranks (78)

(c) Calculate the mean rank (6.5)

(d) Calculate the sum of the ranks for (say) the smaller group: 29.5

(e) Calculate the expected sum of ranks for the smaller group: 5*6.5=32.5

(f) Calculate the deviation of the group rank sum from expected: -3.0

(g) Calculate the variance of the ranks (s_R^2): 12.95

(h) Calculate the sampling variance of the deviation from $s_R^2 \frac{n_1 n_2}{n_1+n_2}$: 37.77

(i) Calculate the Z score: $\frac{-3}{\sqrt{37.77}} = -0.488$. Or using a continuity correction, Z=$\frac{-2.5}{\sqrt{37.77}} = -0.407$.

(j) Using $2 * \text{pnorm}(-\text{abs}(z))$, calculate the two-sided p.value: 0.62 (or 0.68, corrected).

We conclude that there is no evidence of a treatment effect, assuming equal arbitrary distribution shapes.

Note: There is also a method for constructing confidence intervals on the median by adding and subtracting different values to one group until the p.value just hits 0.05.

3. The rank sum test is a specific example of a **permutation (randomization) test**. For any given statistic, e.g., mean difference, we can ask what is the distribution of the statistic across many (all) different ways the treatment assignment could have been made. If the observed statistics is "typical" then we have no reason to think that the randomized treatments differ. If it is "unusual", then we have evidence that the randomization caused the difference.

   The rank-sum test is the permutation test for the difference of the expected sum of the ranks of one group from the observed sum.

4. Aside: **Welch t-test** (default in R) uses an altered df to estimate the null sampling distribution of the t-statistic when the equal variance assumption is violated. Use `var.equal=TRUE` to get the classic t-test that assumes equal variance.

5. **Sign test**: With *paired* data converted to the sign of the difference or with a single set of binary outcomes (e.g., "before" vs. "after"), this quickly tests if one value predominates. One way to express the null hypothesis for the paired data is $\Pr(Y_{i1} > Y_{i2}) = 0.5$. This is equivalent to testing that the median difference is equal to 0 (or some other cutoff).

   This is a non-parametric alternative to the paired t-test.

   Method:

   (a) Discard ties, label the remaining number of cases as n

   (b) The observed number in one group (either) is called K

   (c) The expected number in one group under H_0 is $n/2$ (if the true $p = 0.5$)

   (d) Assuming independence, the sampling variance of K under H_0 is $np(1-p) = n/4$

   (e) Calculate $Z = \frac{K-n/2}{\sqrt{n/4}}$

   (f) Reject H_0 if $|Z| \geq 1.96$ or calculate a p-value.

6. Paired non-parametric test: **Wilcoxon signed-rank test**

   Comments: More powerful version of the sign test, so it's also a non-parametric alternative to the paired t-test. This is similar to the sign test for differences, but takes into account how large the difference is in ranking. This is resistant to outliers.

   Method: Based on sum of ranks for pairs where the difference is in one direction. See textbook for more details, if interested.

7. Breakout and Discussion