Nonparametric Tests

Day 3, Morning

Reminder: Parametric and Nonparametric Procedures

- Parametric methods (bit of a misnomer):
  - Based on assuming a particular underlying population distribution, usually "normal."
  - Can sometimes be used even without that assumption, for large samples.
- Nonparametric methods:
  - Can be used without assuming a distribution.
  - Often not as "powerful" as parametric methods.
- When in doubt, safer to use nonparametric method

Review of Parametric Inference Procedures

- One sample t-test and confidence interval:
  - Estimate and test mean of one population
  - Example: Is pH for Davis rainfall 5.6? Get C.I.
- Paired t-test and confidence interval:
  - Estimate and test mean of differences for pairs
  - Example: Pilots sober and with alcohol
- Independent samples t-test:
  - Are the means of two populations equal? If not, what is a confidence interval for the difference?
  - Example: Dolomite vs limestone wells
  - Example: Mercury concentration in Tomales Bay

Sign Test: One Sample or Paired Data

Example: Pilot and alcohol data. Are population differences equally likely to be positive or negative? Or more likely to be positive?

Pilot	No Alcohol	Alcohol	Difference = No Alcohol − Alcohol
1	261	185	76
2	565	375	190
3	900	310	590
4	630	240	390
5	280	215	65
6	365	420	−55
7	400	405	−5
8	735	205	530
9	430	255	175
10	900	900	0

Five Steps in Hypothesis Test

Step 1: Null and alternative hypotheses

Two ways to state these:

1. One sample or sample of differences, want to test specific value for the population median M.

   Null: H_0: $M = M_0$

   Alternative: H_a: $M > M_0$ *or* H_a: $M < M_0$ *or* H_a: $M \neq M_0$

2. Matched Pairs (X, Y)

   H_0: $P(X > Y) = .5$ [X equally likely to be > or < Y]

   H_a: $P(X > Y) > .5$ *or* H_a: $P(X > Y) < .5$ *or* H_a: $P(X > Y) \neq 5$

Example: No alcohol > Alcohol?

Step 1: Null and alternative hypotheses

Version 1: Is the median population difference 0, or is it greater than 0?

H_0: $M = 0$

H_a: $M > 0$

Version 2: Is P(No alcohol > Alcohol) = .5, or is it greater than .5, so No alcohol values are larger?

H_0: $P(X > Y) = .5$

H_a: $P(X > Y) > .5$

Step 2: Test statistic (no data conditions needed)

S^+ = Number of observations greater than M_0
or Number of observations with x > y
S^- = Number of observations less than M_0
or Number of observations with x < y
Ties are not used, so use $n = S^+ + S^-$.

Ex: There were 2 negative differences and 1 zero difference, so $S^+ = 7$, $S^- = 2$, $n = 9$.

Is that convincing evidence that in population No alcohol values > alcohol values?

Step 3: Finding the *p*-value

Remember, *p*-value is:
- Probability of observing a test statistic as large as or larger than that observed
- in the direction that supports H_a
- *if* the null hypothesis is true.

Example: For $n = 9$, what is the probability of observing 7 or more positive differences, if in fact probability is ½ for each pair?

Analogy: Same as probability of 7 or more heads in 9 flips of a fair coin!

Properties of a *Binomial Experiment*

1. There are *n* "trials" where *n* is determined in advance. (*Sign test*: *n pairs or single data points*)
2. There are *the same two possible outcomes* on each trial, called "success" and "failure" and denoted S and F. (*Sign test: "Success" is X > Y, "Failure" is X < Y*)
3. The *outcomes are independent* from one trial to the next. Knowledge of one does not help predict the next one. (*Sign test*: *Data points or pairs are independent*)
4. The probability of a "success" *remains the same* from one trial to the next, and this probability is denoted by *p*. The probability of a "failure" is $1-p$ for every trial. (*Sign test, under the null p = .5.*)

P-value for the sign test, rationale

- For a binomial experiment, if S^+ = number of successes, then for $k = 0, 1, \ldots, n$

$$\Pr(S^+ = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

- When $p = .5$, this becomes

$$\Pr(S^+ = k) = \frac{n!}{k!(n-k)!}(.5)^n$$

- But we want P(S^+ <u>*or more*</u> successes)

Sign Test *p*-value

Let B = binomial with *n* trials and $p = .5$.
H_a: "greater than" → *p*-value = P(B ≥ S^+)
H_a: "less than" → *p*-value = P(B ≤ S^+)
H_a: "not equal" → *p*-value =
2 × [smaller of P(B ≤ S^+) and P(B ≥ S^+)]
Alcohol Example:
"Greater than" so *p*-value = P(B ≥ S^+)
= P(B is 7, 8 or 9) = .0195

Finding Sign Test *p*-value

R Commander:
Distributions → Discrete distributions
→ Binomial distribution → Binomial tail probabilities
Fill in boxes:
<u>Variable value(s):</u> Fill in S^+ (Ex: 7)
<u>Binomial trials</u>: Fill in *n* (Ex: 9)
<u>Probability of success</u>: Leave default of .5
Click radio button "lower tail" or "upper tail."

Hands-On Activity:
To be given in class

Sign test for atrazine concentrations.

Results shown in class.

Nonparametric test for independent samples: Two-sample Wilcoxon

- Other names are Wilcoxon rank sum test and Mann-Whitney test.
- Assume two populations have approximately the same shape, but shape is not specified.
- Null hypothesis: Centered at same value (so the two distributions are the same)
- Alternative hypothesis: One distribution is shifted to the right (or left) of the other. (One-sided test specifies which direction.)

Picture (From *Statistical Ideas and Methods*, Utts and Heckard)

Figure S2.3 An example of the assumption of same shape but possibly different location

Wilcoxon test, rationale and method

- Assign *ranks* to all of the observations in both samples, with 1 = smallest, 2 = next, ... to N_T largest, where $N_T = n_1 + n_2$
- For ties, use average rank.
- Test statistic W = Sum of ranks for Sample 1 (sometimes subtract minimum it could be).
- The sum of all numbers from 1 to N_T is $\frac{N_T(N_T+1)}{2}$
- The *smallest* W could be is if Sample 1 values are *all* smaller then Sample 2 values, then W = $\frac{n_1(n_1+1)}{2}$

Wilcoxon test rationale, continued

- Remember, test statistic W = Sum of ranks for sample 1, but sometimes use:

  (Sum of ranks for Sample 1) – $\frac{n_1(n_1+1)}{2}$

- Suppose H_0 is true. Then W should be close to the proportion n_1/N_T of sum of all possible ranks, or

$$\frac{n_1}{N_T} \times \frac{N_T(N_T+1)}{2} = \frac{n_1(N_T+1)}{2}$$

- *P*-value complicated, but R commander finds it.

Example: Walker Creek Delta mercury

Step 1:

H_0: Mercury values for Walker Creek Delta and "other" have same distribution.

H_a: Distribution of mercury values for Walker Creek Delta is shifted to the right compared to "other"

Sum of all ranks (1 to 19):

$$\frac{19(19+1)}{2} = 190$$

If H_0 true:

$$W \approx \frac{11}{19} \times 190 = 110$$

Step 2:

Compute W = ranks of values in blue

= sum of 1 to 10 + 12.5 = 67.5.

Minimum possible is sum of 1 to 11 = 66, so alternative version is to use W = 67.5 – 66 = 1.5

1	0.056
2	0.069
3	0.072
4	0.087
5	0.095
6	0.13
7	0.15
8	0.16
9	0.21
10	0.22
11	0.38
12.5	0.62
12.5	0.62
14	0.67
15	0.68
16	0.78
17.5	1.4
17.5	1.4
19	1.6

R Commander

Statistics → Nonparametric tests

→ Two-sample Wilcoxon test

Choose variables, specify alternative.

Results for the example:

```
data:  Mercury by WalkerOther
W = 1.5, p-value = 0.0002592
alternative hypothesis: true location shift is
  less than 0
```

Since *p*-value is so small, reject null hypothesis. Conclude mercury levels are higher in Walker Creek delta. (Note: t-test gave p-value of .0008)

Hands-On Activity: To be given in class

Simulation comparing *t*-test and Wilcoxon test for skewed (log normal) data

Dealing with Non-detects

Left Censored Data = Non-detects

- Non-detects occur when the actual value is below the detection limit of the measuring instrument. They are recorded as < dl, where dl = detection limit.
- Special case of "censored data."
- Simple methods include replacing all non- with a fixed value. Commonly used values are 0, dl/2, dl.
- Simulation studies have shown that this is not a good idea.

Robust Probability Plot Method

- Construct a "probability plot" using data above the reporting limit.
- Use regression to extrapolate what the values would be below the reporting limit.
- Use those values to find summary statistics.
- See pictures on next slide, from Helsel and Hirsch.

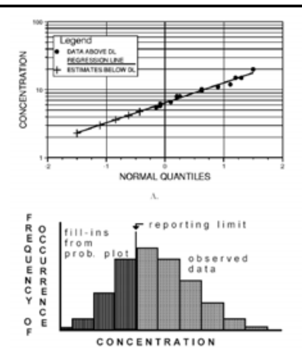

Figure 13.3. Robust (probability plot) method of estimating summary statistics
A) regression of log of concentration versus normal score is used to extrapolate "fill-in" values below the reporting limit.
B) these "fill-ins" are retransformed back to original units, and combined with data above the reporting limit to compute estimates of summary statistics

RPcalc

- Written by Steven Saiz to accompany new regulatory procedures in the CA Ocean Plan to determine if a discharge has the "reasonable potential" to exceed a water quality standard.
- Download it at:

http://www.waterboards.ca.gov/water_issues/programs/ocean/docs/oplans//rpcalc.zip

- The program will estimate summary statistics and calculate the upper tolerance bound even if there are non-detects in the data.
- Uses the robust probability plot method to handle non-detects.

Demonstration and Example of Using RPcalc

- Lead data from San Francisco, from Steve Saiz
- Has non-detects with different detection limits