Wilcoxon signed rank test is similiar, except it ranks ~~[illegible]~~ the differences between paired values.

Resampling or Monte-Carlo approach

Idea: Construct artificial datasets from a collection of real data, by resampling the observations in a manner consistent with the null hypothesis.

~~Take~~

This is used to construct a null distribution, against which the alternative (interesting) hypothesis can be evaluated

You necessarily need a computer to do this, which can randomize the data.

Permutation: Samples are drawn from a distribution without replacement so ~~[illegible]~~ each individual observation is represented only once.

Bootstrap: Samples are drawn from a distribution with replacement so each individual observation may be represented multiple times

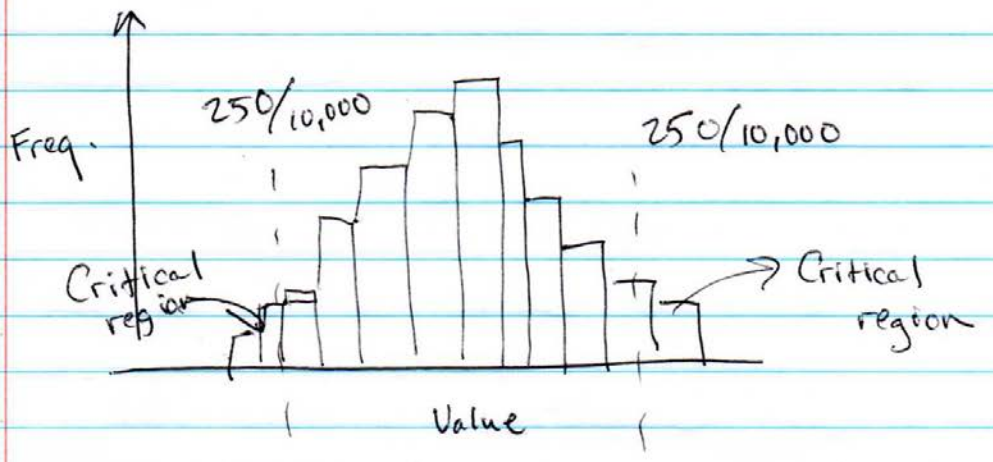
Say we want to sample slips of paper from a hat. Total of n slips of paper. To construct one sample (n, draws) (where $n_1 < n$)

Permutation approach: Each slip of paper is drawn without replacement to construct a sample size of n_1 draws.

Bootstrap approach: After the draw of each slip, it is put back into the hat. Draw again till reach n_1 draws. → less restrictive null hypothesis.

Whatever approach is used, the idea is to generate a distribution from a given amount of estimates. Usually the number of estimates is fairly large to create a robust distribution.

Say 10,000 estimates



In a estimate with 10,000 the critical region occurs near the tail of the distribution. In the critical region, if the original sample falls there it is statistically significant.

Caveat: When using these methods, first need to make sure that the data are not autocorrelated in time (i.e. samples are independent).

Returning to the lightning example discussed earlier with respect to the Mann-Whitney U statistic. (p. 165 Wilks)

Lightning counts from
$n_1 = 12$ seeded storms
$n_2 = 11$ unseeded storms

Compute L-scale statistic, basically gives a measure of spread in a given sample:

$$\lambda_2 = \frac{(n-2)!}{n!} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |x_i - x_j|$$

"x" = # strikes

$x_i - x_j$ summation term gives an overall measure of the difference between all possible pairs in the sample size
Equivalent to variance (F-test, χ^2 test)

Compute this statistic for seeded and unseeded storms.

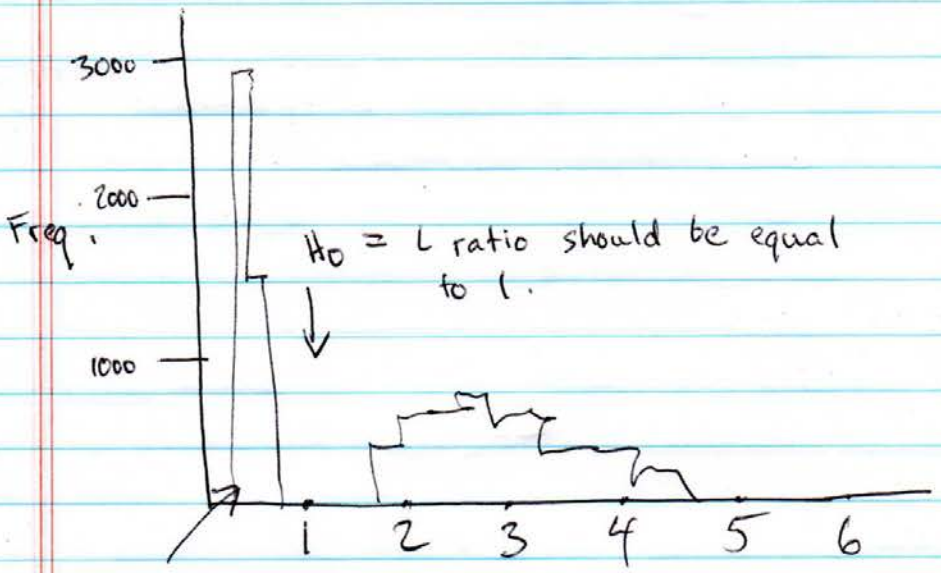
$$L_{scale} = \frac{\lambda_2(\text{seeded})}{\lambda_2(\text{unseeded})} \rightarrow \text{test-statistic}$$

- Seeded storms more variable in lightning if greater than one
- Less variable in lightning if less than one.

To construct null distribution:

- Randomize the entire sample of n (23) storms, pick n_1 and n_2. (permutation → because not replacing after draw)
- Compute the test statistic
- Do this a ~~gi~~ sufficient number of times to generate a frequency distribution (like 10,000 times)

Get Fig 5.7.



Observed ratio falls here (so it is significant)

Smaller than all but 49 of 10,000 permutation tests.

Boot strap example discussion for the lightning data case is given on p.p. 168 - 169.

What the less restrictive null hypothesis can account for in bootstrapping,

- (spread does not necessarily just depend on seeding
- other aspects of the distributions may be different.

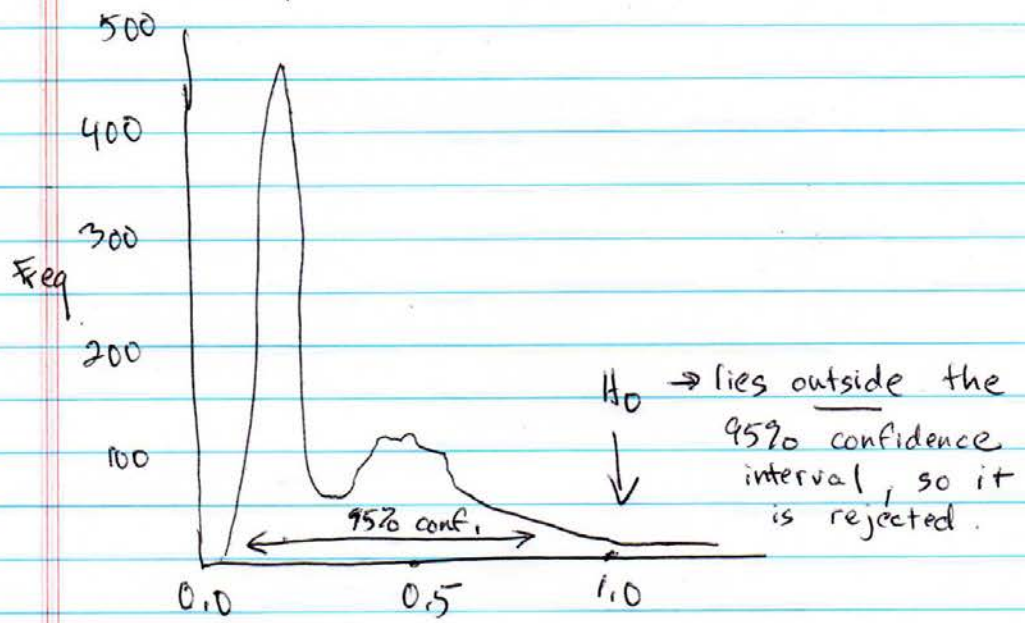


Fig. 5.9.

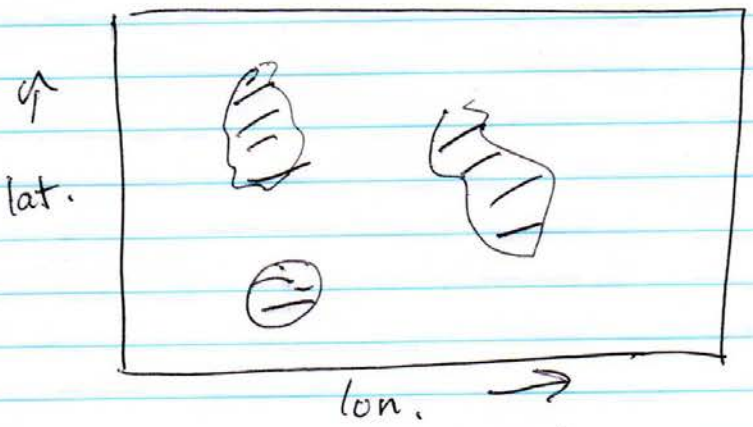
∴ other aspects of the distribution are probably different!

Field significance – A VERY important non-parametric test in geophysical data analysis.

"The" reference for this, which has been cited thousands of times, is Livezey and Chen (1983). (And if you don't do it, Livezey may embarass you at a conference!!)

In geophysical fields, often looking at patterns of significance in space, based on many local tests at the individual gridpoints.

Say you're doing a composite analysis, and you get a map which indicates the level of significance at each point.



Shaded areas represent local significance at the gridpoints exceeding some critical

value. Say the 95% level on a t-test.

Two things you need to think about before you conclude things are statistically significant or not.

1) Are the data related in time? That is, is there autocorrelation present? If so, this will reduce the degrees of freedom used in the local test.

We'll address how to handle this a bit later in the course--

For now, we'll assume that the data are independent in time.

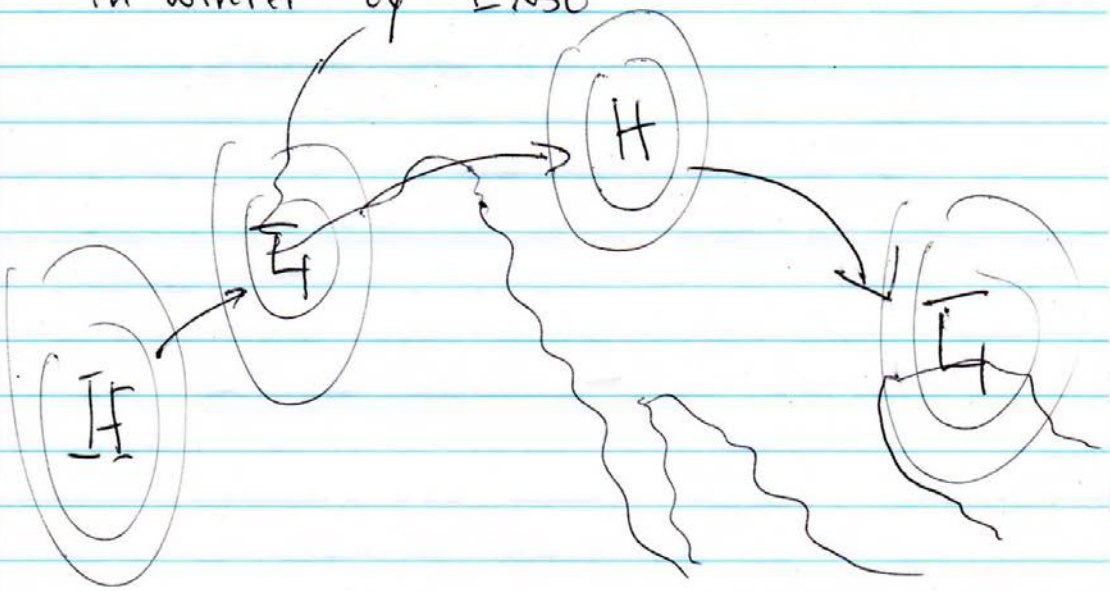
2) Are the data related in space? If so, then it would be expected that just a few dominant spatial patterns account for the significant shaded areas on the map.

Example of spatial patterns: 700-mb heights over the northern hemisphere.

The geopotential height at one point is more often than not related to what happens at other points.

Consider the PNA pattern, which get if composite 700-mb heights in winter by ENSO

PNA

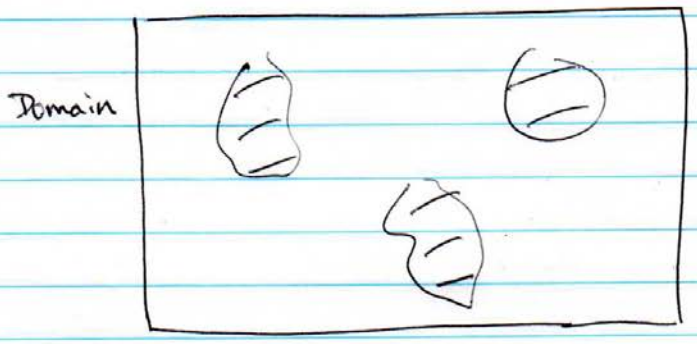


Result of changes in convection in the eastern tropical Pacific → produces a hemispheric pattern of height anomalies, or tele connection.

So because there are teleconnections, or relationships in space, what happens at one point IS NOT independent of what happens at another point – there is spatial correlation!

How to account for this?

Original composite based on n, from a sample of n (e.g. ENSO years)



Procedure to assess significance of shaded regions in the composite example.

1) Calculate the percentage of area within the domain that passes the local significance test.

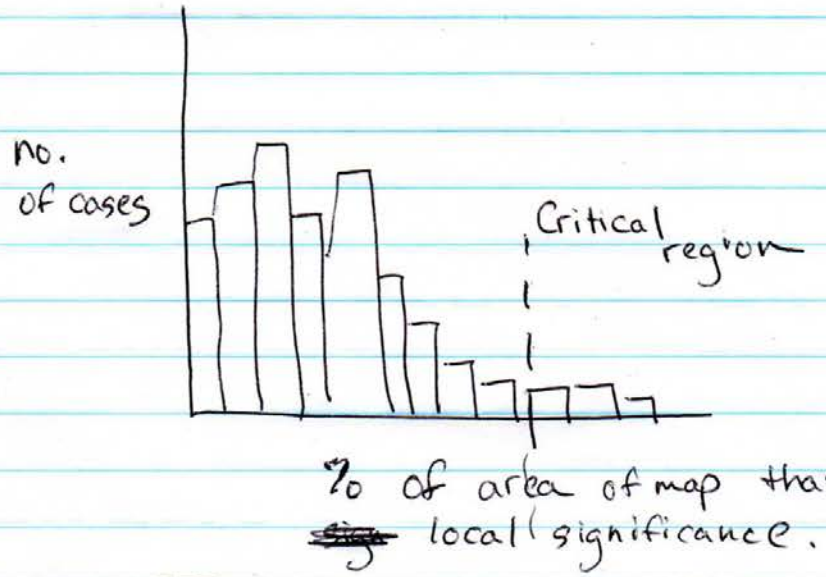
2) ~~Rec~~ Randomly reorder the maps

3) Calculate the percentage of area

that passes the local significance test.

4) Repeat this procedure many times to generate a null distribution of percentage area ~~that~~ on the map that satifies a local test.

~~The~~ Fig. 5.12 from Wilks.



Pass test

5) IF the percentage area on your original map falls within the critical region, then field significance is satisfied.

Fail test

IF the percentage area on original map is below the critical region, then no field significance. In other words, can get the same pattern of sig. shaded regions by a random reordering of the maps.

In terms of how to program

Have a three dimensional array for a given variable:

In FORTRAN

VAR(i , j , t)

i: x-dim, j: y-dim, t: time dimension

To evaluate field significance:

- Keep x-dim & y-dim the same
- Only reshuffle the maps by changing the time index.