Nonparametric methods require minimal assumptions about the form of the distribution of the population of interest.
Ex.: Assume a continuous distribution, but no other assumptions
Ex.: Assume the population distribution depends on location and scale parameters, but no other information regarding the distribution (normal or otherwise) is assumed or specified.

- Chapter 1 will cover some simple nonparametric tests of hypotheses and confidence intervals based on the binomial distribution.
- We will discuss an example where a nonparametric test may be preferred over a popular normal theory based test.

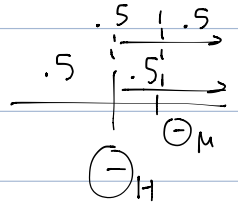
- Suppose: A random sample from a population with continuous F(x).
- Theta_.5
  The median of the population. In other words, theta_.5 is a value such that half the probability is less than theta_.5 and half is greater.

- Want to test: H_0 : theta_.5 = theta_H, H_a : theta_.5 > theta_H
- Example:
  A food product is advertised to contain 75 mg of Na per serving. Some people suspect that servings may actually contain more.

Tests of medians can often be used in the same situation as tests of means.
If a distribution is symmetric and if the mean of the population exists, then the mean and the median are the same value.

Let X_1, X_2,..., X_n be a random sample from a population.
Let B denote the number of X_i's out of n that have values above the hypothesize median theta_H.
If H_0 is true, what is the probability that each X_i has of falling above theta_H? **.5**
Is this the same for all X_i's? If it's a simple random sample, all X's independent, then Yes
Review: What are the assumptions needed for the Binomial setting? Fixed and finite, constant probability across observations, 1 denotes success, 0 denotes failure, independent observations
If null hypothesis is true, what is the distribution of B? B has binomial distribution with # observations = n, probability of success = .5

Want to test if the true median is greater than theta_H
If the true median is greater, then what can we say about *p* (in the context of *B*)? **Greater than** or less than .5?



Want to test if the true median is greater than theta_H
H_0 : p = .5, *H_a : p* > .5

Let's assume n is large enough to use the normal approximation to the binomial. Therefore, we can write the test statistic *Z_B* as:

$$Z_B = \frac{B - n(.5)}{\sqrt{n(.5)(1-.5)}} = \frac{B - n(.5)}{\sqrt{n(.25)}}$$

At a level of significance *a*, we reject *H_0* in favor of *H_a* if

Z_B > z_(1-a). (Why not *Z_B < z_(1-a)*?)

This is a nonparametric test!
We did not have to make any assumptions about the form of the population distribution other than that it is continuous, AND observations are independent (midterm)

H_0 : theta_.5 = 75, H_a : theta_.5 > 75

40 data points
B = data values above 75 = 26

There are 26 data values for which *X_i* > 75 so
Z_B = 26 - 40 * .5 / (sqrt(40*.25)) = 1.9 > 1.645 = z_.95

a = always assumed to be .05

We reject the null hypothesis at a significance level of .05, and conclude that the median of the population is greater than 75 mg. < Expected for all answers

We reject the null hypothesis at a significance level of .05, and conclude that ___ is ____
We don't reject the null hypothesis at a significance level of .05 and conclude that ____ is ____

Order Statistics
Data (*X_i*) are ordered from smallest to largest and denoted as:
X_(1) < X_(2) < ... < X_(n)
Interval of interest:
X_(a) < theta_.5 < X_(b)
Such that
P(X_(a) < theta_.5 < X_(b)) = 1 - a
Where *1 - a* is the desired probability that the interval captures the median.

X_(a) < theta_.5 < X_(b)
At least *a* of the observations must fall less than theta_.5, and at most *b-1* observations must fall less than or equal to theta_.5.
Since theta_.5 is the median and sicne the distribution of the *X's* is continuous:
P(X < theta_.5) = P(X <= theta_.5) = .5

CDF↗

Since the observations are independent, the probability that at least *a* and at most *b-1* of the observations fall less than theta_.5 is given by the binomial probability with *p = .5:*

$$\sum_{k=a}^{b-1} \binom{n}{k} (.5)^k (.5)^{n-k} = \sum_{k=a}^{b-1} \binom{n}{k} (.5)^n$$

To construct a 100(*1-a*)% confidence interval for theta_.5, choose *a* and *b* such that the sum is *1-a*

$$\sum_{k=a}^{b-1} \binom{n}{k} (.5)^k (.5)^{n-k} = \sum_{k=a}^{b-1} \binom{n}{k} (.5)^n = 1 - \alpha$$

Since the binomial distribution is discrete, there may not be exact limits that give exactly the desired

confidence. However, we choose a level of confidence as close to 100(*1-a*)% as possible without going under this value.

For large samples, approximate values of *a* and *b* may be found by using the normal approximation to the binomial distribution.
Without using the continuity correction for this approximation, we may obtain *a* and *b* by solving for them in the equations:

$$\frac{a - n(.5)}{\sqrt{n(.25)}} = -z_{(1-\alpha_1)}, \quad \frac{b-1-n(.5)}{\sqrt{n(.25)}} = z_{(1-\alpha_2)}$$

And rounding to the nearest integer.

If we want a 95% confidence interval and n = 40,

$$\frac{a-20}{\sqrt{10}} = -1.96, \quad \frac{b-1-20}{\sqrt{10}} = 1.96$$

Then *a* = 13.8 and *b* = 27.2. Therefore, we use *X_(14)* = 75.0 and *X_(27)* = 77.1 as the lower and upper confidence limits.

data=c(72.1, 72.8, ...
Load package 'BSDA'

1) The confidence interval is

2) $1 - \alpha$