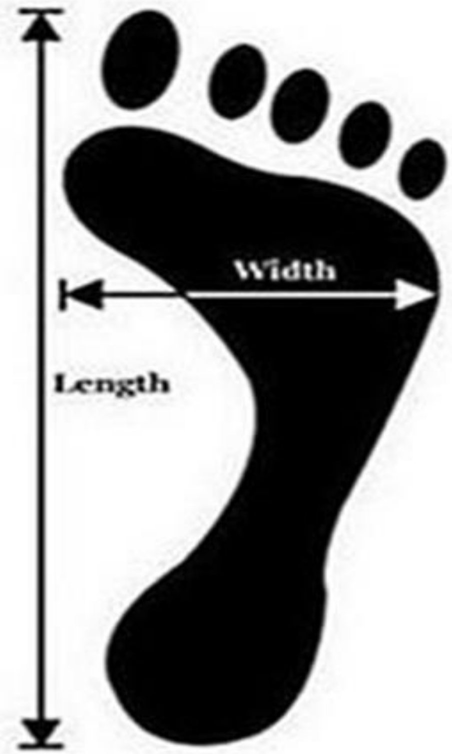
Hand and Foot Measurements



Foot measurements WITHOUT shoe!

To Do

Read Sections 4.6 - 4.7.

Do End-of-Chapter Problems 1-17 in preparation for Tutorial Test 2.

Important Ideas from Last Day

(1) Definition of a Pivotal Quantity

(2) How to Use a Pivotal Quantity to Construct a Confidence Interval

(3) Approximate Pivotal Quantities

(4) Approximate Confidence Intervals for Binomial

Interval Estimation

Interval estimation is important because it gives us a way to quantify how good our point estimate of an unknown parameter is.

We have two ways of finding interval estimates for an unknown parameter:

(1) Use a 100p% likelihood interval.

(2) Use a 100p% confidence interval if an exact pivotal quantity exists or a 100p% approximate confidence interval based on an approximate pivotal quantity (often based on a Central Limit Theorem result).

Example

Suppose *θ* is the proportion of units in a large population who have a specific characteristic.

Suppose *n* = 100 units are randomly selected and *y* = 18 units have the characteristic.

A point estimate of *θ* is

$$\hat{\theta} = \frac{18}{100} = 0.18$$

An approximate 95% confidence interval for *θ* is given by

$$0.18 \pm 1.96\sqrt{\frac{0.18(0.82)}{100}} = 0.18 \pm 0.075$$

$$\text{or } [0.105, 0.255]$$

Example Continued

The approximate 95% confidence interval for θ is $[0.105, 0.255]$.

How the media would report this:

"It is estimated that 18% of the population has the characteristic. This result is accurate to within 7.5%, 19 times out of 20."

Today's Lecture

How to Choose the Sample Size for a Binomial Experiment

How to Choose a Sample Size

We have seen that confidence intervals for a parameter get narrower as the sample size n increases.

When designing a study researchers need to choose a sample size on the basis of:

(i) how narrow they would like a confidence interval to be and

(ii) how much they can afford to spend (it costs time and money to collect data).

How to Choose a Sample Size for a Binomial Experiment

Suppose that to estimate $\boldsymbol{\theta}$, the proportion of units in a large population who have a specific characteristic, we plan to select *n* units at random.

Suppose also that we intend to use the approximate 95% confidence interval for $\boldsymbol{\theta}$ given by

$$\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

The width of this interval is

$$2(1.96)\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

How to Choose a Sample Size for a Binomial Experiment

A criterion that is widely used is to choose the sample size *n* large enough so that the width of the approximate 95% confidence interval is no wider than 2(0.03). That is, choose *n*, such that

$$(1.96)\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \le 0.03$$

or

$$n \ge \left(\frac{1.96}{0.03}\right)^2 \hat{\theta}(1-\hat{\theta})$$

Problem: What value do we use for $\hat{\theta}$ if we have not yet collected the data?

How to Choose a Sample Size for a Binomial Experiment

$$n \geq \left(\frac{1.96}{0.03}\right)^2 \hat{\theta}(1-\hat{\theta})$$

Since 0 < $\hat{\theta}$< 1, the right hand side takes on its largest value when $\hat{\theta}$ = 0.5. (Can you show this?)

So we take

$$n \geq \left(\frac{1.96}{0.03}\right)^2 (0.5)(0.5) = 1067.1$$

If we choose *n* = 1068 then the approximate 95% confidence interval for *θ* will have width less than 0.03 for all values of $\hat{\theta}$.

Polling Results and the Media

When polling results are announced in the media, you will often hear or see "this poll is accurate to within 3 percentage points 19 times out of 20."

Since 19/20 = 0.95 this really means that the estimate given is the centre of an approximate 95% confidence interval

$$\hat{\theta} \pm c$$

for which c = 0.03.

In practice, many polls are based on 1050 - 1100 people which agrees with our calculation of n = 1068.

Exercise

(1) Show that for c = 0.05 you only need n = 385 while for c = 0.02 you need n = 2401.

(2) How do these results change if you use an approximate 99% confidence interval? an approximate 90% confidence interval?