

Inferential statistics

We've seen how operational definition specifies the measurement operations that define a variable. Last week we considered how carrying out such a measurement operation assigns a number—a score; a value—to a variable. Typically one carries out not a single such operation of measurement but several—and this gives us many scores: a “distribution” of scores.

We've seen how “descriptive statistics” can be used to describe such a distribution, in particular its “central tendency” and “dispersion.”

(Notice that we're leaving to one side for a while notions of experimental design and hypothesis-testing. We'll return to these soon.)

In addition to descriptive statistics, we need to understand “inferential statistics”:

Inferential statistics provide a way of:

going from a “sample” to a “population”

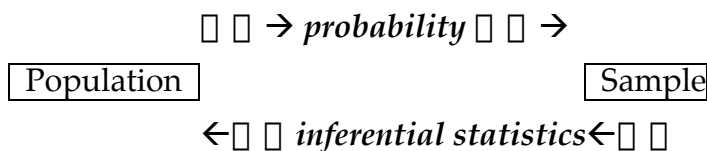
inferring the “parameters” of a population from data on the “statistics” of a sample.

i.e., **parameters** such as μ and σ , from **statistics** such as m and s .

But before we can see what is involved in the move from sample to population we need to understand how to move from population to sample.

The study of obtaining a sample from a population is “probability.”

Probability



For example: The probability of picking a black ball from jar A is one half; the probability of picking a black ball from jar B is one tenth. [This is reasoning about probability.]

jar A:	jar B:
50 black,	90 black,
50 white	10 white

[We would use Inferential statistics to answer questions such as: from which jar is one more likely to get a sample of four black balls?]

Definition:

the probability of an outcome B = $\frac{\text{number of outcomes classified as B}}{\text{total number of possible outcomes}}$

This is written $P(B)$

this is a *proportion*, a *fraction* ≤ 1

And a value of $P(B) = 0$ means the outcome is impossible;

the value $P(B) = 1$ means the outcome is certain.

BUT, for this definition of probability to be accurate, the outcomes must be selected “randomly”.

Definition: Random sample:

1. every selection has an *equal* chance.
(i.e. there is no *bias* to the selection)
2. when there is more than one selection, there must be a *constant probability* each time.
(i.e., we must be “sampling with replacement”)

...or else we'll have a situation like this after we've made our first selection:

jar A: 50 black, 50 white	jar A: 49 black; 50 white
---------------------------------	---------------------------------

Notice: “random” does not mean “chaotic.”

Rather, random means there is a pattern that becomes apparent only when we examine a large number of events. It means there is a pattern that doesn't show itself in a single case, or a few cases.

Probability is the study of the patterns of random processes.

The Normal Distribution

The kind of pattern that emerges from *many* types of random process is “the normal distribution”

See pattern emerging from a random process in a [quincunx](http://www.rand.org/methodology/stat/applets/clt.html) at:

<http://www.rand.org/methodology/stat/applets/clt.html>

There's [another](#) at:

<http://www.users.on.net/zhcchz/java/quincunx/Quincunx.html>

Read the history of how [Francis Galton](#) invented the quincunx at:

<http://www.tld.jcu.edu.au/hist/stats/galton/galton16.html>

A normal distribution is easily described....

[We'll add this in class]

Now we're ready for:

Inferential Statistics

It is usually necessary for a researcher to work with **samples** rather than a whole population.

but one **difficulty** is that a sample is **generally not identical** to the population from which it comes.

specifically, the sample mean will differ from the population mean:

$$\bar{X} \neq \mu$$

and another **difficulty** is that no two samples are the same. How can we know which best describes the population?

We need rules that relate samples to populations

"The Distribution of Sample Means"

Definition: the distribution of sample means is the collection of sample means for *all* the possible random samples of a particular size (n) that can be obtained from a population.

It is **not** a distribution of **scores**, but a distribution of **statistics**.

This distribution **tends to be normal**.

It will be almost perfectly normal if **either**:

1. the population from which the sample is drawn is normal,

or

2. the n of the sample is relatively large (30 or more).

This distribution has a mean that is equal to the population mean;

i.e. $\mu_{\bar{X}} = \mu$

and a standard deviation, $\sigma_{\bar{X}}$, which is called "the standard error of \bar{X} ."

The **standard error** is a very valuable measure. It specifies how well a sample mean estimates the population mean;

i.e. how accurate an estimate a sample provides;

i.e., the error between \bar{X} and μ , on average.

IMPORTANT NOTE: It can be shown that:

$$\sigma_{\bar{x}} = \sigma/\sqrt{n}.$$

This formula shows how it is that the accuracy of the estimate provided by a sample increases as the sample size increases. (This formula—and a related one we'll introduce shortly—is something APA says one needs to know for the licensing exam.)

Now we can turn to...

Hypothesis Testing

Definition: hypothesis testing is an inferential procedure that uses *sample* data to evaluate the credibility of a hypothesis about a *population*.

Remember?

null hypothesis (H_0): the treatment has no effect.

i.e., the experimental group and the control group are drawn from the **same population**.

This will seem confusing, since a good experiment assigns subjects randomly to the two groups. The point here is that the null hypothesis asserts that it is still the case **AFTER THE TREATMENT** that the two groups belong to the same population. If, on the other hand, the treatment *did* have an effect—which of course means that the null hypothesis is false—then the two groups would now come from **different** populations. Comprendas?

i.e., \bar{X}_{expt} and \bar{X}_{control} will not differ by more than random error.

i.e., they will not differ by more than the “standard error.”

How do we know the standard error? Remember, it's the standard deviation of the distribution of sampling means. Actually, we don't know it directly. But we can **estimate** it: our estimate replaces the population parameter σ with the sample statistic, s :

Remember we said that it can be shown that $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.

In the same way:

$$s_{\bar{x}} = s/\sqrt{n}.$$

We use the symbol $s_{\bar{x}}$ indicate that the value is calculated from sample data, rather than from the population parameter.

Now we can say, if the null hypothesis is true, the experimental group mean minus the control group mean will equal the estimated standard error, or less:

$$\text{i.e., If } H_0 \text{ is true, } \bar{X}_{\text{expt}} \text{ minus } \bar{X}_{\text{control}} \leq s_{\bar{x}}$$

$$\text{i.e., } \bar{X}_{\text{expt}} \text{ minus } \bar{X}_{\text{control}} / s_{\bar{x}} \leq 1 \quad (\text{simply rearranging the formula})$$

The left hand side of this equation is called " t ", the ***t*-statistic**.

So, if the null hypothesis is **true**, $t \leq 1$

Conversely, if the null hypothesis is **false**, $t > 1$

How much greater than 1 will t be? We just calculate that from our data. Then we use this value to look in a table of t . This table will tell us how likely it is that we could have obtained our data as a result of chance alone. That's to say, it tells us the probability of the null hypothesis being true, given our data. That means, it tells us how confident we can feel in eliminating the null hypothesis. In other words, it gives us a p value!

===

n.b., the above stuff assumes:

1. random sampling
2. that the experimental treatment doesn't change σ . (I.e., that it only changes μ . This is the assumption of "homogeneity of variance.")
3. that the distribution of sampling means is normal. (We considered this earlier.)