Lecture 7: Statistics

A random sample is a collection of the outcomes in the sample space chosen with the same probability that they occur in the ~~spaem~~ sample space. Random samples are the way to get information indicative of the population without having to ~~talk~~ look at every outcome from every person in the population.

The standard example is a survey, which will extrapolate the views of the population based on the few hundred people it asks.

Statistics studies these samples.

Example: 10 pennies are tossed 10 times. The number of heads is recorded.

results: 4, 5, 4, 5, 5, 3, 5, 4, 4, 5.

The mean is calculated by summing the results and dividing by the number of trials.

$$\frac{1}{10}(4+5+4+5+5+3+5+4+4+5) = 4.4$$

The mean is denoted by $\bar{x}$.

The median is the middle number when the outcomes are written in increasing order.

$$3, 4, 4, 4, 4, 5, 5, 5, 5, 5$$

Here, the median is 4.5. When the number of trials is even, it is necessary to average the middle two.

In the population, the median represents the number where there is an equal probability that any outcome will be greater than or less than this number.

In notation, let X be a random variable

$$p(X \leq (\text{median}(X))) = p(X > \text{median}(X))$$

The mode is the most frequent outcome. In the example, 5 5's occured, so this is the mode. Ties ~~and~~ are possible for the mode. In the population, the mode is the most probable outcome.

$$p(X = x_i) \leq p(X = \text{mode}(X))$$

for any outcome x_i.

~~It is dif~~ When the mean, median, and mode are the same, it is still possible to have different spreads in the outcomes: They could all be the same, ~~sp~~ close together, or far apart. To measure the variation in a sample, we have several tools.

The deviation from the mean is the number $x-\bar{x}$. This number changes with each trial.

There was one 3, with deviation -1.4, four 4's, with deviation $-.4$, and five 5's, with deviation $.6$.

Note that the deviations from the mean from a sample always sum to zero. The list format makes these numbers difficult to deal with.

The variance is the sum of the squares of the deviations from the mean, divided by $n-1$ (where n is the number of trials).

In our example,

$$\frac{1}{9}\left(\underbrace{3}_{\text{for the 3}}(-1.4)^2 + \underbrace{(-.4)^2}_{\text{for the 4's}} \cdot \underbrace{4}_{\text{since there are 4 of them}} + \underbrace{(.6)^2}_{\text{for the 5's}} \cdot \underbrace{5}_{\text{5 5's}} \right)$$

$= .489$

the variance is denoted by s^2.

The <u>standard</u> <u>deviation</u> is the squareroot of the variance and denoted by s.

Here, $s = .699$.

The notation for mean ($\bar{X}$), variance (s^2), and standard deviation (s), is only for these values ~~as a~~ from a random sample. For the population, the mean corresponds to the expected value and can be notated as μ. The population variance, ~~[illegible]~~, is the value $\sigma^2 = E((x-\mu)^2)$, the popution standard deviation, σ, is the squareroot of the population variance.

Example: Roll 3 dice, record the total.

10 trials: 14, 12, 13, 10, 14, 10, 7, 15, 12, 10

The <u>mean</u> is $\bar{x} = 11.7$

The <u>median</u> is 7, 10, 10, 10, 12, 12, 13, 14, 14, 15

$\boxed{12}$

The <u>mode</u> is 10. (10 occurs 3 times.)

~~The variance is~~

The <u>deviations from the mean</u> are:

	7	:	-4.7
(3)	10	:	-1.7
(2)	12	:	$.3$
	13	:	1.3
(2)	14	:	2.3
	15	:	3.3

The <u>variance</u> is

$$\frac{1}{9}\Big((-4.7)^2 + (3)(-1.7)^2 + (2)(.3)^2 + (1.3)^2 + (2)(2.3)^2 + (3.3)^2 \Big)$$

$= 6.01$

The <u>standard deviation</u> is 2.45.