U U

ECE 3530 Statistics Definitions

8 Apr
Neil E

def: samples ≡ actual outcomes of experiments, i.e. measurements
≡ x_i or observations, i.e. actual data.

ex: A company manufactures resistors. There is variation in the actual resistor values. If we measure the resistance of each resistor, we have samples from which we can draw statistical inferences.

Note: We usually assume samples are independent outcomes.

def: statistic ≡ any value that is a function of sample values.
$\equiv f(x_1, \ldots, x_n)$

ex: The average value of a set of samples is a statistic.

Roll die. Outcomes (i.e., samples) are
1, 5, 2, 6, 4, 1, 2, 3

The average value (i.e., sample mean) is
$\frac{1}{8}(1+5+2+6+4+1+2+3) = \frac{1}{8} \cdot 24 = 3.$

ex: The (sample) median value of a set of samples is a statistic.

For above die outcomes (i.e., samples) the median value is between 2 and 3. Say we choose sample median = 2.5.

def: parameter ≡ unknown value that helps to describe the true (and usually unknowable) probability density function underlying sample values.
$\equiv \mu$ or σ^2 or...

ex: The true mean, variance, and type of pdf of rolling a die are parameters.

Comment: The true μ and σ^2 of a prob density func (pdf) is almost never known. The science of statistics is directed at accurately estimating parameters like μ and σ^2.

8 Apr 03
Neil E Cotter

def: sample mean $\equiv \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\equiv$ average value of samples

~~ex~~: Ave of dice rolls (see prev pg) is 3.

def: sample variance $\equiv s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $\equiv$ average squared variation of x_i's relative to sample mean, $\bar{x}$.

Note: We use $\frac{1}{n-1}$ and not $\frac{1}{n}$ if we want an unbiased estimate (i.e., one whose expected value is exactly σ^2).

Why $\frac{1}{n-1}$? It turns out that since we are using $\bar{x}$ to estimate μ, we are using our data twice. This causes $\bar{x}$ to be closer to the x_i on average than μ would be.

pf: Here's the proof that $E\{s^2\} = \sigma^2$:

First, we observe that $E\{\bar{x}\} = E\{\frac{1}{n}\sum_{i=1}^{n} x_i\} = \frac{1}{n}\cdot n\mu = \mu$.

Second, we observe that $E\{(\bar{x}-\mu)^2\} = E\{(\bar{x} - E\{\bar{x}\})^2\}$

$$= \text{var}\{\bar{x}\} = \underbrace{\frac{1}{n^2}\sigma^2 + \ldots + \frac{1}{n^2}\sigma^2}_{n \text{ terms}} = \frac{\sigma^2}{n}$$

Third, $E\{s^2\} = E\left\{\frac{1}{n-1}\sum_{i=1}^{n}[(x_i-\mu)-(\bar{x}-\mu)]^2\right\}$

$$= E\left\{\frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i-\mu)^2 - 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu) + \sum_{i=1}^{n}(\bar{x}-\mu)^2\right]\right\}$$

U U

ECE 3530 Statistics Definitions

8 Apr
Neil E

pf: (cont.)

$$E\{s^2\} = E\left\{ \frac{1}{n-1}\left[\left(\sum_{i=1}^{n}(x_i-\mu)^2\right) - 2(\bar{x}-\mu)(\bar{x}-\mu)n + n(\bar{x}-\mu)^2\right]\right\}$$

$$= E\left\{ \frac{1}{n-1}\left[\left(\sum_{i=1}^{n}(x_i-\mu)^2\right) - n(\bar{x}-\mu)^2\right]\right\}$$

$$= \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} E\{(x_i-\mu)^2\}\right) - n\,E\{(\bar{x}-\mu)^2\}\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right]$$

term arising from $(\bar{x}-\mu)^2$ actually <u>reduces</u> the total variance of our estimate

$$= \frac{1}{n-1}(n-1)\sigma^2$$

$$= \sigma^2$$

Comment: The correlation of the x_i and $\bar{x}$ reduces the variance of s^2.

ex: sample variance of die rolls 1, 5, 2, 6, 4, 1, 2, 3 is
$x_1, \ldots, x_{n=8}$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$ where $\bar{x}$ = sample mean = 3, $n = 8$ die rolls

$$= \frac{1}{7}\left(2^2+2^2+1^2+3^2+1^2+2^2+1^2+0^2\right) = \frac{22}{7} = 3\tfrac{1}{7}$$

compare with $\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$ where $\mu = 3\frac{1}{2}$

use n if true μ available

(Note: $\sigma^2 = \frac{35}{12}$ for die)

$$= \frac{1}{8}\left(\left(\tfrac{5}{2}\right)^2+\left(\tfrac{3}{2}\right)^2+\left(\tfrac{3}{2}\right)^2+\left(\tfrac{5}{2}\right)^2+\left(\tfrac{1}{2}\right)^2+\left(\tfrac{5}{2}\right)^2+\left(\tfrac{3}{2}\right)^2+\left(\tfrac{1}{2}\right)^2\right)$$

$$= \frac{1}{32}\left(25+9+9+25+1+25+9+1\right) = \frac{104}{32} = 3\tfrac{1}{4}$$