

Definition 6.1.1: Random sample

A random sample of size n from the distribution of X is a collection of n independent random variables, each with the same distribution as X .

I det som följer ska vi låta μ och σ beteckna X 's väntevärde och standardavvikelse.

Alltså

Om vi har n oberoende observationer x_1, \dots, x_n av någon variabel x , så tänker vi oss en teoretisk motsvarighet bestående av n oberoende stokastiska variabler X_1, \dots, X_n , alla med samma fördelning som X och vi låter $\mu = E[X]$ samt $\sigma^2 = \text{Var}[X]$.

Stem-and-leaf-diagram (används knappast)

Histogram (se exempel 6.2.1, 6.2.2 (tabell 6.3, figur 6.4 och 6.5))

Cumulative distribution plots (Ogives) (tabell 6.4, figur 6.6 och 6.7)

Definition 6.3.1: Sample mean

Let X_1, X_2, \dots, X_n be a random sample from the distribution of X . The statistic

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

is called the sample mean.

Även medelvärdet av observationerna

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

kallas för stickprovsmedelvärdet. Behöver man vara tydlig kan man säga (eller skriva) det observerade medelvärdet. Man kan också säga det empiriska eller experimentella medelvärdet då man avser \bar{x} .

Definition 6.3.3: Sample variance and sample standard deviation

Let X_1, X_2, \dots, X_n be a random sample from the distribution of X . The statistic

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

is called the sample variance. Furthermore, the statistic $S = \sqrt{S^2}$ is called the sample standard deviation.

Vi kallar även

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

för stickprovsvariansen eller den observerade (uppmätta, experimentella, empiriska) variansen.

Samma terminologi används för standardavvikelsen $s = \sqrt{s^2}$.

Theorem 6.3.1: A computational formula for s^2

$$\begin{aligned} s^2 &= \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right] \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \end{aligned}$$

Definition 6.3.2

Let $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be a sample of observations arranged in order from the smallest to the largest. The sample median is the middle observation if n is odd. It is the average of the two middle observations if n is even. We shall denote the median of a sample by \tilde{x} .

Medianen delar i stickprovet i två halvor, den undre och den övre, med ca 50% av observationerna i varje. Medianen i den undre halvan har egenskapen att ca 25% av all data är mindre och ca 75% är större och kallas därför för den undre kvartilen. Den övre kvartilen definieras analogt.

Definition 6.3.4: Sample range

Let $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be a sample of observations arranged in order from the smallest to the largest. The sample range is $x_{(n)} - x_{(1)}$.

Boxplots (boxplottar, lådogram)

