Search Go DICTIONARY | CLOSE WINDOW

Formulas for descriptive statistics, probability, and random variables

Counting
- Combinations
- Permutations

Descriptive statistics
- Estimated standard deviation of grouped data
- Percentile
- Population standard deviation
- Sample standard deviation

Random variables
- Binomial random variables
- Expectation of a random variable
- Variance of a random variable

Sets
- Distributivity of the intersection over the union
- Distributivity of the union over the intersection

Probability
- Bayes theorem
- Conditional probability
- Independent events
- Law of total probability
- Mutually exclusive events
- Probability of the union of two events

Bayes theorem:

When $P(A) \neq 0$ and $P(B) \neq 0$, we have

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})}.$$

Binomial random variable:

If X is a binomial random variable with parameters n and p, and x is any integer from 0 to n, then the probability that X takes the value x is given by

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The expectation of a binomial random variable with parameters n and p is equal to np, and the standard deviation is equal to $\sqrt{np(1-p)}$.

Combinations:

The number of combinations of n objects taken r at a time (that is, the number of subsets of size r which may be chosen from a set of size n) is given by

$$ {}_nC_r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}. $$

Conditional probability:

When $P(B) \neq 0$, the conditional probability of A given B is defined to be

$$ P(A \mid B) = \frac{P(A \cap B)}{P(B)}. $$

Distributivity of the union over the intersection:

For any sets A, B, and C, we have $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Distributivity of the intersection over the union:

For any sets A, B, and C, we have $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Estimated standard deviation of grouped data:

An estimate of the sample standard deviation s of a data set of size n grouped into k classes is

$$ s \approx \sqrt{\frac{f_1(M_1-\bar{x})^2 + f_2(M_2-\bar{x})^2 + \cdots + f_k(M_k-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{k} f_i(M_i-\bar{x})^2}{n-1}} $$

,

in which $\bar{x}$ is the estimated mean of the data set, f_i is the frequency of class i, and M_i is the midpoint of class i, for $i = 1, 2, ..., k$.

Expectation of a random variable:

Let X be a random variable which takes on the values $x_1, x_2, ..., x_n$ (and only those values). The expectation of X is defined to be

$$E(X) = x_1 P(X=x_1) + x_2 P(X=x_2) + \cdots + x_n P(X=x_n) .$$

Independent events:

Events A and B are independent precisely when $P(A \cap B) = P(A)P(B)$.

Law of total probability:

For any events A and B, we have

$$P(A) = P(A \cap B) + P(A \cap \overline{B}).$$

Mutually exclusive events:

Events A and B are mutually exclusive precisely when $P(A \cap B) = 0$.

Percentile:

We show two methods for calculating a percentile:

Method A:

To find a ***p***th percentile of a sample of n observations:

1. Order the n observations from smallest to largest.
2. Determine the value of $n \times \frac{p}{100}$. This value is used to find the position of the ***p***th percentile in the ordered list.
   - If $n \times \frac{p}{100}$ is not an integer, round it up to the next integer. The observation in this

position is the ***p***th percentile.

- If $n \times \frac{p}{100}$ is an integer, calculate the mean of the observations in this position and the next (higher) position. This mean is the ***p***th percentile.

Method B:

To find a ***p***th percentile of a sample of n observations:

1. Order the n observations from smallest to largest.
2. Determine the value of $(n+1) \times \frac{p}{100}$. This value is used to find the position of the ***p***th percentile in the ordered list.
   - If $(n+1) \times \frac{p}{100}$ is an integer, the observation in this position is a ***p***th percentile.
   - If $(n+1) \times \frac{p}{100}$ is not an integer but equal to *k* plus some proper fraction *d*, the ***p***th percentile is a value lying *d* of the distance between the observations in the *k*th and *(k+1)*th positions.

Permutations:

The number of permutations of n objects taken r at a time (that is, the number of ordered arrangements of size r which may be chosen from a set of size n) is given by

$${}_nP_r = \frac{n!}{(n-r)!}.$$

Population standard deviation:

The standard deviation of a population of n numbers $x_1, x_2, ..., x_n$ is defined to be

$$\sigma = \sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+....+(x_n-\mu)^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}},$$

in which μ is the population mean.

Probability of the union of two events:

For any events A and B, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Sample standard deviation:

The standard deviation of a sample of n numbers $x_1, x_2, ..., x_n$ is defined to be

$$s = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}},$$

in which $\overline{x}$ is the sample mean.

Variance of a random variable:

The variance of the random variable X may be found from the calculation

$$Var(X) = E(X^2) - [E(X)]^2,$$

in which $E(X)$ is the expectation of X.