Let $X_1, \dots, X_n$ be a sample. The **sample mean** is

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

Let $X_1, \dots, X_n$ be a sample. The **sample variance** is the quantity

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

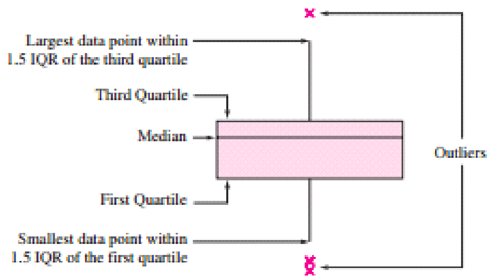
Let $X_1, \dots, X_n$ be a sample. The **sample standard deviation** is the quantit

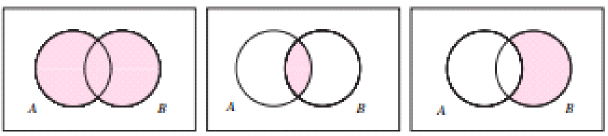
$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

- If $X_1, \dots, X_n$ is a sample and $Y_i = a + bX_i$, where $a$ and $b$ are constants, then $\overline{Y} = a + b\overline{X}$.
- If $X_1, \dots, X_n$ is a sample and $Y_i = a + bX_i$, where $a$ and $b$ are constants, then $s_Y^2 = b^2 s_X^2$, and $s_Y = |b| s_X$.

If $n$ numbers are ordered from smallest to largest:

- If $n$ is odd, the sample median is the number in position $\frac{n+1}{2}$.
- If $n$ is even, the sample median is the average of the numbers in positions $\frac{n}{2}$ and $\frac{n}{2} + 1$.




(a) (b) (c)

(a) $A \cup B$, (b) $A \cap B$, (c) $B \cap A^c$.

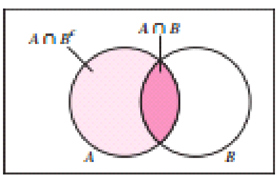
- The events $A$ and $B$ are said to be **mutually exclusive** if they have no outcomes in common.
- More generally, a collection of events $A_1, A_2, \dots, A_n$ is said to be mutually exclusive if no two of them have any outcomes in common.

**The Axioms of Probability**

1. Let $\mathcal{S}$ be a sample space. Then $P(\mathcal{S}) = 1$.
2. For any event $A$, $0 \le P(A) \le 1$.
3. If $A$ and $B$ are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$. More generally, if $A_1, A_2, \dots$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$.

Let $A$ and $B$ be any events. Then

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




The number of permutations of $n$ objects is $n!$.

The number of permutations of $k$ objects chosen from a group of $n$ objects is

$$\frac{n!}{(n-k)!}$$

The number of combinations of $k$ objects chosen from a group of $n$ objec

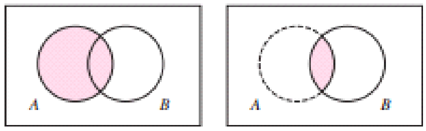
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The number of ways of dividing a group of $n$ objects into groups of $k_1, \dots, k_r$ objects, where $k_1 + \cdots + k_r = n$, is

$$\frac{n!}{k_1! \cdots k_r!} \tag{2.13}$$

Let $A$ and $B$ be events with $P(B) \neq 0$. The conditional probability of $A$ given $B$ is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.14}$$




(a) (b)

Two events $A$ and $B$ are **independent** if the probability of each event remains the same whether or not the other occurs.

*In symbols:* If $P(A) \neq 0$ and $P(B) \neq 0$, then $A$ and $B$ are independent if

$$P(B|A) = P(B) \quad \text{or, equivalently,} \quad P(A|B) = P(A) \tag{2.15}$$

If either $P(A) = 0$ or $P(B) = 0$, then $A$ and $B$ are independent.

Events $A_1, A_2, \dots, A_n$ are independent if the probability of each remains the same no matter which of the others occur.

*In symbols:* Events $A_1, A_2, \dots, A_n$ are independent if for each $A_i$, and each collection $A_{j1}, \dots, A_{jm}$ of events with $P(A_{j1} \cap \cdots \cap A_{jm}) \neq 0$,

$$P(A_i | A_{j1} \cap \cdots \cap A_{jm}) = P(A_i) \tag{2.16}$$

If $A$ and $B$ are two events with $P(B) \neq 0$, then

$$P(A \cap B) = P(B)P(A|B) \tag{2.17}$$

If $A$ and $B$ are two events with $P(A) \neq 0$, then

$$P(A \cap B) = P(A)P(B|A) \tag{2.18}$$

If $P(A) \neq 0$ and $P(B) \neq 0$, then Equations (2.17) and (2.18) both hold.

If $A$ and $B$ are independent events, then

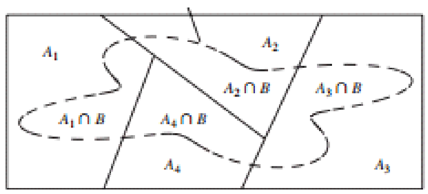
$$P(A \cap B) = P(A)P(B) \tag{2.19}$$

This result can be extended to any number of events. If $A_1, A_2, \dots, A_n$ are independent events, then for each collection $A_{j1}, \dots, A_{jm}$ of events

$$P(A_{j1} \cap A_{j2} \cap \cdots \cap A_{jm}) = P(A_{j1})P(A_{j2}) \cdots P(A_{jm}) \tag{2.20}$$

In particular,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n) \tag{2.21}$$




**Law of Total Probability**

If $A_1, \dots, A_n$ are mutually exclusive and exhaustive events, and $B$ is any event, then

$$P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B) \tag{2.23}$$

Equivalently, if $P(A_i) \neq 0$ for each $A_i$,

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n) \tag{2.24}$$

**Bayes' Rule**

**Special Case:** Let $A$ and $B$ be events with $P(A) \neq 0$, $P(A^c) \neq 0$, and $P(B) \neq 0$. Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \tag{2.27}$$

**General Case:** Let $A_1, \ldots, A_n$ be mutually exclusive and exhaustive events with $P(A_i) \neq 0$ for each $A_i$. Let $B$ be any event with $P(B) \neq 0$. Then

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} \tag{2.28}$$

Let $X$ be a discrete random variable. Then

- The probability mass function of $X$ is the function $p(x) = P(X = x)$.
- The cumulative distribution function of $X$ is the function $F(x) = P(X \leq x)$.
- $F(x) = \sum_{t \leq x} p(t) = \sum_{t \leq x} P(X = t)$.
- $\sum_x p(x) = \sum_x P(X = x) = 1$, where the sum is over all the possible values of $X$.

Let $X$ be a discrete random variable with probability mass function $p(x) = P(X = x)$.

The mean of $X$ is given by

$$\mu_X = \sum_x x P(X = x) \tag{2.29}$$

where the sum is over all possible values of $X$.

The mean of $X$ is sometimes called the expectation, or expected value, of $X$ and may also be denoted by $E(X)$ or by $\mu$.

Let $X$ be a discrete random variable with probability mass function $p(x) = P(X = x)$. Then

- The variance of $X$ is given by

$$\sigma_X^2 = \sum_x (x - \mu_X)^2 P(X = x) \tag{2.30}$$

- An alternate formula for the variance is given by

$$\sigma_X^2 = \sum_x x^2 P(X = x) - \mu_X^2 \tag{2.31}$$

- The variance of $X$ may also be denoted by $V(X)$ or by $\sigma^2$.
- The standard deviation is the square root of the variance: $\sigma_X = \sqrt{\sigma_X^2}$.

Let $X$ be a continuous random variable with probability density function $f(x)$. Let $a$ and $b$ be any two numbers, with $a < b$. Then

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) = \int_a^b f(x)\,dx$$

In addition,

$$P(X \leq b) = P(X < b) = \int_{-\infty}^{b} f(x)\,dx \tag{2.32}$$

$$P(X \geq a) = P(X > a) = \int_a^{\infty} f(x)\,dx \tag{2.33}$$

Let $X$ be a continuous random variable with probability density function $f(x)$. Then

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Let $X$ be a continuous random variable with probability density function $f(x)$. The cumulative distribution function of $X$ is the function

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt \tag{2.34}$$

Let $X$ be a continuous random variable with probability density function $f(x)$. Then the mean of $X$ is given by

$$\mu_X = \int_{-\infty}^{\infty} x f(x)\,dx \tag{2.35}$$

The mean of $X$ is sometimes called the expectation, or expected value, of $X$ and may also be denoted by $E(X)$ or by $\mu$.

Let $X$ be a continuous random variable with probability density function $f(x)$. Then

- The variance of $X$ is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\,dx \tag{2.36}$$

- An alternate formula for the variance is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu_X^2 \tag{2.37}$$

- The variance of $X$ may also be denoted by $V(X)$ or by $\sigma^2$.
- The standard deviation is the square root of the variance: $\sigma_X = \sqrt{\sigma_X^2}$.

Let $X$ be a continuous random variable with probability density function $f(x)$. Then

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Let $X$ be a continuous random variable with probability density function $f(x)$. The cumulative distribution function of $X$ is the function

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt \tag{2.34}$$

Let $X$ be a continuous random variable with probability density function $f(x)$. Then the mean of $X$ is given by

$$\mu_X = \int_{-\infty}^{\infty} x f(x)\,dx \tag{2.35}$$

The mean of $X$ is sometimes called the expectation, or expected value, of $X$ and may also be denoted by $E(X)$ or by $\mu$.

Let $X$ be a continuous random variable with probability density function $f(x)$. Then

- The variance of $X$ is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\,dx \tag{2.36}$$

- An alternate formula for the variance is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu_X^2 \tag{2.37}$$

- The variance of $X$ may also be denoted by $V(X)$ or by $\sigma^2$.
- The standard deviation is the square root of the variance: $\sigma_X = \sqrt{\sigma_X^2}$.

Let $X$ be a continuous random variable with probability mass function $f(x)$ and cumulative distribution function $F(x)$.

- The median of $X$ is the point $x_m$ that solves the equation $F(x_m) = P(X \leq x_m) = \int_{-\infty}^{x_m} f(x)\,dx = 0.5$.
- If $p$ is any number between 0 and 100, the $p$th percentile is the point $x_p$ that solves the equation $F(x_p) = P(X \leq x_p) = \int_{-\infty}^{x_p} f(x)\,dx = p/100$.
- The median is the 50th percentile.

**Chebyshev's Inequality**

Let $X$ be a random variable with mean $\mu_X$ and standard deviation $\sigma_X$. Then

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}$$

If $X$ is a random variable and $a$ is a constant, then

$$\mu_{aX} = a\mu_X$$

If $X$ is a random variable and $a$ is a constant, then

$$\sigma_{aX}^2 = a^2\sigma_X^2$$

$$\sigma_{aX} = |a|\sigma_X$$

If $X$ is a random variable, and $a$ and $b$ are constants, then

$$\mu_{aX+b} = a\mu_X + b$$

$$\sigma_{aX+b}^2 = a^2\sigma_X^2$$

$$\sigma_{aX+b} = |a|\sigma_X$$

If $X_1, X_2, \ldots, X_n$ are random variables, then the mean of the sum $X_1 + X_2 + \cdots + X_n$ is given by

$$\mu_{X_1+X_2+\cdots+X_n} = \mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_n} \tag{2.47}$$

If $X$ and $Y$ are random variables, and $a$ and $b$ are constants, then

$$\mu_{aX+bY} = \mu_{aX} + \mu_{bY} = a\mu_X + b\mu_Y \tag{2.48}$$

More generally, if $X_1, X_2, \ldots, X_n$ are random variables and $c_1, c_2, \ldots, c_n$ are constants, then the mean of the linear combination $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ is given by

$$\mu_{c_1X_1+c_2X_2+\cdots+c_nX_n} = c_1\mu_{X_1} + c_2\mu_{X_2} + \cdots + c_n\mu_{X_n} \tag{2.49}$$

If $X$ and $Y$ are **independent** random variables, and $S$ and $T$ are sets of numbers, then

$$P(X \in S \text{ and } Y \in T) = P(X \in S)P(Y \in T) \tag{2.50}$$

More generally, if $X_1, \ldots, X_n$ are independent random variables, and $S_1, \ldots, S_n$ are sets, then

$$P(X_1 \in S_1 \text{ and } X_2 \in S_2 \text{ and} \cdots \text{ and } X_n \in S_n) = P(X_1 \in S_1)P(X_2 \in S_2)\cdots P(X_n \in S_n) \tag{2.51}$$

If $X_1, X_2, \ldots, X_n$ are *independent* random variables and $c_1, c_2, \ldots, c_n$ are constants, then the variance of the linear combination $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ is given by

$$\sigma^2_{c_1X_1+c_2X_2+\cdots+c_nX_n} = c_1^2\sigma^2_{X_1} + c_2^2\sigma^2_{X_2} + \cdots + c_n^2\sigma^2_{X_n} \tag{2.53}$$

If $X$ and $Y$ are *independent* random variables with variances $\sigma_X^2$ and $\sigma_Y^2$, then the variance of the sum $X + Y$ is

$$\sigma^2_{X+Y} = \sigma_X^2 + \sigma_Y^2 \tag{2.54}$$

The variance of the difference $X - Y$ is

$$\sigma^2_{X-Y} = \sigma_X^2 + \sigma_Y^2 \tag{2.55}$$

If $X_1, \ldots, X_n$ is a simple random sample from a population with mean $\mu$ and variance $\sigma^2$, then the sample mean $\overline{X}$ is a random variable with

$$\mu_{\overline{X}} = \mu \tag{2.56}$$

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \tag{2.57}$$

The standard deviation of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \tag{2.58}$$

If $X$ and $Y$ are jointly discrete random variables:

- The joint probability mass function of $X$ and $Y$ is the function

$$p(x, y) = P(X = x \text{ and } Y = y)$$

- The marginal probability mass functions of $X$ and of $Y$ can be obtained from the joint probability mass function as follows:

$$p_X(x) = P(X = x) = \sum_y p(x, y) \qquad p_Y(y) = P(Y = y) = \sum_x p(x, y)$$

where the sums are taken over all the possible values of $Y$ and of $X$, respectively.

- The joint probability mass function has the property that

$$\sum_x \sum_y p(x, y) = 1$$

where the sum is taken over all the possible values of $X$ and $Y$.

If $X$ and $Y$ are jointly continuous random variables, with joint probability density function $f(x, y)$, and $a < b$, $c < d$, then

$$P(a \le X \le b \text{ and } c \le Y \le d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$$

The joint probability density function has the following properties:

$$f(x, y) \ge 0 \text{ for all } x \text{ and } y$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 1$$

If $X$ and $Y$ are jointly continuous with joint probability density function $f(x, y)$, then the marginal probability density functions of $X$ and of $Y$ are given, respectively, by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

Let $X$ be a random variable, and let $h(X)$ be a function of $X$. Then

- If $X$ is discrete with probability mass function $p(x)$, the mean of $h(X)$ is given by

$$\mu_{h(X)} = \sum_x h(x)p(x) \tag{2.59}$$

where the sum is taken over all the possible values of $X$.

- If $X$ is continuous with probability density function $f(x)$, the mean of $h(X)$ is given by

$$\mu_{h(X)} = \int_{-\infty}^{\infty} h(x)f(x)\, dx \tag{2.60}$$

If $X$ is a random variable, and $a$ and $b$ are constants, then

$$\mu_{aX+b} = a\mu_X + b \tag{2.61}$$

$$\sigma^2_{aX+b} = a^2\sigma_X^2 \tag{2.62}$$

$$\sigma_{aX+b} = |a|\sigma_X \tag{2.63}$$

If $X$ and $Y$ are jointly distributed random variables, and $h(X, Y)$ is a function of $X$ and $Y$, then

- If $X$ and $Y$ are jointly discrete with joint probability mass function $p(x, y)$,

$$\mu_{h(X,Y)} = \sum_x \sum_y h(x, y)p(x, y) \tag{2.64}$$

where the sum is taken over all the possible values of $X$ and $Y$.

- If $X$ and $Y$ are jointly continuous with joint probability density function $f(x, y)$,

$$\mu_{h(X,Y)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)f(x, y)\, dx\, dy \tag{2.65}$$

Let $X$ and $Y$ be jointly discrete random variables, with joint probability mass function $p(x, y)$. Let $p_X(x)$ denote the marginal probability mass function of $X$ and let $x$ be any number for which $p_X(x) > 0$.

The conditional probability mass function of $Y$ given $X = x$ is

$$p_{Y|X}(y \mid x) = \frac{p(x, y)}{p_X(x)} \tag{2.66}$$

Note that for any particular values of $x$ and $y$, the value of $p_{Y|X}(y \mid x)$ is just the conditional probability $P(Y = y \mid X = x)$.

The conditional probability density function of $Y$ given $X = x$ is

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \tag{2.67}$$

If $X$ and $Y$ are independent random variables, then

- If $X$ and $Y$ are jointly discrete, and $x$ is a value for which $p_X(x) > 0$, then

$$p_{Y|X}(y \mid x) = p_Y(y)$$

- If $X$ and $Y$ are jointly continuous, and $x$ is a value for which $f_X(x) > 0$, then

$$f_{Y|X}(y \mid x) = f_Y(y)$$

Let $X$ and $Y$ be random variables with means $\mu_X$ and $\mu_Y$. The covariance of $X$ and $Y$ is

$$\text{Cov}(X, Y) = \mu_{(X-\mu_X)(Y-\mu_Y)} \qquad \text{Cov}(X, Y) = \mu_{XY} - \mu_X\mu_Y$$

Let $X$ and $Y$ be jointly distributed random variables with standard deviations $\sigma_X$ and $\sigma_Y$. The correlation between $X$ and $Y$ is denoted $\rho_{X,Y}$ and is given by

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X\sigma_Y} \tag{2.70}$$

For any two random variables $X$ and $Y$:

$$-1 \le \rho_{X,Y} \le 1$$

For any random variable $X$, $\text{Cov}(X, X) = \sigma_X^2$ and $\rho_{X,X} = 1$.

If $X_1, \ldots, X_n$ are random variables and $c_1, \ldots, c_n$ are constants, then

$$\mu_{c_1X_1+\cdots+c_nX_n} = c_1\mu_{X_1} + \cdots + c_n\mu_{X_n} \tag{2.71}$$

$$\sigma^2_{c_1X_1+\cdots+c_nX_n} = c_1^2\sigma^2_{X_1} + \cdots + c_n^2\sigma^2_{X_n} + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} c_ic_j \,\text{Cov}(X_i, X_j) \tag{2.72}$$

$$\sigma^2_{X+Y} = \sigma_X^2 + \sigma_Y^2 + 2\,\text{Cov}(X, Y)$$

$$\sigma^2_{X-Y} = \sigma_X^2 + \sigma_Y^2 - 2\,\text{Cov}(X, Y)$$

If $X_1, \dots, X_n$ are *independent* random variables and $c_1, \dots, c_n$ are constants, then

$$\sigma^2_{c_1X_1+\cdots+c_nX_n} = c_1^2\sigma^2_{X_1} + \cdots + c_n^2\sigma^2_{X_n} \tag{2.73}$$

In particular,

$$\sigma^2_{X_1+\cdots+X_n} = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n} \tag{2.74}$$

$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$

$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$

If $X_1, \dots, X_n$ is a simple random sample from a population with mean $\mu$ and variance $\sigma^2$, then the sample mean $\overline{X}$ is a random variable with

$$\mu_{\overline{X}} = \mu \tag{2.79}$$

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \tag{2.80}$$

The standard deviation of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \tag{2.81}$$

If $X$ is a measurement and $c$ is a constant, then

$$\sigma_{cX} = |c|\sigma_X \tag{3.3}$$

If $X_1, \dots, X_n$ are independent measurements and $c_1, \dots, c_n$ are constants, then

$$\sigma_{c_1X_1+\cdots+c_nX_n} = \sqrt{c_1^2\sigma^2_{X_1} + \cdots + c_n^2\sigma^2_{X_n}} \tag{3.4}$$

If $X_1, \dots, X_n$ are $n$ *independent* measurements, each with mean $\mu$ and uncertainty $\sigma$, then the sample mean $\overline{X}$ is a measurement with mean

$$\mu_{\overline{X}} = \mu \tag{3.5}$$

and with uncertainty

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \tag{3.6}$$

If $X$ is a measurement whose uncertainty $\sigma_X$ is small, and if $U$ is a function of $X$, then

$$\sigma_U \approx \left|\frac{dU}{dX}\right|\sigma_X \tag{3.10}$$

In practice, we evaluate the derivative $dU/dX$ at the observed measurement $X$.

If $X_1, X_2, \dots, X_n$ are *independent* measurements whose uncertainties $\sigma_{X_1}, \sigma_{X_2}, \dots, \sigma_{X_n}$ are small, and if $U = U(X_1, X_2, \dots, X_n)$ is a function of $X_1, X_2, \dots, X_n$, then

$$\sigma_U \approx \sqrt{\left(\frac{\partial U}{\partial X_1}\right)^2\sigma^2_{X_1} + \left(\frac{\partial U}{\partial X_2}\right)^2\sigma^2_{X_2} + \cdots + \left(\frac{\partial U}{\partial X_n}\right)^2\sigma^2_{X_n}} \tag{3.12}$$

In practice, we evaluate the partial derivatives at the point $(X_1, X_2, \dots, X_n)$.

If $X_1, X_2, \dots, X_n$ are measurements whose uncertainties $\sigma_{X_1}, \sigma_{X_2}, \dots, \sigma_{X_n}$ are small, and if $U = U(X_1, X_2, \dots, X_n)$ is a function of $(X_1, X_2, \dots, X_n)$, then a conservative estimate of $\sigma_U$ is given by

$$\sigma_U \le \left|\frac{\partial U}{\partial X_1}\right|\sigma_{X_1} + \left|\frac{\partial U}{\partial X_2}\right|\sigma_{X_2} + \cdots + \left|\frac{\partial U}{\partial X_n}\right|\sigma_{X_n} \tag{3.13}$$

In practice, we evaluate the partial derivatives at the point $(X_1, X_2, \dots, X_n)$. The inequality (3.13) is valid in almost all practical situations; in principle it can fail if some of the second partial derivatives of $U$ are quite large.

If $X_1, \dots, X_n$ are measurements whose relative uncertainties are $\sigma_{X_1}/X_1, \dots, \sigma_{X_n}/X_n$, and $U = X_1^{m_1} \cdots X_n^{m_n}$, where $m_1, \dots, m_n$ are any exponents, then the relative uncertainty in $U$ is

$$\frac{\sigma_U}{U} = \sqrt{\left(m_1\frac{\sigma_{X_1}}{X_1}\right)^2 + \cdots + \left(m_n\frac{\sigma_{X_n}}{X_n}\right)^2} \tag{3.14}$$

If $X \sim \text{Bernoulli}(p)$, then

$\mu_X = p$

$\sigma^2_X = p(1-p)$

If $X \sim \text{Bin}(n, p)$, then

$\mu_X = np$

$\sigma^2_X = np(1-p)$

If a total of $n$ Bernoulli trials are conducted, and

- The trials are independent
- Each trial has the same success probability $p$
- $X$ is the number of successes in the $n$ trials

then $X$ has the binomial distribution with parameters $n$ and $p$, denoted $X \sim \text{Bin}(n, p)$.

If $X \sim \text{Bin}(n, p)$, the probability mass function of $X$ is

$$p(x) = P(X = x) = \begin{cases} \dfrac{n!}{x!(n-x)!}p^x(1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

If $X \sim \text{Bin}(n, p)$, then the sample proportion $\widehat{p} = X/n$ is used to estimate success probability $p$.

- $\widehat{p}$ is unbiased.
- The uncertainty in $\widehat{p}$ is

$$\sigma_{\widehat{p}} = \sqrt{\frac{p(1-p)}{n}} \quad (4$$

In practice, when computing $\sigma_{\widehat{p}}$, we substitute $\widehat{p}$ for $p$, since $p$ is unknown.

If $X \sim \text{Poisson}(\lambda)$, then

- $X$ is a discrete random variable whose possible values are the non-negative integers.
- The parameter $\lambda$ is a positive constant.
- The probability mass function of $X$ is

$$p(x) = P(X = x) = \begin{cases} e^{-\lambda}\dfrac{\lambda^x}{x!} & \text{if } x \text{ is a non-negative integer} \\ 0 & \text{otherwise} \end{cases}$$

- The Poisson probability mass function is very close to the binomial probability mass function when $n$ is large, $p$ is small, and $\lambda = np$.

If $X \sim \text{Poisson}(\lambda)$, then

$\mu_X = \lambda$

$\sigma^2_X = \lambda$

we estimate the rate $\lambda$ with $\widehat{\lambda} = \dfrac{X}{t}$.

The uncertainty in $\widehat{\lambda}$ is $\sigma_{\widehat{\lambda}} = \sqrt{\dfrac{\lambda}{t}}$

$X \sim \text{H}(N, R, n)$.

$$p(x) = \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}}$$

If $X \sim \text{H}(N, R, n)$, then

$$\mu_X = \frac{nR}{N}$$

$$\sigma^2_X = n\left(\frac{R}{N}\right)\left(1 - \frac{R}{N}\right)\left(\frac{N-n}{N-1}\right)$$

If $X \sim \text{Geom}(p)$, then

$$p(x) = p(1-p)^{x-1} \qquad x = 1, 2, \dots$$

$$\mu_X = \frac{1}{p} \qquad \sigma^2_X = \frac{1-p}{p^2}$$

If $X \sim \text{NB}(r, p)$, then

$$p(x) = \binom{x-1}{r-1}p^r(1-p)^{x-r} \qquad x = r, r+1, \dots$$

$$\mu_X = \frac{r}{p} \qquad \sigma^2_X = \frac{r(1-p)}{p^2}$$

If $X \sim \text{Exp}(\lambda)$

**PDF**

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

**CDF**

$$F(x) = P(X \le x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

$$\mu_X = \frac{1}{\lambda} \qquad \sigma^2_X = \frac{1}{\lambda^2}$$

If $X \sim N(\mu, \sigma^2)$, then

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{(x-\mu)^2/(2\sigma^2)} \qquad \mu_X = \mu \qquad \sigma^2_X = \sigma^2 \qquad z = \frac{x-\mu}{\sigma}$$

Let $X \sim N(\mu, \sigma^2)$, and let $a \ne 0$ and $b$ be constants. Then

$$aX + b \sim N(a\mu + b, a^2\sigma^2). \tag{4.25}$$

Let $X_1, \dots, X_n$ be independent and normally distributed with mean $\mu$ and variance $\sigma^2$. Then

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \tag{4.27}$$

Let $X$ and $Y$ be independent, with $X \sim N(\mu_X, \sigma^2_X)$ and $Y \sim N(\mu_Y, \sigma^2_Y)$. Then

$$X + Y \sim N(\mu_X + \mu_Y, \sigma^2_X + \sigma^2_Y) \tag{4.28}$$

$$X - Y \sim N(\mu_X - \mu_Y, \sigma^2_X + \sigma^2_Y) \tag{4.29}$$

**The Central Limit Theorem**

Let $X_1, \ldots, X_n$ be a simple random sample from a population with mean $\mu$ and variance $\sigma^2$.

Let $\overline{X} = \dfrac{X_1 + \cdots + X_n}{n}$ be the sample mean.

Let $S_n = X_1 + \cdots + X_n$ be the sum of the sample observations.

Then if $n$ is sufficiently large,

$$\overline{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right) \quad \text{approximately} \tag{4.55}$$

and

$$S_n \sim N(n\mu,\ n\sigma^2) \quad \text{approximately} \tag{4.56}$$

If $X \sim \text{Bin}(n,\ p)$, and if $np > 10$ and $n(1-p) > 10$, then

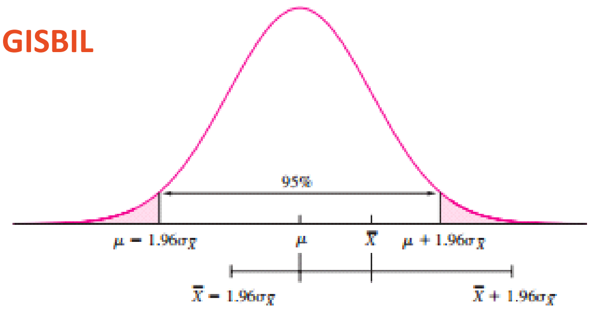
$$X \sim N(np,\ np(1-p)) \quad \text{approximately}$$

$$\widehat{p} \sim N\left(p,\ \frac{p(1-p)}{n}\right) \quad \text{approximately}$$

If $X \sim \text{Poisson}(\lambda)$, where $\lambda > 10$, then

$$X \sim N(\lambda,\ \lambda) \quad \text{approximately}$$

## ÖRYGGISBIL




Let $X_1, \ldots, X_n$ be a *large* ($n > 30$) random sample from a population with mean $\mu$ and standard deviation $\sigma$, so that $\overline{X}$ is approximately normal. Then a level $100(1-\alpha)\%$ confidence interval for $\mu$ is

$$\overline{X} \pm z_{\alpha/2}\sigma_{\overline{X}} \tag{5.1}$$

where $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. When the value of $\sigma$ is unknown, it can be replaced with the sample standard deviation $s$.

Let $X_1, \ldots, X_n$ be a *large* ($n > 30$) random sample from a population with mean $\mu$ and standard deviation $\sigma$, so that $\overline{X}$ is approximately normal. Then level $100(1-\alpha)\%$ lower confidence bound for $\mu$ is

$$\overline{X} - z_\alpha\sigma_{\overline{X}} \tag{5.2}$$

and level $100(1-\alpha)\%$ upper confidence bound for $\mu$ is

$$\overline{X} + z_\alpha\sigma_{\overline{X}} \tag{5.3}$$

where $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. When the value of $\sigma$ is unknown, it can be replaced with the sample standard deviation $s$.

Let $X$ be the number of successes in $n$ independent Bernoulli trials with success probability $p$, so that $X \sim \text{Bin}(n,\ p)$.

Define $\tilde{n} = n + 4$, and $\tilde{p} = \dfrac{X+2}{\tilde{n}}$. Then a level $100(1-\alpha)\%$ confidence interval for $p$ is

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \tag{5.5}$$

If the lower limit is less than 0, replace it with 0. If the upper limit is greater than 1, replace it with 1.

**T-dreifingu** = $\dfrac{\overline{X} - \mu}{s/\sqrt{n}}$ Oryggisbil = $\overline{X} \pm t_{n-1,\alpha/2}\dfrac{s}{\sqrt{n}}$

**Use $z$, Not $t$, If $\sigma$ Is Known**

Let $X$ and $Y$ be independent, with $X \sim N(\mu_X,\ \sigma_X^2)$ and $Y \sim N(\mu_Y,\ \sigma_Y^2)$. Then

Stór úrtak:

$$X + Y \sim N(\mu_X + \mu_Y,\ \sigma_X^2 + \sigma_Y^2) \tag{5.14}$$

$$X - Y \sim N(\mu_X - \mu_Y,\ \sigma_X^2 + \sigma_Y^2) \tag{5.15}$$

Öryggisbil: $\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}$

Hlutföll: a level $100(1-\alpha)\%$ confidence interval for the difference $p_X - p_Y$ is

$$\tilde{p}_X - \tilde{p}_Y \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}_X(1-\tilde{p}_X)}{\tilde{n}_X} + \frac{\tilde{p}_Y(1-\tilde{p}_Y)}{\tilde{n}_Y}} \tag{5.18}$$

Lítil úrtak: $\dfrac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}}$ $\quad \nu = \dfrac{\left(\dfrac{s_X^2}{n_X} + \dfrac{s_Y^2}{n_Y}\right)^2}{\dfrac{(s_X^2/n_X)^2}{n_X - 1} + \dfrac{(s_Y^2/n_Y)^2}{n_Y - 1}}$

Let $X_1, \ldots, X_{n_X}$ be a random sample of size $n_X$ from a *normal* population with mean $\mu_X$, and let $Y_1, \ldots, Y_{n_Y}$ be a random sample of size $n_Y$ from a *normal* population with mean $\mu_Y$. Assume the two samples are independent.

If the populations are known to have nearly the same variance, a level $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is

$$\overline{X} - \overline{Y} \pm t_{n_X+n_Y-2,\alpha/2} \cdot s_p\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \tag{5.22}$$

The quantity $s_p$ is the pooled standard deviation, given by

$$s_p = \sqrt{\frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}} \tag{5.23}$$

Let $D_1, \ldots, D_n$ be a *small* random sample ($n \le 30$) of differences of pairs. If the population of differences is approximately normal, then a level $100(1-\alpha)\%$ confidence interval for the mean difference $\mu_D$ is given by

$$\overline{D} \pm t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}} \tag{5.24}$$

where $s_D$ is the sample standard deviation of $D_1, \ldots, D_n$. Note that this interval is the same as that given by expression (5.9).

If the sample size is large, a level $100(1-\alpha)\%$ confidence interval for the mean difference $\mu_D$ is given by

$$\overline{D} \pm z_{\alpha/2}\sigma_{\overline{D}} \tag{5.25}$$

In practice $\sigma_{\overline{D}}$ is approximated with $s_D/\sqrt{n}$. Note that this interval is the same as that given by expression (5.1).

Let $X_1, \ldots, X_n$ be a sample from a *normal* population. Let $Y$ be another item to be sampled from this population, whose value has not been observed. A $100(1-\alpha)\%$ prediction interval for $Y$ is

$$\overline{X} \pm t_{n-1,\alpha/2}\, s\sqrt{1 + \frac{1}{n}} \tag{5.26}$$

The probability is $1 - \alpha$ that the value of $Y$ will be contained in this interval.

**Steps in Performing a Hypothesis Test**

1. Define $H_0$ and $H_1$.
2. Assume $H_0$ to be true.
3. Compute a **test statistic**. A test statistic is a statistic that is used to assess the strength of the evidence against $H_0$.
4. Compute the $P$-value of the test statistic. The $P$-value is the probability, assuming $H_0$ to be true, that the test statistic would have a value whose disagreement with $H_0$ is as great as or greater than that actually observed. The $P$-value is also called the **observed signi cance level**.
5. State a conclusion about the strength of the evidence against $H_0$.

**Summary**

Let $X_1, \ldots, X_n$ be a *large* (e.g., $n > 30$) sample from a population with mean $\mu$ and standard deviation $\sigma$.

To test a null hypothesis of the form $H_0: \mu \le \mu_0$, $H_0: \mu \ge \mu_0$, or $H_0: \mu = \mu_0$:

- Compute the $z$-score: $z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$.
  If $\sigma$ is unknown it may be approximated with $s$.
- Compute the $P$-value. The $P$-value is an area under the normal curve, which depends on the alternate hypothesis as follows:

| **Alternate Hypothesis** | ***P*-value** |
|---|---|
| $H_1: \mu > \mu_0$ | Area to the right of $z$ |
| $H_1: \mu < \mu_0$ | Area to the left of $z$ |
| $H_1: \mu \ne \mu_0$ | Sum of the areas in the tails cut off by $z$ and $-z$ |