Formula Sheet

Definition

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

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (1.1)$$

Definition

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

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad (1.2)$$

An equivalent formula, which can be easier to compute, is

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right) \qquad (1.3)$$

Definition

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

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} \qquad (1.4)$$

An equivalent formula, which can be easier to compute, is

$$s = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)} \qquad (1.5)$$

The sample standard deviation is the square root of the sample variance.

Definition

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 $\frac{n}{2}+1$.

Chapter 2

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 $A_1, A_2, \ldots$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$.

For any event A,

$$P(A^c) = 1 - P(A) \qquad (2.1)$$

Let $\emptyset$ denote the empty set. Then

$$P(\emptyset) = 0 \qquad (2.2)$$

Summary

Let A and B be any events. Then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (2.5)$$

Definition

For any positive integer n, $n! = n(n-1)(n-2)\cdots(3)(2)(1)$.

Also, we define $0! = 1$.

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

Summary

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

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

Summary

The number of combinations of k objects chosen from a group of n objects is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \qquad (2.12)$$

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

$$\frac{n!}{k_1! \cdots k_r!} \qquad (2.13)$$

Definition

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

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

Definition

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

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

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

Definition

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

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

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

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

$$P(A \cap B) = P(B)P(A|B) \qquad (2.17)$$

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

$$P(A \cap B) = P(A)P(B|A) \qquad (2.18)$$

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

$$P(A \cap B) = P(A)P(B) \qquad (2.19)$$

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

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

In particular,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n) \qquad (2.21)$$

Law of Total Probability

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

$$P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B) \qquad (2.23)$$

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

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n) \qquad (2.24)$$

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

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

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

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

Summary

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 $F(x) = P(X \le x)$.
- $F(x) = \sum_{t \le x} p(t) = \sum_{t \le x} P(X = t)$.
- $\sum_x p(x) = \sum_x P(X = x) = 1$, where the sum is over all the possible values of X.

Definition

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 μ.

Summary

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 σ^2.
- The standard deviation is the square root of the variance: $\sigma_X = \sqrt{\sigma_X^2}$.

Summary

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

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

In addition,

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

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

Summary

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

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

Definition

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

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

Definition

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 μ.

Definition

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 σ^2.
- The standard deviation is the square root of the variance: $\sigma_X = \sqrt{\sigma_X^2}$.

Definition

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

Chebyshev's Inequality

Let X be a random variable with mean μ_X and standard deviation σ_X. Then

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

Summary

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

$$\mu_{X+b} = \mu_X + b \tag{2.39}$$

$$\sigma_{X+b}^2 = \sigma_X^2 \tag{2.40}$$

Summary

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

$$\mu_{aX} = a\mu_X \tag{2.41}$$

Summary

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

$$\sigma_{aX}^2 = a^2\sigma_X^2 \tag{2.42}$$

$$\sigma_{aX} = |a|\sigma_X \tag{2.43}$$

Summary

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

$$\mu_{aX+b} = a\mu_X + b \tag{2.44}$$

$$\sigma_{aX+b}^2 = a^2\sigma_X^2 \tag{2.45}$$

$$\sigma_{aX+b} = |a|\sigma_X \tag{2.46}$$

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

$$c_1X_1 + \cdots + c_nX_n$$

is called a **linear combination** of $X_1, \ldots, X_n$.

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

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

Definition

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, then the variance of the sum $X_1 + X_2 + \cdots + X_n$ is given by

$$\sigma^2_{X_1+X_2+\cdots+X_n} = \sigma^2_{X_1} + \sigma^2_{X_2} + \cdots + \sigma^2_{X_n} \tag{2.52}$$

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 σ_X^2 and σ_Y^2, then the variance of the sum $X + Y$ is

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y \tag{2.54}$$

The variance of the difference $X - Y$ is

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y \tag{2.55}$$

Summary

If $X_1, \ldots, X_n$ is a simple random sample from a population with mean μ and variance σ^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

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

Summary

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.

Summary

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:

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

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

Summary

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

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

Definition

- If the random variables $X_1, \ldots, X_n$ are jointly discrete, the joint probability mass function is

  $$p(x_1, \ldots, x_n) = P(X_1 = x_1, \ \ldots, \ X_n = x_n)$$

- If the random variables $X_1, \ldots, X_n$ are jointly continuous, they have a joint probability density function $f(x_1, \ldots, x_n)$, where

  $$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n) = \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_n)\,dx_1 \cdots dx_n$$

  for any constants $a_1 \le b_1, \ldots, a_n \le b_n$.

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

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

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

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

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 $f(x, y)$,

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

Definition

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

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

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

Definition

Let X and Y be jointly continuous random variables, with joint probability density function $f(x, y)$. Let $f_X(x)$ denote the marginal probability density function of X and let x be any number for which $f_X(x) > 0$.

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

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

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

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

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

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

Definition

Two random variables X and Y are independent, provided that

- If X and Y are jointly discrete, the joint probability mass function is equal to the product of the marginals:

$$p(x,y) = p_X(x)p_Y(y)$$

- If X and Y are jointly continuous, the joint probability density function is equal to the product of the marginals:

$$f(x,y) = f_X(x)f_Y(y)$$

Random variables $X_1, \ldots, X_n$ are independent, provided that

- If $X_1, \ldots, X_n$ are jointly discrete, the joint probability mass function is equal to the product of the marginals:

$$p(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$

- If $X_1, \ldots, X_n$ are jointly continuous, the joint probability density function is equal to the product of the marginals:

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

Definition

Let X and Y be random variables with means μ_X and μ_Y. The covariance of X and Y is

$$\text{Cov}(X,Y) = \mu_{(X-\mu_X)(Y-\mu_Y)} \tag{2.68}$$

An alternate formula is

$$\text{Cov}(X,Y) = \mu_{XY} - \mu_X\mu_Y \tag{2.69}$$

Summary

Let X and Y be jointly distributed random variables with standard deviations σ_X and σ_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:

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

Summary

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

Summary

- If $\text{Cov}(X,Y) = \rho_{X,Y} = 0$, then X and Y are said to be uncorrelated.
- If X and Y are independent, then X and Y are uncorrelated.
- It is mathematically possible for X and Y to be uncorrelated without being independent. This rarely occurs in practice.

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

$$c_1X_1 + \cdots + c_nX_n$$

is called a **linear combination** of $X_1, \ldots, X_n$.

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

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

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

If $X_1, \ldots, X_n$ are *independent* random variables and $c_1, \ldots, c_n$ 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,

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

If X and Y are random variables, then

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\,\text{Cov}(X,Y) \tag{2.75}$$

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\,\text{Cov}(X,Y) \tag{2.76}$$

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

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y \tag{2.77}$$

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y \tag{2.78}$$

If $X_1, \ldots, X_n$ is a simple random sample from a population with mean μ and variance σ^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

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

Chapter 3

Summary

- A measured value is a random variable with mean μ and standard deviation σ.
- The bias in the measuring process is the difference between the mean measurement and the true value:

$$\text{Bias} = \mu - \text{true value}$$

- The uncertainty in the measuring process is the standard deviation σ.
- The smaller the bias, the more accurate the measuring process.
- The smaller the uncertainty, the more precise the measuring process.

Summary

Let $X_1, \ldots, X_n$ be independent measurements, all made by the same process on the same quantity.

- The sample standard deviation s can be used to estimate the uncertainty.
- Estimates of uncertainty are often crude, especially when based on small samples.
- If the true value is known, the sample mean $\overline{X}$ can be used to estimate the bias: Bias $\approx \overline{X}-$ true value.
- If the true value is unknown, the bias cannot be estimated from repeated measurements.

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

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

If $X_1, \ldots, X_n$ are independent measurements and $c_1, \ldots, c_n$ 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 $X_1, \ldots, X_n$ are n *independent* measurements, each with mean μ and uncertainty σ, then the sample mean $\overline{X}$ is a measurement with mean

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

and with uncertainty

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

Summary

If X and Y are *independent* measurements of the same quantity, with uncertainties σ_X and σ_Y, respectively, then the weighted average of X and Y with the smallest uncertainty is given by $c_{\text{best}}X + (1 - c_{\text{best}})Y$, where

$$c_{\text{best}} = \frac{\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2} \qquad 1 - c_{\text{best}} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2} \tag{3.7}$$

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

$$\sigma_{c_1X_1+\cdots+c_nX_n} \le |c_1|\sigma_{X_1} + \cdots + |c_n|\sigma_{X_n} \tag{3.8}$$

Chapter 4

Summary

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

$$\mu_X = p \tag{4.1}$$

$$\sigma_X^2 = p(1-p) \tag{4.2}$$

Summary

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

Summary

Assume that a finite population contains items of two types, successes and failures, and that a simple random sample is drawn from the population. Then if the sample size is no more than 5% of the population, the binomial distribution may be used to model the number of successes.

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

$$p(x) = P(X = x) = \begin{cases} \dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

Summary

If $X \sim \text{Bin}(n, p)$, then the mean and variance of X are given by

$$\mu_X = np \tag{4.5}$$
$$\sigma_X^2 = np(1-p) \tag{4.6}$$

Summary

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

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

$$\sigma_{\widehat{p}} = \sqrt{\frac{p(1-p)}{n}} \tag{4.7}$$

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

Summary

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

- X is a discrete random variable whose possible values are the non-negative integers.
- The parameter λ 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 $\lambda = np$.

Summary

If $X \sim \text{Poisson}(\lambda)$, then the mean and variance of X are given by

$$\mu_X = \lambda \tag{4.10}$$

$$\sigma_X^2 = \lambda \tag{4.11}$$

Summary

Let λ denote the mean number of events that occur in one unit of time or space. Let X denote the number of events that are observed to occur in t units of time or space. Then if $X \sim \text{Poisson}(\lambda t)$, λ is estimated with $\widehat{\lambda} = X/t$.

Summary

If $X \sim \text{Poisson}(\lambda t)$, we estimate the rate λ with $\widehat{\lambda} = \dfrac{X}{t}$.

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

$$\sigma_{\widehat{\lambda}} = \sqrt{\frac{\lambda}{t}} \tag{4.12}$$

In practice, we substitute $\widehat{\lambda}$ for λ in Equation (4.12), since λ is unknown.

Summary

Assume a finite population contains N items, of which R are classified as successes and $N - R$ are classified as failures. Assume that n items are sampled from this population, and let X represent the number of successes in the sample. Then X has the hypergeometric distribution with parameters N, R, and n, which can be denoted $X \sim \text{H}(N, R, n)$.

The probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \dfrac{\dbinom{R}{x}\dbinom{N-R}{n-x}}{\dbinom{N}{n}} & \max(0, R + n - N) \le x \le \min(n, R) \\ 0 & \text{otherwise} \end{cases} \tag{4.15}$$

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

$$\mu_X = \frac{nR}{N} \tag{4.16}$$

$$\sigma_X^2 = n\left(\frac{R}{N}\right)\left(1 - \frac{R}{N}\right)\left(\frac{N-n}{N-1}\right) \tag{4.17}$$

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

$$p(x) = P(X = x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

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

$$\mu_X = \frac{1}{p} \tag{4.18}$$

$$\sigma_X^2 = \frac{1-p}{p^2} \tag{4.19}$$

If $X \sim \text{NB}(r, p)$, then the probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \dbinom{x-1}{r-1} p^r (1-p)^{x-r} & x = r, r+1, \dots \\ 0 & \text{otherwise} \end{cases}$$

Summary

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

$$X = Y_1 + \cdots + Y_r$$

where $Y_1, \dots, Y_r$ are independent random variables, each with the $\text{Geom}(p)$ distribution.

Summary

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

$$\mu_X = \frac{r}{p} \tag{4.20}$$

$$\sigma_X^2 = \frac{r(1-p)}{p^2} \tag{4.21}$$

Assume n independent trials are performed, each of which results in one of k possible outcomes. Let $x_1, \dots, x_k$ be the numbers of trials resulting in outcomes $1, 2, \dots, k$, respectively. The number of arrangements of the outcomes among the n trials is

$$\frac{n!}{x_1!\, x_2! \cdots x_k!}$$

If $X_1, \dots, X_k \sim \text{MN}(n, p_1, \dots, p_k)$, then the probability mass function of $X_1, \dots, X_k$ is

$$p(x_1, \dots, x_k) = P(X_1 = x_1, \dots, X_k = x_k)$$
$$= \begin{cases} \dfrac{n!}{x_1!\, x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} & x_i = 0, 1, 2, \dots, n \text{ and } \sum x_i = n \\ 0 & \text{otherwise} \end{cases} \tag{4.22}$$

If $X_1, \dots, X_k \sim \text{MN}(n, p_1, \dots, p_k)$, then for each i

$$X_i \sim \text{Bin}(n, p_i)$$

Summary

If $X \sim N(\mu, \sigma^2)$, then the mean and variance of X are given by

$$\mu_X = \mu$$
$$\sigma_X^2 = \sigma^2$$

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

Summary

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

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

Summary

Let $X_1, X_2, \ldots, X_n$ be independent and normally distributed with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$. Let $c_1, c_2, \ldots, c_n$ be constants, and $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ be a linear combination. Then

$$c_1X_1+c_2X_2+\cdots+c_nX_n \sim N(c_1\mu_1+c_2\mu_2+\cdots+c_n\mu_n,\ c_1^2\sigma_1^2+c_2^2\sigma_2^2+\cdots+c_n^2\sigma_n^2) \tag{4.26}$$

Summary

Let $X_1, \ldots, X_n$ be independent and normally distributed with mean μ and variance σ^2. Then

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

Summary

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

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2) \tag{4.28}$$

$$X - Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2) \tag{4.29}$$

Summary

- If $X \sim N(\mu, \sigma^2)$, then the random variable $Y = e^X$ has the lognormal distribution with parameters μ and σ^2.
- If Y has the lognormal distribution with parameters μ and σ^2, then the random variable $X = \ln Y$ has the $N(\mu, \sigma^2)$ distribution.

$$f(x) = \begin{cases} \dfrac{1}{\sigma x\sqrt{2\pi}} \exp\left[-\dfrac{1}{2\sigma^2}(\ln x - \mu)^2\right] & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases} \tag{4.30}$$

$$E(Y) = e^{\mu+\sigma^2/2} \qquad V(Y) = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} \tag{4.31}$$

Definition

The probability density function of the exponential distribution with parameter $\lambda > 0$ is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{4.32}$$

Summary

If $X \sim \text{Exp}(\lambda)$, the cumulative distribution function of X is

$$F(x) = P(X \le x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{4.33}$$

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

$$\mu_X = \frac{1}{\lambda} \tag{4.34}$$

$$\sigma_X^2 = \frac{1}{\lambda^2} \tag{4.35}$$

If events follow a Poisson process with rate parameter λ, and if T represents the waiting time from any starting point until the next event, then $T \sim \text{Exp}(\lambda)$.

Lack of Memory Property

If $T \sim \text{Exp}(\lambda)$, and t and s are positive numbers, then

$$P(T > t + s \mid T > s) = P(T > t)$$

Summary

If $X_1, \ldots, X_n$ is a random sample from $\text{Exp}(\lambda)$, then the parameter λ is estimated with

$$\widehat{\lambda} = \frac{1}{\overline{X}} \tag{4.36}$$

This estimator is biased. The bias is approximately equal to λ/n. The uncertainty in $\widehat{\lambda}$ is estimated with

$$\sigma_{\widehat{\lambda}} \approx \frac{1}{\overline{X}\sqrt{n}} \tag{4.37}$$

This uncertainty estimate is reasonably good when the sample size is more than 20.

Definition

The probability density function of the continuous uniform distribution with parameters a and b is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases} \tag{4.41}$$

If X is a random variable with probability density function $f(x)$, we say that X is uniformly distributed on the interval (a, b).

Let $X \sim U(a, b)$. Then

$$\mu_X = \frac{a+b}{2} \tag{4.42}$$

$$\sigma_X^2 = \frac{(b-a)^2}{12} \tag{4.43}$$

Definition

For $r > 0$, the gamma function is defined by

$$\Gamma(r) = \int_0^\infty t^{r-1}e^{-t}\,dt \tag{4.44}$$

The gamma function has the following properties:

1. If r is an integer, then $\Gamma(r) = (r-1)!$.
2. For any r, $\Gamma(r+1) = r\Gamma(r)$.
3. $\Gamma(1/2) = \sqrt{\pi}$.

Definition

The probability density function of the gamma distribution with parameters $r > 0$ and $\lambda > 0$ is

$$f(x) = \begin{cases} \dfrac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{4.45}$$

Summary

If $X_1, \ldots, X_r$ are independent random variables, each distributed as $\text{Exp}(\lambda)$, then the sum $X_1 + \cdots + X_r$ is distributed as $\Gamma(r, \lambda)$.

If $X \sim \Gamma(r, \lambda)$, then

$$\mu_X = \frac{r}{\lambda} \tag{4.46}$$

$$\sigma_X^2 = \frac{r}{\lambda^2} \tag{4.47}$$

If $T \sim \Gamma(r, \lambda)$, and r is a positive integer, the cumulative distribution function of T is given by

$$F(x) = P(T \le x) = \begin{cases} 1 - \displaystyle\sum_{j=0}^{r-1} e^{-\lambda x}\frac{(\lambda x)^j}{j!} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{4.48}$$

The Weibull Distribution

$$f(x) = \begin{cases} \alpha\beta^\alpha x^{\alpha-1}e^{-(\beta x)^\alpha} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{4.49}$$

$$F(x) = P(X \le x) = \begin{cases} \displaystyle\int_0^x \alpha\beta^\alpha t^{\alpha-1}e^{-(\beta t)^\alpha}\,dt = 1 - e^{-(\beta x)^\alpha} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{4.50}$$

If $X \sim \text{Weibull}(\alpha, \beta)$, then

$$\mu_X = \frac{1}{\beta}\Gamma\left(1+\frac{1}{\alpha}\right) \tag{4.51}$$

$$\sigma_X^2 = \frac{1}{\beta^2}\left\{\Gamma\left(1+\frac{2}{\alpha}\right) - \left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2\right\} \tag{4.52}$$

In the special case that $1/\alpha$ is an integer, then

$$\mu_X = \frac{1}{\beta}\left[\left(\frac{1}{\alpha}\right)!\right] \qquad \sigma_X^2 = \frac{1}{\beta^2}\left\{\left(\frac{2}{\alpha}\right)! - \left[\left(\frac{1}{\alpha}\right)!\right]^2\right\}$$

Definition

Let θ be a parameter, and $\widehat{\theta}$ an estimator of θ. The mean squared error (MSE) of $\widehat{\theta}$ is

$$\text{MSE}_{\widehat{\theta}} = (\mu_{\widehat{\theta}} - \theta)^2 + \sigma_{\widehat{\theta}}^2 \tag{4.53}$$

An equivalent expression for the MSE is

$$\text{MSE}_{\widehat{\theta}} = \mu_{(\widehat{\theta}-\theta)^2} \tag{4.54}$$

Definition

Let $X_1, \ldots, X_n$ have joint probability density or probability mass function $f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_k)$, where $\theta_1, \ldots, \theta_k$ are parameters, and $x_1, \ldots, x_n$ are the values observed for $X_1, \ldots, X_n$. The values $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$ that maximize f are the maximum likelihood estimates of $\theta_1, \ldots, \theta_k$.

If the random variables $X_1, \ldots, X_n$ are substituted for $x_1, \ldots, x_n$, then $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$ are called maximum likelihood estimators.

The abbreviation MLE is often used for both maximum likelihood estimate and maximum likelihood estimator.

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be a simple random sample from a population with mean μ and variance σ^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

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

For most populations, if the sample size is greater than 30, the Central Limit Theorem approximation is good.

Summary

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

$$X \sim N(np, np(1-p)) \quad \text{approximately} \tag{4.57}$$

$$\widehat{p} \sim N\left(p, \frac{p(1-p)}{n}\right) \quad \text{approximately} \tag{4.58}$$

To compute $P(45 \le X \le 55)$, the areas of the rectangles corresponding to 45 and to 55 should be included. To approximate this probability with the normal curve, compute the area under the curve between 44.5 and 55.5.

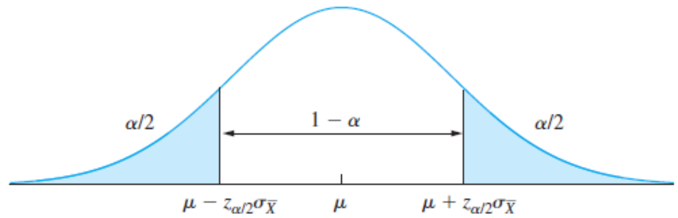
To compute $P(45 < X < 55)$, the areas of the rectangles corresponding to 45 and to 55 should be excluded. To approximate this probability with the normal curve, compute the area under the curve between 45.5 and 54.5.

Summary

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

$$X \sim N(\lambda, \lambda) \quad \text{approximately} \tag{4.59}$$

Chapter 5



Summary

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

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

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

- $\overline{X} \pm \dfrac{s}{\sqrt{n}}$ is a 68% confidence interval for μ.
- $\overline{X} \pm 1.645\dfrac{s}{\sqrt{n}}$ is a 90% confidence interval for μ.
- $\overline{X} \pm 1.96\dfrac{s}{\sqrt{n}}$ is a 95% confidence interval for μ.
- $\overline{X} \pm 2.58\dfrac{s}{\sqrt{n}}$ is a 99% confidence interval for μ.
- $\overline{X} \pm 3\dfrac{s}{\sqrt{n}}$ is a 99.7% confidence interval for μ.

Summary

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

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

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

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

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

Summary

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.

Summary

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 $100(1-\alpha)\%$ lower confidence bound for p is

$$\tilde{p} - z_\alpha\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \tag{5.6}$$

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

$$\tilde{p} + z_\alpha\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \tag{5.7}$$

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

Summary

The Traditional Method for Computing Confidence Intervals for a Proportion (widely used but not recommended)

Let $\widehat{p}$ be the proportion of successes in a *large* number n of independent Bernoulli trials with success probability p. Then the traditional level $100(1-\alpha)\%$ confidence interval for p is

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \tag{5.8}$$

The method cannot be used unless the sample contains at least 10 successes and 10 failures.

Summary

Let $X_1, \ldots, X_n$ be a *small* (e.g., $n < 30$) sample from a *normal* population with mean μ. Then the quantity

$$\frac{\overline{X}-\mu}{s/\sqrt{n}}$$

has a Student's t distribution with $n-1$ degrees of freedom, denoted t_{n-1}.

When n is large, the distribution of the quantity $(\overline{X}-\mu)/(s/\sqrt{n})$ is very close to normal, so the normal curve can be used, rather than the Student's t.

Summary

Let $X_1, \ldots, X_n$ be a *small* random sample from a *normal* population with mean μ. Then a level $100(1-\alpha)\%$ confidence interval for μ is

$$\overline{X} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \tag{5.9}$$

Let $X_1, \ldots, X_n$ be a *small* random sample from a *normal* population with mean μ. Then a level $100(1-\alpha)\%$ upper confidence bound for μ is

$$\overline{X} + t_{n-1,\alpha}\frac{s}{\sqrt{n}} \tag{5.10}$$

and a level $100(1-\alpha)\%$ lower confidence bound for μ is

$$\overline{X} - t_{n-1,\alpha}\frac{s}{\sqrt{n}} \tag{5.11}$$

Summary

Let $X_1, \ldots, X_n$ be a random sample (of any size) from a *normal* population with mean μ. If the standard deviation σ is known, then a level $100(1-\alpha)\%$ confidence interval for μ is

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{5.12}$$

Summary

Let X be a single value sampled from a *normal* population with mean μ. If the standard deviation σ is known, then a level $100(1-\alpha)\%$ confidence interval for μ is

$$X \pm z_{\alpha/2}\sigma \tag{5.13}$$

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

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

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

Summary

Let $X_1, \ldots, X_{n_X}$ be a *large* random sample of size n_X from a population with mean μ_X and standard deviation σ_X, and let $Y_1, \ldots, Y_{n_Y}$ be a *large* random sample of size n_Y from a population with mean μ_Y and standard deviation σ_Y. If the two samples are independent, then a level $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is

$$\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \tag{5.16}$$

When the values of σ_X and σ_Y are unknown, they can be replaced with the sample standard deviations s_X and s_Y.

Summary

Let X be the number of successes in n_X independent Bernoulli trials with success probability p_X, and let Y be the number of successes in n_Y independent Bernoulli trials with success probability p_Y, so that $X \sim \text{Bin}(n_X, p_X)$ and $Y \sim \text{Bin}(n_Y, p_Y)$. Define $\tilde{n}_X = n_X + 2$, $\tilde{n}_Y = n_Y + 2$, $\tilde{p}_X = (X+1)/\tilde{n}_X$, and $\tilde{p}_Y = (Y+1)/\tilde{n}_Y$.

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

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

If the lower limit of the confidence interval is less than -1, replace it with -1. If the upper limit of the confidence interval is greater than 1, replace it with 1.

Summary

The Traditional Method for Computing Confidence Intervals for the Difference Between Proportions (widely used but not recommended)

Let $\widehat{p}_X$ be the proportion of successes in a *large* number n_X of independent Bernoulli trials with success probability p_X, and let $\widehat{p}_Y$ be the proportion of successes in a *large* number n_Y of independent Bernoulli trials with success probability p_Y. Then the traditional level $100(1-\alpha)\%$ confidence interval for $p_X - p_Y$ is

$$\widehat{p}_X - \widehat{p}_Y \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n_X} + \frac{\widehat{p}_Y(1-\widehat{p}_Y)}{n_Y}} \tag{5.19}$$

This method cannot be used unless both samples contain at least 10 successes and 10 failures.

Summary

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

If the populations do not necessarily have the same variance, a level $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is

$$\overline{X} - \overline{Y} \pm t_{\nu,\alpha/2}\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} \tag{5.21}$$

The number of degrees of freedom, ν, is given by

$$\nu = \frac{\left(\dfrac{s_X^2}{n_X} + \dfrac{s_Y^2}{n_Y}\right)^2}{\dfrac{(s_X^2/n_X)^2}{n_X - 1} + \dfrac{(s_Y^2/n_Y)^2}{n_Y - 1}}$$

rounded down to the nearest integer.

Summary

Let $X_1, \ldots, X_{n_X}$ be a random sample of size n_X from a *normal* population with mean μ_X, and let $Y_1, \ldots, Y_{n_Y}$ be a random sample of size n_Y from a *normal* population with mean μ_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

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

Summary

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 μ_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 μ_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).

Summary

Let $X_1, \ldots, X_n$ be a random sample from a normal population with variance σ^2. The sample variance is $s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$. The quantity

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

has a chi-square distribution with $n-1$ degrees of freedom, denoted χ^2_{n-1}.

Summary

Let $X_1, \ldots, X_n$ be a random sample from a *normal* population with variance σ^2. Let s^2 be the sample variance. A level $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \quad \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} \right)$$

A level $100(1-\alpha)\%$ confidence interval for the standard deviation σ is

$$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}}, \quad \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}} \right)$$

Summary

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

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

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

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 $100(1-\alpha)\%$ upper prediction bound for Y is

$$\overline{X} + t_{n-1,\alpha}\, s\sqrt{1+\frac{1}{n}} \qquad (5.27)$$

and a level $100(1-\alpha)\%$ lower prediction bound for Y is

$$\overline{X} - t_{n-1,\alpha}\, s\sqrt{1+\frac{1}{n}} \qquad (5.28)$$

Summary

Let $X_1, \ldots, X_n$ be a sample from a *normal* population. A tolerance interval containing at least $100(1-\gamma)\%$ of the population with confidence $100(1-\alpha)\%$ is

$$\overline{X} \pm k_{n,\alpha,\gamma}\, s \qquad (5.29)$$

Of all the tolerance intervals that are computed by this method, $100(1-\alpha)\%$ will actually contain at least $100(1-\gamma)\%$ of the population.

Chapter 6

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 significance 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 μ and standard deviation σ.
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: $z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$.
  If σ 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 $-z$

Summary

- The smaller the P-value, the more certain we can be that H_0 is false.
- The larger the P-value, the more plausible H_0 becomes, but we can never be certain that H_0 is true.
- A rule of thumb suggests to reject H_0 whenever $P \le 0.05$. While this rule is convenient, it has no scientific basis.

Summary

Let α be any value between 0 and 1. Then, if $P \le \alpha$,

- The result of the test is said to be statistically significant at the $100\alpha\%$ level.
- The null hypothesis is rejected at the $100\alpha\%$ level.
- When reporting the result of a hypothesis test, report the P-value, rather than just comparing it to 5% or 1%.

Summary

Let X be the number of successes in n independent Bernoulli trials, each with success probability p; in other words, let $X \sim \text{Bin}(n, p)$.
To test a null hypothesis of the form $H_0: p \le p_0$, $H_0: p \ge p_0$, or $H_0: p = p_0$, assuming that both np_0 and $n(1-p_0)$ are greater than 10:

- Compute the z-score: $z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$.
- 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**
$H_1: p > p_0$	Area to the right of z
$H_1: p < p_0$	Area to the left of z
$H_1: p \ne p_0$	Sum of the areas in the tails cut off by z and $-z$

Summary

Let $X_1, \ldots, X_n$ be a sample from a *normal* population with mean μ and standard deviation σ, where σ is unknown.
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 test statistic $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$.
- Compute the P-value. The P-value is an area under the Student's t curve with $n-1$ degrees of freedom, which depends on the alternate hypothesis as follows:

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

- If σ is known, the test statistic is $z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$, and a z test should be performed.

Summary

Let $X_1, \ldots, X_{n_X}$ and $Y_1, \ldots, Y_{n_Y}$ be *large* (e.g., $n_X > 30$ and $n_Y > 30$) samples from populations with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. Assume the samples are drawn independently of each other.

To test a null hypothesis of the form $H_0: \mu_X - \mu_Y \le \Delta_0$, $H_0: \mu_X - \mu_Y \ge \Delta_0$, or $H_0: \mu_X - \mu_Y = \Delta_0$:

- Compute the z-score: $z = \dfrac{(\overline{X} - \overline{Y}) - \Delta_0}{\sqrt{\sigma_X^2/n_X + \sigma_Y^2/n_Y}}$. If σ_X and σ_Y are unknown they may be approximated with s_X and s_Y, respectively.
- 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
$H_1: \mu_X - \mu_Y > \Delta_0$	Area to the right of z
$H_1: \mu_X - \mu_Y < \Delta_0$	Area to the left of z
$H_1: \mu_X - \mu_Y \ne \Delta_0$	Sum of the areas in the tails cut off by z and $-z$

Summary

Let $X \sim \text{Bin}(n_X, p_X)$ and let $Y \sim \text{Bin}(n_Y, p_Y)$. Assume that there are at least 10 successes and 10 failures in each sample, and that X and Y are independent.

To test a null hypothesis of the form $H_0: p_X - p_Y \le 0$, $H_0: p_X - p_Y \ge 0$, or $H_0: p_X - p_Y = 0$:

- Compute $\widehat{p}_X = \dfrac{X}{n_X}$, $\widehat{p}_Y = \dfrac{Y}{n_Y}$, and $\widehat{p} = \dfrac{X+Y}{n_X + n_Y}$.
- Compute the z-score: $z = \dfrac{\widehat{p}_X - \widehat{p}_Y}{\sqrt{\widehat{p}(1-\widehat{p})(1/n_X + 1/n_Y)}}$.
- 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
$H_1: p_X - p_Y > 0$	Area to the right of z
$H_1: p_X - p_Y < 0$	Area to the left of z
$H_1: p_X - p_Y \ne 0$	Sum of the areas in the tails cut off by z and $-z$

Summary

Let $X_1, \ldots, X_{n_X}$ and $Y_1, \ldots, Y_{n_Y}$ be samples from *normal* populations with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. Assume the samples are drawn independently of each other.

If σ_X and σ_Y are not known to be equal, then, to test a null hypothesis of the form $H_0: \mu_X - \mu_Y \le \Delta_0$, $H_0: \mu_X - \mu_Y \ge \Delta_0$, or $H_0: \mu_X - \mu_Y = \Delta_0$:

- Compute $\nu = \dfrac{[(s_X^2/n_X) + (s_Y^2/n_Y)]^2}{[(s_X^2/n_X)^2/(n_X-1)] + [(s_Y^2/n_Y)^2/(n_Y-1)]}$, rounded down to the nearest integer.
- Compute the test statistic $t = \dfrac{(\overline{X} - \overline{Y}) - \Delta_0}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}}$.
- Compute the P-value. The P-value is an area under the Student's t curve with ν degrees of freedom, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	P-value
$H_1: \mu_X - \mu_Y > \Delta_0$	Area to the right of t
$H_1: \mu_X - \mu_Y < \Delta_0$	Area to the left of t
$H_1: \mu_X - \mu_Y \ne \Delta_0$	Sum of the areas in the tails cut off by t and $-t$

Summary

Let $X_1, \ldots, X_{n_X}$ and $Y_1, \ldots, Y_{n_Y}$ be samples from *normal* populations with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. Assume the samples are drawn independently of each other.

If σ_X and σ_Y are known to be equal, then, to test a null hypothesis of the form $H_0: \mu_X - \mu_Y \le \Delta_0$, $H_0: \mu_X - \mu_Y \ge \Delta_0$, or $H_0: \mu_X - \mu_Y = \Delta_0$:

- Compute $s_p = \sqrt{\dfrac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}}$.
- Compute the test statistic $t = \dfrac{(\overline{X} - \overline{Y}) - \Delta_0}{s_p\sqrt{1/n_X + 1/n_Y}}$.
- Compute the P-value. The P-value is an area under the Student's t curve with $n_X + n_Y - 2$ degrees of freedom, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	P-value
$H_1: \mu_X - \mu_Y > \Delta_0$	Area to the right of t
$H_1: \mu_X - \mu_Y < \Delta_0$	Area to the left of t
$H_1: \mu_X - \mu_Y \ne \Delta_0$	Sum of the areas in the tails cut off by t and $-t$

Summary

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of ordered pairs whose differences $D_1, \ldots, D_n$ are a sample from a *normal* population with mean μ_D. Let s_D be the sample standard deviation of $D_1, \ldots, D_n$.

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

- Compute the test statistic $t = \dfrac{\overline{D} - \mu_0}{s_D/\sqrt{n}}$.
- Compute the P-value. The P-value is an area under the Student's t curve with $n - 1$ degrees of freedom, which depends on the alternate hypothesis as follows:

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

- If the sample is large, the D_i need not be normally distributed, the test statistic is $z = \dfrac{\overline{D} - \mu_0}{s_D/\sqrt{n}}$, and a z test should be performed.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (6.6)$$

Summary

To conduct a fixed-level test:

- Choose a number α, where $0 < \alpha < 1$. This is called the significance level, or the level, of the test.
- Compute the P-value in the usual way.
- If $P \le \alpha$, reject H_0. If $P > \alpha$, do not reject H_0.

If α is the significance level that has been chosen for the test, then the probability of a type I error is never greater than α.

Summary

When conducting a fixed-level test at significance level α, there are two types of errors that can be made. These are

- Type I error: Reject H_0 when it is true.
- Type II error: Fail to reject H_0 when it is false.

The probability of a type I error is never greater than α.

$$\int e^{ax}\,dx = \frac{1}{a}e^{ax}$$

$$\int xe^{ax}\,dx = \frac{e^{ax}}{a^2}(ax - 1)$$

$$\int x^2e^{ax}\,dx = \frac{e^{ax}}{a^3}(a^2x^2 - 2ax + 2)$$

$$\int e^{ax}\sin bx\,dx = \frac{e^{ax}}{a^2+b^2}(a\sin bx - b\cos bx)$$

$$\int e^{ax}\cos bx\,dx = \frac{e^{ax}}{a^2+b^2}(a\cos bx + b\sin bx)$$

$$\int \frac{1}{x^2+a^2}\,dx = \frac{1}{a}\tan^{-1}\frac{x}{a}$$

$$\int \frac{x}{x^2+a^2}\,dx = \frac{1}{2}\ln(x^2+a^2)$$