Mathematical Statistics

Hacker-Code.J

July 4, 2020

Contents

1 Introduction

2 Organization and Description of Data

2.1 Main Types of Data

2.2 Describing Data by Tables and Graphs

2.3 Measures of Center

2.4 Measures of Variation

Besides locating the center of the data, any descriptive study of data must numerically measure the extent of variation around the center.

2.4.1 Deviation

$$\begin{aligned}\text{Deviation} &= \text{Observation} - (\text{Sample mean}) \\ &= x - \bar{x}.\end{aligned}$$

For any data set, the total deviation is 0, that is,

$$\sum \text{Deviations} = \sum (x - \bar{x}) = 0.$$

Check. Since $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, *i.e.*, $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n = n\bar{x}$,

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \bar{x}) &= (x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x}) \\ &= x_1 + x_2 + \cdots + x_n - n\bar{x} \\ &= n\bar{x} - n\bar{x} = 0.\end{aligned}$$

2.4.2 Sample Variance and Sample Standard Deviation

Sample Variance of n observations:

$$s^2 = \frac{\text{Sum of squared deviations}}{n-1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Remark. Although the sample variance is conceptualized as the **average squared deviation**, notice that the divisor is $n-1$ rather than n. The divisor, $n-1$, is called the degrees of *freedom*[1] associated with s^2.

Sample Standard Deviation

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

[1] The deviations add to 0 so a specification of any $n-1$ deviations allows us to recover the one that is left out. In definition of s^2, the divisor $n-1$ represents the number of deviations that can be viewed as free quantities.

An alternative formula for the sample variance is

$$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$$

Proof. Note that

$$\begin{aligned}\sum (x_i - \bar{x})^2 &= \sum x_i^2 - \sum 2x_i\bar{x} + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2.\end{aligned}$$

Then, since $\bar{x} = \sum x_i/n$,

$$\begin{aligned}\sum (x_i - \bar{x})^2 &= \sum x_i^2 - 2\frac{\sum x_i}{n}\sum x_i + n\left(\frac{\sum x_i}{n}\right)^2 \\ &= \sum x_i^2 - 2\frac{(\sum x_i)^2}{n} + \frac{(\sum x_i)^2}{n} \\ &= \sum x_i^2 - \frac{(\sum x_i)^2}{n}.\end{aligned}$$

□

It does not require the calculation of the individual deviations.

2.4.3 Population Mean and Sample Mean

	Population(N)	**Sample**(n)
Mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n-1}$
Variance	$\sigma^2 = \frac{\sum (x-\mu)^2}{N}$	$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$

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

Proof.

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2&=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2\\&=\frac{1}{n}\sum_{i=1}^{n}\left[(x_i-\bar{x})^2+2(x-\bar{x})(x-\mu)+(\bar{x}-\mu)^2\right]\\&=\frac{1}{n}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+2(x-\mu)\underbrace{\sum_{i=1}^{n}(x-\bar{x})}_{=0}+\sum_{i=1}^{n}(\bar{x}-\mu)^2\right]\\&=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2+(\bar{x}-\mu)^2\end{aligned}$$

□

2.4.4 Other Measure of Variation

Sample range = Largest observation - Smallest observation

The range gives the length of the interval spanned by the observations.

3 Descriptive Study of Bivariate Data

3.1 Simpson's Paradox

3.2 Scatter Diagram of Bivariate Measurement Data

3.3 The Correlation Coefficient - A Measure of Linear Relation

Our visual impression of the closeness of the scatter to a linear relation can be quantified by calculating a numerical measure, called the **correlation coefficient** and denoted "r".

Important features of correlation coefficient:

3.3.1 Calculation of r

The sample correlation is calculated from n pairs of observations on the two characteristics

$$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n).$$

The correlation coefficient is best interpreted in terms of the **standardized observation**, or **sample z values**

$$\frac{\text{Observation}-\text{Sample mean}}{\text{Sample standard deviation}}=\frac{x_i-\bar{x}}{s_x}$$

where $s_x=\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1)}$ and the subscript x on s distinguishes the sample the sample standard deviation of the x observation from the sample standard deviation s_y of the y observations.

The **sample correlation coefficient** is the sum of the products of the standardized x observation times the standardized y observation divided by $n-1$.

Sample Correlation Coefficient

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$$

Calculation Formula for the Sample Correlation Coefficient

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

where

$$S_{xy} = \sum(x-\bar{x})(y-\bar{y})$$
$$S_{xx} = \sum(x-\bar{x})^2, \quad S_{yy} = \sum(y-\bar{y})^2$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n},\ S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n},\ \text{and } S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}.$$

Proof. Since $(x-\bar{x})(y-\bar{y}) = xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y}$,

$$\sum(x-\bar{x})(y-\bar{y}) = \sum xy - \bar{y}\sum x - \bar{x}\sum y + n\bar{x}\bar{y}.$$

Then, since $\bar{x} = \sum x/n$ and $\bar{y} = \sum y/n$,

$$\begin{aligned}\sum(x-\bar{x})(y-\bar{y}) &= \sum xy - \frac{\sum y}{n}\sum x - \frac{\sum x}{n}\sum y + n\frac{\sum x}{n}\frac{\sum y}{n}\\ &= \sum xy - \frac{2\sum x\sum y}{n} + \frac{\sum xy}{n}\\ &= \sum xy - \frac{\sum x\sum y}{n}\end{aligned}$$

□

e.g.) Calculation of Sample Correlation Calculate r for the $n = 4$ pairs of observations

$$(2,5)\quad (1,3)\quad (5,6)\quad (0,2)$$

Sol.

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$	x^2	y^2	xy
	2	5	0	1	0	1	0	4	25	10
	1	3	-1	-1	1	1	1	1	9	3
	5	6	3	2	9	4	6	25	36	30
	0	2	-2	-2	4	4	4	0	4	0
Total	8	16	0	0	14	10	11	30	74	43
	$\bar{x}=2$	$\bar{y}=4$			S_{xx}	S_{yy}	S_{xy}	$\sum x^2$	$\sum y^2$	$\sum xy$

Consequently,

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{11}{\sqrt{14}\sqrt{10}} = .930$$

The value .930 is large and it implies a strong associate where both x and y tend to be small or both tend to be large. □

3.3.2 Population Correlation Coefficient(ρ) and Sample Correlation Coefficient(r)

$$\rho = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{c_{1i}-\mu_1}{\sigma_1}\right)\left(\frac{c_{2i}-\mu_2}{\sigma_2}\right)$$

- $-1 \le \rho \le 1$

Proof.

$$\begin{aligned}
0 &\le \frac{1}{N}\sum_{i=1}^{N}\left(\frac{c_{1i}-\mu_1}{\sigma_1} - \rho\frac{c_{2i}-\mu_2}{\sigma_2}\right)^2 \\
&= \frac{1}{\sigma_1^2}\cdot\frac{1}{N}\sum_{i=1}^{N}(c_{1i}-\mu_1)^2 - 2\rho\cdot\frac{1}{N}\sum_{i=1}^{N}\left(\frac{c_{1i}-\mu_1}{\sigma_1}\right)\left(\frac{c_{2i}-\mu_2}{\sigma_2}\right) + \frac{\rho^2}{\sigma_2^2}\cdot\frac{1}{N}\sum_{i=1}^{N}(c_{2i}-\mu_2)^2 \\
&= 1 - 2\rho^2 + \rho^2 \\
&= 1 - \rho^2.
\end{aligned}$$

Thus, $\rho^2 \le 1$, i.e, $-1 \le \rho \le 1$. □

3.3.3 Correlation and Caution

An observed correlation between two variables may be **spurious**. That is, it may be caused by the influence of a third variable.

4 Probability

4.1 Probability of an Event

An **experiment** is the process of observing a phenomenon that has variation in its outcomes.

- The **sample space** S associated with an experiment is the collection of all possible distinct outcomes of the experiment.
  - Each outcome is called an elementary outcome, a simple event, or an element of the sample space, say, $e_1, e_2, \cdots$.
- An **event** A, B is the set of elementary outcomes possessing a designated feature. $(A, B \subseteq S)$
  - An event A occurs when any one of the elementary outcomes in A occurs.

The **probability of an event** is a numerical value that represents the proportion of times the event is expected to occur when the experiment is repeated many times under identical conditions. (Relative Frequency)

The probability of event A is denoted by P(A).

Probability must be satisfy:

1. $0 \leq P(A) \leq 1$ for all events A
2. $P(A) = \sum_{\text{all } e \in A} P(e)$
3. $P(S) = \sum_{\text{all } e \in S} P(e) = 1$

4.2 Methods of Assigning Probability

4.2.1 Equally Likely Elementary Outcomes_ The Uniform Probability Model

e.g.) Consider the experiment of rolling a fair die and recording top face. Then

$$S = \{e_1, e_2, \cdots, e_6\}$$

$$P(e_1) = P(e_2) = \cdots = P(e_6) = \frac{1}{6}$$

Q. What is the probability of getting a number higher than 4?
A. Note that $A = \{e_5, e_6\}$.

$$P(A) = P(e_5) + P(e_6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

4.2.2 Probability as The Long-Run Relative Frequency

Letting N denote the number of repetition (or trials) of an experiment, we set

$$\begin{aligned}\text{Relative frequency of event } A \text{ in } N \text{ trials} &= \frac{\text{No. of times } A \text{ occurs in } N \text{ trials}}{N} \\ &= P_N.\end{aligned}$$

Then

$$P(A) = \lim_{N \to \infty} P_N.$$

We define $P(A)$, the probability of an event A, as the value to which the relative frequency stabilizes with increasing number of trials.

Although we will never know $P(A)$ exactly, it can be estimated accurately by repeating the experiment many times.

4.3 Event Relation and Two Laws of Probability

For event A,

- complement: A^C, $\bar{A}$
- union: $A \cup B$

- intersection: $A \cap B$, AB

Two events A, B are mutually disjoint $\iff$ $A \cap B = \varnothing$, *i.e.*, $AB = \varnothing$

(**Law of Complement**) $P(A) = 1 - P(\bar{A})$

(**Addition Law**) $P(A \cup B) = P(A) + P(B) - P(AB)$

4.* Axioms of Probability

If $P(A)$ satisfies as follows:

1. For any event A, $0 \leq P(A) \leq 1$.
2. $P(S) = 1$, where S is sample space.
3. For events that are mutually disjoint, $A_1, A_2, \cdots$,

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots .$$

4.4 Conditional Probability and Independence

The probability of A when it is known that B has occurred is called the conditional probability of A given B and is denoted by $P(A|B)$.

The **conditional probability** of A given B is denoted by $P(A|B)$ and defined by the formula

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Equivalently, this formula can be written

$$P(AB) = P(B)P(A|B)$$

This latter version is called the **multiplication law of probability**.

Two events A and B are **independent** if

$$P(A|B) = P(A)$$

Equivalent conditions are

$$P(B|A) = P(B)$$

or

$$P(AB) = P(A)P(B)$$

- -

Check.

$$P(A) = P(A|B) = \frac{P(AB)}{P(B)} \implies P(AB) = P(A)P(B)$$

Caution. Do not confuse the terms "incompatible events" and "independent events". We say A and B are incompatible when their intersection AB is empty, so $P(AB) = 0$. On the other hand, if A and B are independent, $P(AB) = P(A)P(B)$. Both these properties cannot hold as long as A and B have non-zero probabilities.

4.5 Bayes' Theorem

An event A can occur either when an event B occurs or when it does not occur. That is, A can be written as the disjoint union of AB and $A\bar{B}$. Consequently,

$$P(A) = P(AB) + P(A\bar{B}).$$

(**Rule of Total Probability**) $P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$

Extend. For events $A_1, A_2, \cdots, A_n$, where $A_1 \cup A_2 \cup \cdots \cup A_n = S$ and $A_i \cap A_j = \varnothing$ for $i \neq j$, $P(A_i) > 0$,

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$

- -

- (**prior probability**) Suppose the two events A and B can occur together and, before observing either, we know the probability

  $$P(B) \text{ so } P(\bar{B}) = 1 - P(B).$$

  When we also know the two conditional probabilities $P(A|B)$ and $P(A|\bar{B})$, the probability of B can be updated when we observe the status of A.

- (**posterior probability**) Once we know A has occurred, the updated or *posterior probability* of B is given by the condition probability $P(B|A) = \dfrac{P(AB)}{P(A)}$.

(**Bayes' Theorem**)

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$$

The posterior probability of $\bar{B}$ is then $P(\bar{B}|A) = 1 - P(B|A)$

5 Probability Distributions

5.1 Random Variables

A **random variable** X associates a numerical value with each outcome of an experiment.

$$X : S \to \mathbb{R},$$

where S is a sample space and $\mathbb{R}$ is a real number.

5.2 Probability Distribution of a Discrete Random Variable

The **probability distribution**, or simply the **distribution**, of a discrete random variable X is a list of the distinct numerical values of X along with their associated probabilities. (Often, a formula can be used in place of a detailed list.)

Consider the distinct values of a random variable X. The probability that a particular value x_i occurs

will be denoted by $f(x_i)$. If X can take k possible values $x_1, \cdots, x_k$ with the corresponding probabilities $f(x_1), \cdots, f(x_k)$, the probability distribution of X can be displayed in the format of below table.

Value of x	Probability $f(x)$
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	$\vdots$
x_k	$f(x_k)$
Total	1

The **probability distribution** of a discrete of a random variable X is described as the function

$$f(x_i) = P(X = x_i)$$

which gives the probability for each value and satisfies:

1. $0 \leq f(x_i) \leq 1$ for each value x_i of X
2. $\sum_{i=1}^{k} f(x_i) = 1$

5.3 Expectation(Mean) and Standard Deviation of a Probability Distribution

The sample mean is calculated as

$$\bar{x} = \frac{\text{Sum of the observations}}{\text{Sample size}}.$$

The another calculation of sample mean illustrates the formula

$$\text{Sample mean } \bar{x} = \sum(\text{Value } \times \text{ Relative frequency}).$$

If we imagine a huge number of trials, the relative frequencies will approach the probability. The mean of the (infinite) collection of trials should be calculated as

$$\sum(\text{Value} \times \text{Probability}).$$

Define the mean of a random variable X or its probability distribution as

$$\sum(\text{Value} \times \text{Probability}) \quad \text{or} \quad \sum x_i f(x_i).$$

where x_i's denote the distinct values of X. The mean of a probability distribution is also called the population mean for the variable X.

The mean of a random values X is called its **expected value**.

The **mean** of X or **population mean**

$$\begin{aligned} E[X] &= \mu \\ &= \sum(\text{Value } \times \text{ Probability}) = \sum x_i f(x_i) \end{aligned}$$

Here the sum extends over all the distinct values x_i of X.

Since the mean μ is the center of the distribution of X, we express variation of X in term of the deviation $X - \mu$. We define the variance of X as the expected value of the squared deviation $(X - \mu)^2$

Variance and Standard Deviation of X

$$\sigma^2 = \text{Var}[X] = \sum (x_i - \mu)^2 f(x_i)$$
$$\sigma = \text{sd}[X] = +\sqrt{\text{Var}[X]}$$

Alternative Formula for Hand calculation

$$\sigma^2 = \sum x_i^2 f(x_i) - \mu^2$$

e.g.) Calculating a Population Variance and Standard Deviation Calculate the variance and the standard deviation of the distribution of X that appears in the left two columns of below table.

x	$f(x)$	$xf(x)$	$(x-\mu)$	$(x-\mu)^2$	$(x-\mu)^2 f(x)$	$x^2 f(x)$
0	.1	0	-2	4	.4	0
1	.2	.2	-1	1	.2	0.2
2	.4	.8	0	0	.0	1.6
3	.2	.6	1	1	.2	1.8
4	.1	.4	2	4	.4	1.6
Total	1.0	$2.0 = \mu$			$1.2 = \sigma^2$	$5.2 = \sum x^2 f(x)$

$$\text{Var}(X) = \sigma^2 = 1.2 \qquad \sigma^2 = 5.2 - (2.0)^2 = 1.2$$
$$\text{sd}(X) = \sigma = \sqrt{1.2} = 1.095 \qquad \sigma = \sqrt{1.2} = 1.095$$

5.4 Successes and Failures - Bernoulli Trails

5.4.1 Bernoulli Trials

- The sample space $S = \{ \text{S}, \text{F} \}$.
- The probability of success p =P(S), the probability of failure q =P(F).
- $0 \le p \le 1$, $q = 1 - p$.

Bernoulli Trials

1. Each trial yields one of two outcomes, technically called success (S) and failure (F).
2. For each trial, the probability of success $P(\text{S})$ is the same and is denoted by $p = P(\text{S})$. The probability of failure is then $P(\text{F}) = 1 - p$ for each trial and is denoted by q, so that $p + q = 1$.
3. Trials are independent. The probability of success in a trial remains unchanged given the outcomes of all the other trials.

5.4.2 Bernoulli Random Variable

- The random variable, $X(\text{S}) = 1$ and $X(\text{F})$=0, in $S = \{\ \text{S},\ \text{F}\ \}$.
- **Bernoulli Distribution**: The probability distribution of Bernoulli random variable.

x	0	1
$p(x)$	$1-p$	p

5.5 The Binomial Distribution

A **probability model** is an assumed form of the probability distribution that describes the chance behavior for a random variable X.
Probabilities are expressed in terms of relevant population quantities, called the **parameters**.

The Binomial Distribution

Denote

$$\begin{aligned} n &= \text{a fixed number of Bernolli trials} \\ p &= \text{the probability of success in each trial} \\ X &= \text{the (random) number of successes in } n \text{ trials} \end{aligned}$$

The random variable X called a **binomial random variable**. Its distribution is called a **binomial distribution**.

e.g.) An Example of the Binomial Distribution The elementary outcomes of 4 samples, the associated probabilities, and the value of X are listed as follows.

FFFF	SFFF	SSFF	SSSF	SSSS
	FSFF	SFSF	SSFS	
	FFSF	SFFS	SFSS	
	FFFS	FSSF	FSSS	
		FSFS		
		FFSS		

Value of X	0	1	2	3	4
Probability of each outcome	q^4	pq^3	p^2q^2	p^3q	p^4
Number of outcomes	$1 = \binom{4}{0}$	$4 = \binom{4}{1}$	$6 = \binom{4}{2}$	$4 = \binom{4}{1}$	$1 = \binom{4}{4}$

Value x	0	1	2	3	4
Probability $f(x)$	$\binom{4}{0}p^0q^4$	$\binom{4}{1}p^1q^3$	$\binom{4}{2}p^2q^2$	$\binom{4}{1}p^3q^1$	$\binom{4}{4}p^4q^0$

The **binomial distribution** with n trails and success probability p is described by the function

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

for the possible values $x = 0, 1, \cdots, n$.

5.5.1 The Mean and Standard Deviation of the Binomial Distribution

$$X = X_1 + X_2 + \cdots + x_n \sim B(n, p)$$

- $E[X] = E[X_1] + \cdots + E[X_n] = np$
- $\text{Var}[X] = \text{Var}[X_1] + \cdots + \text{Var}[X_n] = npq$

The binomial distribution with n trials and success probability p has

$$\begin{aligned} \text{Mean} &= np \\ \text{Variance} &= npq = np(1-p) \\ \text{sd} &= \sqrt{npq} \end{aligned}$$

5.6 Covariance and Correlation Coefficient of Two Random Variables X, Y

Let X, Y be a random variables. Then

1. The covariance of them:
$$Cov(X, Y) = E[(X - \mu_1)(Y - \mu_2)]$$
2. The correlation coefficient of them:
$$Corr(X, Y) = E\left[\left(\frac{X - \mu_1}{\sigma_1}\right)\left(\frac{Y - \mu_2}{\sigma_2}\right)\right] = \frac{Cov(X, Y)}{\text{sd}(X)\text{sd}(Y)}$$

Note that $-1 \leq Corr(X, Y) \leq 1$ and

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_1)(Y - \mu_2)] \\ &= E[XY - \mu_2 X - \mu_1 Y + \mu_1\mu_2] \\ &= E[XY] - \mu_2 E[X] - \mu_1 E[Y] + \mu_1\mu_2 \\ &= E[XY] - \mu_1\mu_2. \end{aligned}$$

That is, $Cov(X, Y) = E[XY] - \mu_1\mu_2$.

Properties of Covariance and Correlation Coefficient

1. $Cov(aX+b, cY+d) = ac \cdot Cov(X,Y)$
2. $Corr(aX+b, cY+d) = \begin{cases} Corr(X,Y) & \text{if } ac > 0 \\ -Corr(X,Y) & \text{if } ac < 0 \end{cases}$

Proof. (1)

$$\begin{aligned} Cov(aX+b, cY+d) &= E[(aX+b) - (a\mu_x + b) \cdot (cY + d - (c\mu_y + d))] \\ &= E[a(X-\mu_x) \cdot c(Y-\mu_y)] \\ &= acE[(X-\mu_x)(Y-\mu_y)] \\ &= ac \cdot Cov(X,Y). \end{aligned}$$

(2) Note that $\sigma_{aX+b} = \sqrt{\text{Var}(aX+b)} = \sqrt{a^2 \text{Var}(X)} = |a|\,\sigma_X$. Similarly $\sigma_{cY+d} = |c|\,\sigma_Y$.

$$\begin{aligned} Corr(aX+b, cY+d) &= \frac{Cov(aX+b, cY+d)}{\sigma_{aX+b}\sigma_{cY+d}} \\ &= \frac{ac \cdot Cov(X,Y)}{|a|\,\sigma_X |c|\,\sigma_Y} \\ &= \frac{ac}{|ac|} Corr(X,Y). \end{aligned}$$

Hence, $Corr(aX+b, cY+d) = \begin{cases} Corr(X,Y) & \text{if } ac > 0 \\ -Corr(X,Y) & \text{if } ac < 0 \end{cases}$ □

Distribution of Sum of Two Probability Variables

1. $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2Cov(X,Y)$
2. $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2Cov(X,Y)$

Two Probability Variables are Independent

1. $E[XY] = E[X] \cdot E[Y]$
2. $Cov(X, Y) = 0$, $Corr(X, Y) = 0$
3. $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$

Proof. (1)

$$\begin{aligned} E[XY] &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j p(x_i, y_j) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j p_1(x_i) p_2(y_j) \\ &= \sum_{i=1}^{\infty} x_i p_1(x_i) \sum_{j=1}^{\infty} y_j p_2(y_j) \\ &= E[X] \cdot E[Y]. \end{aligned}$$

(2) $Cov(X, Y) = E[XY] - E[X] \cdot E[Y] = 0$. □

6 The Normal Distribution

6.1 Probability Model for a Continuous Random Variable

Recall that a relative frequency histogram has the following properties:

1. The total area under the histogram is 1.
2. For two points a and b such that each is a boundary point of some class, the relative frequency of measurements in the interval a to b is the **area** under the histogram above this interval.

Because probability is interpreted as long-run relative requency, the curve obtained as the limiting form of the relative frequency histograms represents the manner in which the total probability 1 is distributed over the interval of possible values of the random variable X. This curve is called the **probability density curve** of the continuous random variable X. The mathematical function $f(x)$ whose graph produces this curve is called the **probability density function** of the continuous random variable X.

The **probability density function** $f(x)$ describes the distribution of probability for a continuous random variable. It has the properties:

1. The total area under the probability density curve is 1.
2. $P[a \leq X \leq b]$ = area under the probability density curve between a and b.
3. $f(x) \geq 0$ for all x.

With a continuous random variable, the probability that $X = x$ is **always** 0. It is only meaningful to speak about the probability that X lies in an interval.

For a continuous random variable X, $p(x)$ is called **probability density function of** X if $p(x)$ satisfies:

1. $p(x) \geq 0$, $\int_{-\infty}^{\infty} p(x)\ dx = 1$,

2. $P(a \leq X \leq b) = \int_{a}^{b} p(x)\ dx$.

Note that

- For any constant c, $\int_{c}^{c} p(x)\ dx = 0$.
- $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$.

Expectation of a Continuous Random Variable

- Expectation(or Mean) of a Random Variable X
  - a Discrete Random variable: $E[X] = \sum_{i=1}^{\infty} x_i p(x_i)$
  - a Continuous Random variable: $E[X] = \int_{-\infty}^{\infty} xp(x)\ dx$
- Expectation and Median of a Continuous Random Variable X
  - Expectation($\mu = E[X]$): the balance point of the probability mass.
  - Median: the value of X that divides the area under the curve into halves.

The population **100p-th percentile** is an x value that support area p to its left and $1 - p$ to its right.

Lower (first) quartile = 25th percentile
Second quartile (or median) = 50th percentile
Upper (third) quartile = 75th percentile

The **standardized variable**

$$Z = \frac{X - \mu}{\sigma} = \frac{\text{Variable - Mean}}{\text{Standard deviation}}$$

has mean 0 and sd 1.

6.2 The Normal Distribution - Its General Features

Notation. The normal distribution with a mean of μ and a standard deviation of σ is denoted by $N(\mu, \sigma^2)$.

6.3 The Standard Normal Distribution

The **standard normal distribution** has a bell-shaped density with

$$\text{Mean } \mu = 0$$
$$\text{Standard deviation } \sigma = 1$$

The standard normal distribution is denoted by $N(0, 1)$.

6.4 Probability Calculations with Normal Distributions

6.5 The Normal Approximation to the Binomial

(**The Normal Approximation to the Binomial**) When np and $np(1-p)$ are both large, say, greater than 15, the binomial distribution is well approximated by the normal distribution having mean $= np$ and sd $= \sqrt{np(1-p)}$. That is,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \text{ is approximately } N(0,1).$$

7 Variation in Repeated Samples - Sampling Distributions

A numerical feature of a population is called a **parameter**.

A **statistic** is a numerical valued function of the sample observations.

7.1 The Sampling Distribution of a Statistic

The probability distribution of a statistic is called its **sampling distribution**.

7.2 Distribution of the Sample Mean and the Central Limit Theorem

(**Mean and Standard Deviation of $\overline{X}$**) The distribution of the sample mean, based on a random sample of size n, has

$$E[\overline{X}] = \mu \qquad (= \text{Population mean})$$

$$\text{Var}[\overline{X}] = \frac{\sigma^2}{n} \qquad \left(= \frac{\text{Population variance}}{\text{Sample size}}\right)$$

$$\text{sd}[\overline{X}] = \frac{\sigma}{\sqrt{n}} \qquad \left(= \frac{\text{Population standard deviation}}{\sqrt{\text{Sample size}}}\right)$$

($\overline{X}$ **is Normal when Sampling from a Normal Population**) In random sampling from a **normal** population with mean μ and standard deviation σ, the sample mean $\overline{X}$ has the normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

8 Drawing Inferences from Large Samples

8.1 Introduction

Statistical inference deals with drawing conclusions about population parameters from an analysis of the sample data.

Any inference about a population parameter will involve some uncertainty because it is based on a sample rather than the entire population. To be meaningful, a statistical inference must include a specification of the uncertainty that is determined using the ideas of probability and the sampling distribution of the statistic.

The two most important types of inferences are (1) **estimation of parameter(s)** and (2) **testing of statistical hypotheses**. The true value of a parameter is an unknown constant that can be correctly ascertained only by an exhaustive study of the population, if indeed that were possible. Our objective may be to obtain a guess or an estimate of the unknown true value along with a determination of its accuracy. This type of inference is called **estimation of parameters**. An alternative objective may be to examine whether the sample data support or contradict the investigators conjecture about the true value of the parameter. This latter type of inference is called **testing of statistical hypotheses**.

Types of Statistical Inference

- **Estimation: What is the value of the population parameter?**
- **Hypothesis Testing: Is the parameter equal to a specific value?**

8.2 Point Estimation of a Population Mean

A statistic intended for estimating a parameter is called a **point estimator**(or simply an **estimator**). The standard deviation of an estimator is called its **standard error**: S.E.

When we estimate a population mean from a random sample, perhaps the most intuitive estimator is the sample mean,

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

In order to study the properties of the sample mean $\overline{X}$ as an estimator of the population mean m, let us review the C.L.T.

1. $E[\overline{X}] = \mu$.
2. $\text{sd}(\overline{X}) = \dfrac{\sigma}{\sqrt{n}}$ so $\text{S.E.}(\overline{X}) = \dfrac{\sigma}{\sqrt{n}}$.
3. With large n, $\overline{X}$ is nearly normally distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, i.e., $\overline{X} \sim N(\mu, \sigma^2/n)$.

Recall that, in a normal distribution, the interval running two standard deviations on either side of the mean contains probability .954. Thus, prior to sampling, the probability is .954 that the estimator $\overline{X}$ will be at most a distance $2\sigma/\sqrt{n}$ from the true parameter value μ. When we are estimating μ by $\overline{X}$, the 95.4% **error margin** is $2\sigma/\sqrt{n}$.

Notation.

$$z_{\alpha/2} = \text{Upper } \alpha/2 \text{ point of standard normal distribution}$$

That is, the area to the right of $z_{\alpha/2}$, and the area between $-z_{\alpha/2}$ and $z_{\alpha/2}$ is $1-\alpha$ (see Figure 1).

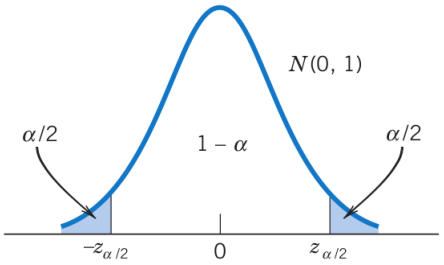

Figure 1: The notation $z_{\alpha/2}$.

$1-\alpha$	.80	.85	.90	0.95	0.99
$z_{\alpha/2}$	1.28	1.44	1.645	1.96	2.58

Point Estimation of the Mean

Parameter: Population mean μ.
Data: $X_1, X_2, \cdots, X_n$ (a random sample of size n)
Estimator: $\overline{X}$ (sample mean)

$$\text{S.E.}(\overline{X}) = \frac{\sigma}{\sqrt{n}} \qquad \text{Estimated S.E.}(\overline{X}) = \frac{S}{\sqrt{n}}$$

For large n, the $100(1-\alpha)\%$ error margin is $z_{\alpha/2}\sigma/\sqrt{n}$. (If σ is unknown, use S in place of σ.)

Note that $S = \sqrt{\dfrac{\sum(X-\overline{X})}{n-1}}$.

*** Determining the Sample Size**

Let

$$d = \text{ Desired error margin}$$

and

$$1-\alpha = \text{ Probability associated with error margin.}$$

Referring to the expression for a $100(1-\alpha)\%$ error margin, we then equate:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = d.$$

To be $100(1-\alpha)\%$ sure that error of estimation $\left|\overline{X}-\mu\right|$ does not exceed d, the **required sample size** is

$$n = \left[\frac{z_{\alpha/2}\sigma}{d}\right]^2$$

8.3 Confidence Interval for a Population Mean

Ideally, we would like to be able to collect a sample and then use it to calculate an interval that would definitely contain the true value of the parameter. This goal, however, is not achievable because of sample-to-sample variation. Instead, we insist that before sampling the proposed interval will contain the true value with a specified high probability. This probability, called **the level of confidence**, is typically taken as .90, .95, or .99.

The normal table shows that the probability is .95 that a normal random variable will lie within 1.96 standard deviation from its mean. For $\overline{X}$, we have

$$P\left[\mu - 1.96\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + 1.96\frac{\sigma}{\sqrt{n}}\right] = .95 \iff P\left[\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right] = .95.$$

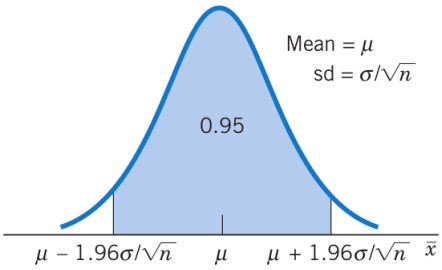

Figure 2: Normal distribution of $\overline{X}$.

We say that the interval

$$\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \quad \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

or its realization $\left(\overline{x} - 1.96\sigma/\sqrt{n},\ \overline{x} + 1.96\sigma/\sqrt{n}\right)$ is a **95% confidence interval for μ** when the population is normal and σ known.

We need not always tie our discussion of confidence intervals to the choice of a 95% level of confidence. An investigator may wish to specify a different high probability. We denote this probability by $1 - \alpha$ and speak of a $100(1 - \alpha)\%$ **confidence interval**.

In summary, when the population is normal and σ is known, a $100(1 - \alpha)\%$ confidence interval for μ is given by

$$\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

* Large Sample Confidence Intervals for μ

The central limit theorem then tells us that $\overline{X}$ is nearly normal whatever the form of the population. We find that the large sample confidence interval for μ has the form

$$\text{Estimate} \pm (z \text{ value})(\text{Estimated standard error}).$$

Large Sample Confidence Interval for μ

When n is large, a $100(1 - \alpha)\%$ confidence interval for μ is given by

$$\left(\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \quad \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

where S is the sample standard deviation.

*** Confidence Interval for a Parameter**

The concept of a confidence interval applies to any parameter, not just the mean.

Definition of a Confidence Interval for a Parameter

An interval (L, U) is a $100(1-\alpha)\%$ confidence interval for a parameter if

$$P\left[L < \text{ Parmeter } < U\right] = 1 - \alpha$$

and the endpoints L and U are computable from the sample.

8.4 Testing Hypotheses about a Population Mean

The formulation of a hypotheses testing problem and then the steps for solving it require a number of definitions and concepts. We will introduce these key statistical concepts

Null hypothesis and the **alternative hypothesis**
Type I and **Type II errors**
Level of significance
Rejection region
P-value

in the context a specific problem to help integrate them with intuitive reasoning.

*** Formulating the Hypothesis**

In the language of statistics, the claim or the research hypothesis that we wish to establish is called the **alternative hypothesis** H_1. The opposite statement, one that nullifies the research hypothesis, is called **the null hypothesis** H_0.

Formulation of H_0 and H_1

When our goal is to establish an assertion with substantive support obtain from the sample, the negation of the assertion is taken to be the null hypothesis H_0 and the assertion itself is taken to be the alternative hypothesis H_1.

Our initial question, "Is there strong evidence in support of the claim?" now translates to "Is there strong evidence for rejecting H_0?"

A **decision rule**, or a **test of the null hypothesis**, specifies a course of action by stating what sample information is to be used and how it is to be used in making a decision. Bear in mind that we are to make one of the following two decisions:

Decisions

Either

Reject H_0 and conclude that H_1 is substantiated

or

Retain H_0 and conclude that H_1 fails to be substantiated

*** Test Criterion and Rejection Region**

A reasonable decision rule should be of the form

$$\text{Reject } H_0 \text{ if } \overline{X} \leq c$$
$$\text{Retain } H_0 \text{ if } \overline{X} > c$$

This decision rule is conveniently expressed as $R : \overline{X} \leq c$, where R stands for the rejection of H_0. The set of outcomes $[\overline{X} \leq c]$ is called **rejection region** or **critical region**, and the cut-off point c is called the **critical value**.

The random variable $\overline{X}$ whose value serves to determine the action is called the **test statistic**.
A **test of the null hypothesis** is a course of action specifying the set of values of a test statistic $\overline{X}$, for which H_0 is to be rejected.
This set is called the **rejection region** of the test.

*** Two Types of Error and their Probabilities**

Decision Based on Sample	Unknown True Situation	
	H_0 True	H_1 True
Reject H_0	Wrong rejection of H_0 (Type I error)	Correct decision
Retain H_0	Correct decision	Wrong retention of H_0 (Type II error)

Two Types of Error

Type I error: Rejection of H_0 when H_0 is true
Type II error: Non-rejection of H_0 when H_1 is true

$\alpha =$ Probability of making a Type I error
$\beta =$ Probability of making a Type II error

Here, α also called the **level of significance**.

We hold a at a predetermined low level such as .10, .05, or .01 when choosing a rejection region. We will not pursue the evaluation of β, but we do note that if the β turns out to be uncomfortably large, the sample size must be increased.

*** P-value: How Strong Is a Rejection of H_0?**

The *P*-**value**(or **significance probability**) is the probability, calculated under H_0, that the test statistic takes a value equal to or more extreme than the value actually observed.

The *P*-**value** serves as a measure of the strength of evidence against H_0. A **small** *P*-**value** means that the null hypothesis is strongly rejected or the result is **highly statistically significant**.

The Steps for Testing Hypotheses

1. Formulate the null hypothesis H_0 and the alternative hypothesis H_1.
2. Test criterion: State the test statistic and the form of the rejection region.
3. With a specified a, determine the rejection region.
4. Calculate the test statistic from the data.
5. Draw a conclusion: State whether or not H_0 is rejected at the specified a and interpret the conclusion in the context of the problem. Also, it is a good statistical practice to calculate the P-value and strengthen the conclusion.

Large Sample Test for μ

When the sample size is large, a *Z* **test** concerning μ is based on the normal test statistic

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

The rejection region is one- or two-sided depending on the alternative hypothesis. Specifically,

$$\begin{array}{lll} H_1 : \mu > \mu_0 & & R : Z \geq z_\alpha \\ H_1 : \mu < \mu_0 & \text{requires} & R : Z \geq -z_\alpha \\ H_1 : \mu \neq \mu_0 & & R : |Z| \geq z_{\alpha/2} \end{array}$$

8.5 Inferences about a Population Mean

When n elements are randomly sampled from the population, the data will consist of the count X of the number of sampled elements possessing the characteristic. Common sense suggests the sample proportion

$$\hat{p} = \frac{X}{n}$$

as an estimator of p. The hat notation reminds us that $\hat{p}$ is a statistic.

When n is large, the binomial variable X is well approximated by a normal with mean np and standard deviation $\sqrt{npq}$. That is,

$$Z = \frac{X - np}{\sqrt{npq}}$$

is approximately standard normal. In summary, when n is large, $X \sim B(n, p) \approx N(np, np(1-p))$. Specially,

$$Z = \frac{(X - np)/n}{(\sqrt{npq})/n} = \frac{\hat{p} - p}{\sqrt{pq/n}} \approx N(0, 1).$$

It shows that $\hat{p}$ is approximately normally distributed with mean p and standard deviation $\sqrt{pq/n}$.

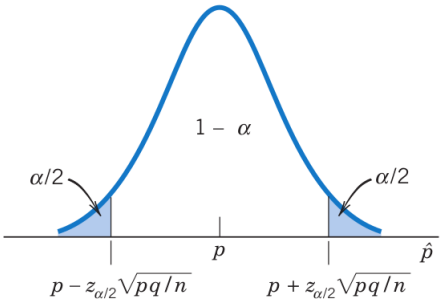

Figure 3: Approximate normal distribution of $\hat{p}$.

* Point Estimation of p

When the count X has a binomial distribution,

$$E[X] = np \qquad \text{sd}[X] = \sqrt{npq}.$$

Since $\hat{p} = X/n$, the properties of expectation give

$$E(\hat{p}) = p \qquad \text{sd}(\hat{p}) = \sqrt{pq/n}.$$

The second result shows that

$$\text{S.E.}(\hat{p}) = \sqrt{\frac{pq}{n}} \qquad \text{and} \qquad \text{estimated S.E.}(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

When n is large, prior to sampling, the probability is approximately .954 that the error of estimation $|\hat{p} - p|$ will be less than $2 \times$ (estimated S.E.). And when n is large

$$P\left[-z_{\alpha/2} \le \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \le z_{\alpha/2}\right] \approx 1 - \alpha.$$

Point Estimation of a Population Proportion

Parameter: Population proportion p.
Data: X = Number having the characteristic in a random sample of size n
Estimator: $\hat{p} = \dfrac{X}{n}$

$$\text{S.E.}(\hat{p}) = \sqrt{\frac{pq}{n}} \qquad \text{and} \qquad \text{estimated S.E.}(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

For large n, an approximate $100(1-\alpha)\%$ error margin is $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.

* Confidence Interval for p

Large Sample Confidence Interval for p

For large n, a $100(1-\alpha)\%$ confidence interval for p is given by

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}, \quad \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

* Determining the Sample Size

The required sample size is obtained by equating

$$z_{\alpha/2}\sqrt{pq/n} = d,$$

where d is the specified error margin. We then obtain

$$n = pq\left[\frac{z_{\alpha/2}}{d}\right]^2.$$

Without prior knowledge of p, pq can be replaced by its maximum possible value $1/4$ and n determined from the relation

$$n = \frac{1}{4}\left[\frac{z_{\alpha/2}}{d}\right]^2.$$

* Large Sample Tests about p

We consider testing $H_0 : p = p_0$ versus $H_1 : p \neq P_0$. With a large number of trials n, the sample proportion

$$\hat{p} = \frac{X}{n}$$

is approximately normally distributed. Under the null hypothesis, p has the specified value p_0 and the distribution of $\hat{p}$ is approximately $N(p_0, \sqrt{p_0 q_0/n})$. Consequently, the standardized statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0/n}}$$

has the $N(0, 1)$ distribution. Note that

(a) $H_1 : p > p_0 \Rightarrow P = P(Z \geq z), \quad R : z \geq z_\alpha$

(b) $H_1 : p < p_0 \Rightarrow P = P(Z \leq z), \quad R : z \leq -z_\alpha$

(c) $H_1 : p \neq p_0 \Rightarrow P = P(|Z| \geq |z|), \quad R : |z| \geq z_{\alpha/2}$

9 Small Sample Inferences for Normal Populations

9.1 Unbiased Estimators

(**Unbiased Estimators for Expectation and Variance**) Suppose $X_1, X_2, \cdots, X_n$ is a random sample from a distribution with finite expectation μ and finite variance σ^2. Then

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is an *unbiased estimator for* μ, i.e., $E[\overline{X}_n] = \mu$ and

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

is an *unbiased estimator for* σ^2, i.e., $E[S_n^2] = \sigma^2$.

Recall. Let $X_1, X_2, \cdots, X_n$ are random sample. Then the followings hold:

(i) $X_1, \cdots, X_n$ are mutually independent.

(ii) Distribution of all X_i = Population distribution.

This samples called **Independent and Identically Distributed (*i.i.d.*)** samples. Note that $\forall 1 \leq i \leq n$,

$$E[X_i] = \mu \quad \text{and} \quad \text{Var}[X_i] = \sigma^2.$$

Recall. $E[\overline{X}_n] = \mu$.

Proof.

$$\begin{aligned} E[\overline{X}_n] &= \frac{1}{n}E[X_1 + \cdots + X_n] \\ &= \frac{1}{n}(E[X_1] + \cdots + E[X_n]) \\ &= \frac{1}{n}(\mu + \cdots + \mu) = \mu. \end{aligned}$$

□

Now, we show that $E[S_n^2] = \sigma^2$.

Proof. Note that

$$E[S_n^2] = \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \overline{X}_n)^2].$$

Since $E[X_n] = \mu$, we have $E[X_i - \overline{X}_n] = 0$. Note that for any random variable Y with $E[Y] = 0$, we have

$$\text{Var}(Y) = E[Y^2] - (E[Y])^2 = E[Y^2].$$

Applying this to $Y = X_i - \overline{X}_n$, it follows that

$$E[(X_i - \overline{X}_n)^2] = \text{Var}(X_i - \overline{X}_n).$$

Note that we can write

$$X_i - \overline{X}_n = X_i - \frac{\sum X_i}{n} = \frac{nX_i - X_i - \sum_{j\neq i} X_j}{n} = \frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i} X_j.$$

Then

$$\begin{aligned}
\mathrm{Var}(X_i - \overline{X}_n) &= \mathrm{Var}\left(\frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i} X_j\right) \\
&= \frac{(n-1)^2}{n^2}\mathrm{Var}(X_i) + \frac{1}{n^2}\sum_{j\neq i}\mathrm{Var}(X_j) \text{ by independence of } X_j\text{'s} \\
&= \left[\frac{(n-1)^2}{n^2} + \frac{n-1}{n^2}\right]\sigma^2 = \frac{n-1}{n}\sigma^2.
\end{aligned}$$

Hence,

$$\begin{aligned}
E[S_n^2] &= \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \overline{X}_n)^2] = \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{Var}(X_i - \overline{X}_n) \\
&= \frac{1}{n-1}\cdot n \cdot \frac{n-1}{n}\sigma^2 = \sigma^2.
\end{aligned}$$

□

9.2 Student's t Distribution

Student's t Distribution

If $X_1, \cdots, X_n$ is a random sample from a normal population $N(\mu, \sigma)$ and

$$\overline{X} = \frac{1}{n}\sum X_i \qquad \text{and} \qquad S^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}$$

then the distribution of

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

is called **Student's t distribution with $n-1$ degrees of freedom**.

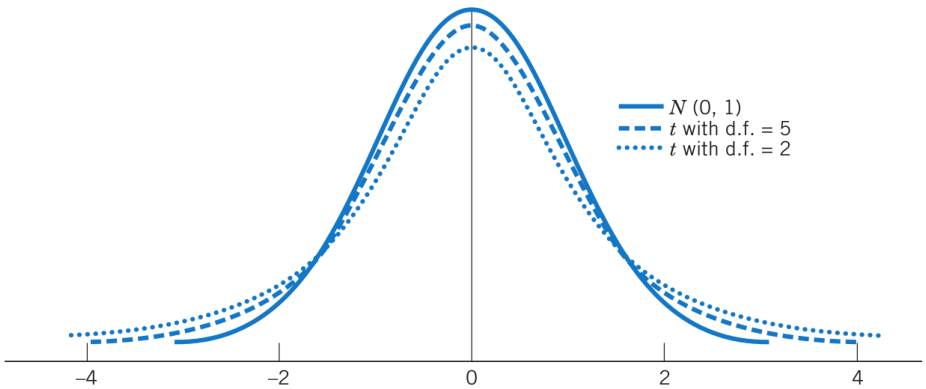

Figure 4: Comparison of $N(0,1)$ and t density curves.

9.3 Inferences about μ - Small Sample Size

9.3.1 Confidence Interval for μ

The distribution of

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

provides the key for determining a **confidence interval for μ**, the mean of a normal population.

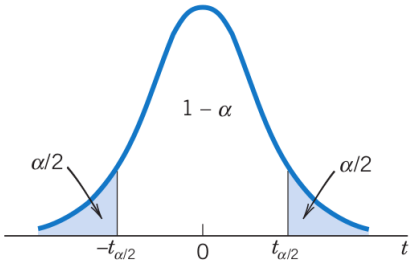

Figure 5: $t_{\alpha/2}$ and the probabilities.

Since $\dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ has the t distribution with d.f. $= n - 1$, we have

$$\begin{aligned} 1 - \alpha &= P\left[-t_{\alpha/2} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}\right] = P\left[-t_{\alpha/2}\frac{S}{\sqrt{n}} < \overline{X} - \mu < t_{\alpha/2}\frac{S}{\sqrt{n}}\right] \\ &= P\left[\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right]. \end{aligned}$$

A $100(1-\alpha)\%$ Confidence Interval for a Normal Population Mean

$$\left(\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \quad \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

where $t_{\alpha/2}$ is the upper $\alpha/2$ point of the t distribution with d.f. $= n - 1$.

9.3.2 Hypothesis Tests for μ

Hypotheses Tests for μ - Small Samples

To test $H_0 : \mu = \mu_0$ concerning the mean of a **normal population**, the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

which has Student's t distribution with $n - 1$ degrees of freedom:

$$\begin{aligned} H_1 &: \mu > \mu_0 \qquad R : T \geq t_\alpha \\ H_1 &: \mu < \mu_0 \qquad R : T \leq -t_\alpha \\ H_1 &: \mu \neq \mu_0 \qquad R : |T| \geq t_{\alpha/2} \end{aligned}$$

The test is called a **Student's t test** or simply a **t test**.

9.4 Relationship between Tests and Confidence Intervals

A $100(1-\alpha)\%$ confidence interval for μ is $\left(\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}},\ \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$. On the other hand, the rejection region of a level α test for $H_0 : \mu = \mu_0$ versus the two-sided alternative $H_1 : \mu \neq \mu_0$ is

$$R : \left|\frac{\overline{X} - \mu_0}{S/\sqrt{n}}\right| \geq t_{\alpha/2}.$$

Reversing the inequality in R, we obtain Acceptance region: $-t_{\alpha/2} < \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} < t_{\alpha/2}$ which can also be written as

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu_0 < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}.$$

9.5 Inferences about the Standard Deviation σ (The χ^2 Distribution)

χ^2 Distribution

Let $X_1, \cdots, X_n$ be a random sample from a normal population $N(\mu, \sigma)$. Then the distribution of

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

is called the **χ^2 distribution with $n-1$ degrees of freedom**.

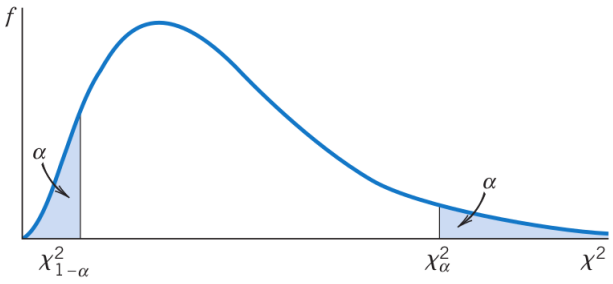

Figure 6: Probability density curve of a χ^2 distribution.

The χ^2 is the basic distribution for constructing intervals for σ^2 or σ. We outline the steps in terms of a 95% confidence interval for σ^2. Dividing the probability $\alpha = .05$ equally between th two tails of the χ^2 distribution,

$$P\left[\chi^2_{.975} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{.025}\right] = .95$$

where the percentage points are read from the χ^2 table at d.f. $= n-1$. Also we have

$$P\left[\frac{(n-1)S^2}{\chi^2_{.025}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{.975}}\right] = .95.$$

Therefore, for a confidence level .95, the interval for σ becomes

$$\left(S\sqrt{\frac{n-1}{\chi^2_{.025}}},\ S\sqrt{\frac{n-1}{\chi^2_{.975}}}\right).$$

9.6 Robustness of Inference Procedures

The small sample methods for both confidence interval estimation and hypothesis testing presuppose that the sample is obtained from a normal population. Users of these methods would naturally ask:

1. What method can be used to determine if the population distribution is nearly normal?
2. What can go wrong if the population distribution is non-normal?
3. What procedures should be used if it is non-normal?
4. If the observations are not independent, is this serious?

1. To answer the first question, we could construct the dot diagram or normal-scores plot.

2. Fortunately, the effects on inferences about μ using the t statistic are not too serious if the sample size is at least moderately large (say, 15). In larger samples, such disturbances tend to disappear due to the central limit theorem. We express this fact by saying that **inferences about μ using the t statistic are reasonably "robust"**.

3.

4.