MATH3200 Intermediate Statistics and Data Analysis

Albert Peng

May 10, 2022

Contents

1 Basic Statistical Concepts

1.1 Averages

Let $v_1, ..., v_N$ denote the values in the population. The *population average (mean)* is

$$\mu = \frac{1}{N}\sum_{i=1}^{N} v_i$$

Population mean can also be described as the *expected value* of X, where X is a random variable, value of a randomly selected population.

$$E(X) = \mu$$

Let $x_1, ..., x_n$ denote the values of our variable of interest in a random sample. The *sample mean* or *sample average* is

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

1.2 Variance and Standard Deviation

The *population variance* mesures the mount of intrinsic variability in the population.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(v_i - \mu)^2$$

The *sample variance* measures the amount of variability in the sample.

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

2 Introduction to Probability

2.1 Sample Spaces, Events, and Set Operations

An *experiment* is any action whose outcome is random and results in well-defined outcome.

A *sample space* is the set of all possible outcomes of an experiment, denoted by S.

An *event* is a subset of the sample space.

- Event with one outcome is a *simple event*
- Event that contains more than one outcome is *compound event*

Set operations are also used to represent events:

- *Union* of events is represented by $A \cup B$
- *Intersection* of events is represented by $A \cap B$
- *Complement* of event A is represented by A^c
- *Difference* of events is represented by $A - B$ or $A \cap B^c$
- *Disjoint* or *mutually exclusive* events if they have no outcomes in common, $A \cap B = \emptyset$
- A is a *subset* of B ($A \subseteq B$) if outcomes of A are also in B

Commutative Laws:

$$A \cup B = B \cup A \text{ and } A \cap B = B \cap A$$

Associative Laws:

$$(A \cup B) \cup C = A \cup (B \cup C) \text{ and } (A \cap B) \cap C = A \cap (B \cap C)$$

Distributive Laws:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \text{ and } (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

DeMorgan's Laws:

$$(A \cup B)^c = A^c \cap B^c \text{ and } (A \cap B)^c = A^c \cup B^c$$

2.2 Equally Likely Outcomes

The *probability* of an event E is the likelihood of the occurrence of E denoted as $P(E)$

There are N outcomes and $N(E)$ denotes the number of outcomes in event E. Then:

$$P(E) = \frac{N(E)}{N}$$

Permutations have ordered outcomes. Number of permutations of k units selected from a group of n units is denoted by $P_{k,n}$:

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

Combinations have unordered outcomes. Number of combinations of k units selected from a group of n units is denoted by $\binom{n}{k}$

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

2.3 Axioms and Properties of Probability

For an experiment with sample space S, probability is a function that assigns a number $P(E)$ to an event so that the following axioms hold:

1. $0 \leq P(E) \leq 1$

2. $P(S) = 1$

3. For any sequence of disjoint events $E_1, E_2, ...$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

In addition, the following properties of probability are also useful:

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

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

2.4 Conditional Probability

For any two events A and B with $P(A) > 0$, the *conditional probability* of B given A, denoted by $P(B|A)$, is

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

The following properties also apply:

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

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B)$$

Theorem. *Law of Total Probability* is a formula for computing the probability of an event B when B arises in connection with a partition of the sample space, such that if $A_1, ..., A_k$ constitute a partition of the sample space,

$$P(B) = P(A_1)P(B|A_1) + ... + P(A_k)P(B|A_k)$$

Theorem. *Bayes Theorem* is used in the same context as the law of total probability.

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

2.5 Independent Events

Events A and B are *independent* if knowledge that A occurred does not change the probability of B occurring. In that case,

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

Some properties include:

- If A and B are independent, so are A and B^c
- S and $\emptyset$ are independent of every other event
- Disjoint events are not independent unless the probability of one of them is 0

2.5.1 Mutual Independence

$E_1, ..., E_n$ are *mutually independent* if for every subset $E_{i_1}, ..., E_{i_k}$, $k \leq n$,

$$P(E_{i_1} \cap ... \cap E_{i_k}) = P(E_{i_1})...P(E_{i_k})$$

3 Chapter 3: Random Variables and Their Distributions

3.1 Random variables

A *random variable* is a function that associates a number with each outcome of the sample space of a random experiment.

The *probability distribution* of a random variable specifies how the total probability is distributed.

3.1.1 Cumulative Distributive Function

Cumulative Distribution Function of a random variable X is a function

$$F(x) = P(X \leq x)$$

CDF has the following properties:

- $F(x)$ is a non-decreasing function
- $F(-\infty) = 0$ and $F(\infty) = 1$
- If $a \leq b$, then $P(\leq X \leq b) = F(b) - F(a)$

The *probability density function (PDF)* of a continuous random variable X is a nonnegative function f such that

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

The PDF can also be obtained by the CDF, where

$$f(x) = F'(x) = \frac{d}{dx}F(x)$$

3.2 Parameters of Probability Distributions

3.2.1 Expected Value, Variance, and Standard Deviations

Th *expected value* for a continuous random variable X with PDF $f(x)$ is:

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

If X is continuous with PDF $f_X(x)$, the expected value of $Y = h(x)$ is:

$$E(h(X)) = \int_{-\infty}^{\infty} h(x) f_X(x)$$

Note: If $Y = aX + b$, then $E(Y) = aE(X) + b$. The *variance* σ_X^2 or $\text{Var}(X)$, of a random variable X is

$$\sigma_X^2 = E[(X - \mu_X)^2] = E(X^2) - E(X)^2$$

The *Standard Deviation* of X is the positive square root of the variance such that $\sigma_X = \sqrt{\sigma_X^2}$

3.2.2 Population Percentiles

Let X be a continuous random variable with CDF F and let α be a number between 0 and 1. The $100(1-\alpha)$-th *percentile* of X is x_a such that

$$F(x_a) = P(X \leq x_a) = 1 - x_a$$

3.3 Model for Discrete Random Variables

3.3.1 Bernoulli Distribution

A *Bernoulli trial* is an experiment whose outcome is either a success or a failure, where 1 stands for success and 0 stands for failure. This is denoted by $X \sim Bern(p)$.

3.3.2 Binomial Distribution

A *binomial experiment* is when n Bernoulli experiments, each having probability of success p, are performed independently.

The *binomial random variable* Y is the number of successes in the n Bernoulli trials, denoted as $Y \sim Bin(n, p)$, where:

$$p(y) = P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

For binomial distribution, the expected value and variance are calculated as

$$E(Y) = np \qquad \sigma_Y^2 = np(1-p)$$

R command for computing PMF and CDF is `dbinom`(y, n, p) and `pbinom`(y, n, p)

3.3.3 Hypergeometric Distribution

Suppose a population consists of M_1 objects labeled 1 and M_2 objects labeled 0, and that a sample of size n is selected at random **without replacement**.

The *hypergeometric random variable* X is the number of objects labeled 1 in the sample, denoted as $X \sim Hyp(M_1, M_2, n)$, where:

$$p(x) = P(X = x) = \frac{\binom{M_1}{x}\binom{M_2}{n-x}}{\binom{M_1+M_2}{n}}$$

In this case, the sample space of X:

$$S_x = \{\max(0, n - M_2), ..., \min(n, M_1)\}$$

When $N = M_1 + M_2$, the expected value and the variance is

$$E(X) = \frac{nM_1}{N} \qquad \sigma_X^2 = \frac{nM_1}{N}\left(1 - \frac{M_1}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Note: For large population size N, the difference between sampling with and without replacement is very small.

R command for computing PMF and CDF is `dhyper`(x, M_1, M_2, n) and `phyper`(x, M_1, M_2, n)

3.3.4 Geometric Distribution

In a *geometric experiment*, individual Bernoulli trials with probability of success p are performed until the first success occurs.

The *geometric random variable* X is the number of trials up to and including the first success, denoted as $X \sim Geo(p)$, where

$$p(x) = P(X = x) = (1-p)^{x-1}p$$

$$F(x) = P(X \leq x) = 1 - (1-p)^x$$

The expected value and variance would be:

$$E(x) = \frac{1}{p} \qquad \sigma_X^2 = \frac{1-p}{p^2}$$

3.3.5 Negative Binomial Distribution

In a *negative binomial experiment*, independent Bernoulli trials, each with probability of success p, are performed until the rth success occurs.

The *negative binomial random variable* Y is the total number of trials up to and including the rth success, denoted as $Y \sim NB(r, p)$, where:

$$p(y) = P(Y = y) = \binom{y-1}{r-1}p^r(1-p)^{y-r}$$

The expected value and variance are

$$E(Y) = \frac{r}{p} \qquad \sigma_Y^2 = \frac{r(1-p)}{p^2}$$

R command for computing PMF and CDF is `dnbinom`$(y - r, r, p)$ and `pnbinom`$(y - r, r, p)$

3.3.6 Poisson Distribution

The Poisson distribution is used to model the probability that a number of events occur in an interval of time or space.

The *poisson random variable* X denotes the number of events that occurred, denoted by $X \sim Poisson(\lambda)$, where

$$p(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

The expected value and variance are

$$E(X) = \lambda \qquad \sigma_X^2 = \lambda$$

R command for computing PMF and CDF is `dpois`(x, λ) and `ppois`(x, λ)
We can also find a sample of n Poissoon random variables using `rpois`(n, λ)

3.4 Models for Continuous Random Variables

3.4.1 Exponential Distribution

The *exponential distribution* is often used to model lifetimes of equipment or waiting times until events are over, denoted as $X \sim Exp(\lambda)$.

The PDF is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, 0 \leq x \leq 1 \\ 0, \text{ otherwise} \end{cases}$$

Therefore, the CDF is

$$F(x) = 1 - e^{-\lambda x}$$

The expected value and variance are

$$E(X) = \frac{1}{\lambda} \qquad \sigma_X^2 = \frac{1}{\lambda^2}$$

The *memoryless property* of exponential random variable X:

$$P(X > s + t | X > s) = P(X > t)$$

3.4.2 Normal Distribution

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \text{ and } \Phi(z) = \int_{-\infty}^{z} \phi(y)dy$$

A random variable X has a *normal distribution* with parameters μ and σ^2 is denoted as $X \sim N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sigma} \cdot \phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(x - \mu)^2}{2\sigma}\right)$$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

The mean and variance is therefore:

$$E(X) = \mu \qquad Var(X) = \sigma^2$$

Suppose $X \sim N(\mu, \sigma^2)$. Then $a + bX \sim N(a + b\mu, b^2\sigma^2)$ has

$$E(a + bX) = a + bE(X) = a + b\mu \quad Var(a + bX) = b^2Var(X) = b^2\sigma^2$$

R command for computing

- the PDF and CDF is `dnorm`(x, μ, σ) and `pnorm`(x, μ, σ)
- the s100th percentile is `qnorm`(s, μ, σ)
- a random sample of size n is `rnorm`(n, μ, σ)

Q-Q Plots plot the sample percentiles against the percentiles from the normal distribution.

4 Chapter 4: Joint Probability Distributions

4.1 Describing Joint Variable Distributions

4.1.1 Joint and Marginal PMF

The *joint probability mass function (joint PMF)* or the jointly discrete random variables X and Y is

$$p(x,y) = P(X = x, Y = y)$$

If the sample space of (X,Y) is $S = \{(x_1,y_1),(x_2,y_2),...\}$, then

$$p(x_i,y_i) \geq 0 \text{ for all } i \quad \textbf{and} \quad \sum_{(x_i,y_i)\in S} p(x_i,y_i) = 1$$

$$P(a < X \leq b, c < Y \leq d) = \sum_{i: a < x_i \leq b, c < y_i \leq d} p(x_i,y_i)$$

The distributions of the individual random variables are called *marginal distributions* and can be found using joint PMF:

$$p_X(x) = \sum_{y\in S_Y} p(x,y) \qquad p_Y(y) = \sum_{x\in S_X} p(x,y)$$

4.1.2 Joint and Marginal PDFs

The *joint probability distribution density function* of the jointly continuous random variables X and Y is the non-negative function $F(X,Y)$ with the property that

$$P((X,Y)\in A) = \int\int_A f(x,y)dx\,dy$$

$f(x,y)$ has to satisfy the condition that

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

Marginal PDF can be found using the joint PDF

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

Example. Let (X,Y) be jointly continuous random variables with the following joint PDF. Verify it is a valid PDF and find $P(X > Y)$

$$f(x,y) \begin{cases} \frac{12}{7}(x^2+xy), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & elsewhere \end{cases}$$

$$\int_0^1 \int_0^1 \frac{12}{7}(x^2 + xy)\,dx\,dy = \frac{12}{7}\int_0^1 \left(\frac{x^3}{3} + \frac{x^2 y}{2}\right)\Bigg|_0^1 dy = \frac{12}{7}\left(\frac{1}{3}y + \frac{y^2}{4}\right)\Bigg|_0^1 = 1$$

To find $P(X > Y)$, $0 \leq Y < X \leq 1$:

$$\int_0^1 \int_y^1 \frac{12}{7}(x^2 + xy)\,dx\,dy = \int_0^1 \int_0^x \frac{12}{7}(x^2 + xy)\,dy\,dx$$

$$= \frac{12}{7}\int_0^1 \left[x^2 y + \frac{1}{2}xy^2\right]_{y=0}^{y=x} dx = \frac{12}{7}\int_0^1 \frac{3}{2}x^3\,dx = 0.643$$

To find the marginal PDF of X:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{12}{7}(x^2 + xy)\,dy = \frac{12}{7}\left(x^2 + \frac{1}{2}x\right)$$

4.2 Conditional Distributions: PMF and PDF

For jointly discrete random variables X and Y, the *conditional PMF* of Y given $X = x$ is

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

Note:

$$p_Y(y) = \sum_{x \in S_x} p(x,y) = \sum_{x \in S_x} p_{y|X=x}(y) \cdot p_X(x)$$

For jointly continuous random variables X and Y, the *conditional PDF* of Y given $X = x$ is

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

Note:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X=x}(y) f_X(x)\,dx$$

4.3 Independent Random Variables

Two random variables are *independent* if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

This could be expanded to discrete and continuous cases, where X and Y are independent only if

$$p(x,y) = p_X(x) \cdot p_Y(y) \qquad f(x,y) = f_X(x) \cdot f_Y(y)$$

Theorem. If X and Y are jointly discrete, then X and Y are independent if and only if (This also holds for jointly continuous random variables PDF):

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

Theorem. Let X and Y be independent. Then

- $E(Y|X = x) = E(Y)$ does not depend on the value of x.

- $g(X)$ and $h(Y)$ are independent

- $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

If $X_1, ..., X_n$ are *independent and identically distributed*, or *iid*, if they are independent and have the same distribution.

4.4 Expected Value of Functions of Random Variables

We can model the expected value as following:

$$E[h(x,y)] = \int_{\infty}^{\infty}\int_{\infty}^{\infty} h(x,y)f(x,y)\,dx\,dy$$

Corollary. If $X_1, ..., X_n$ have the same mean $\mu = E(X_i)$, then

$$E\left(\sum_{i=1}^{n} X_i\right) = n\mu \qquad E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu$$

Corollary. Suppose $X_1, ..., X_n$ are iid Bern(p). Then $E(\hat{p}) = p$, where $\hat{p}$ is the sample proportion of successes (number of successes in $X_1, ..., X_n$ divided by n) .

4.5 Covariance

Var$(X + Y)$ =Var(X)+Var(Y) only if X and Y are independent.

When random variables X and Y are dependent, computing Var$(X + Y)$ involves the *covariance*.

$$\text{Cov}(X,Y) = \sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y$$

If Cov$(X,Y) > 0$, then greater values of X mainly correspond to greater values of Y.
If Cov$(X,Y) < 0$, then greater values of X mainly correspond to lesser values of Y.

Properties

- Cov(X,Y) =Cov(Y,X)

- Cov(X,X)=Var(X)

- If X and Y are independent, then Cov$(X,Y) = 0$

- Cov$(aX + b, cY + d) = ac$Cov(X,Y) for any real numbers a, b, c, and d.

4.5.1 Variance of Sums and Random Variables

Let σ_1^2 and σ_2^2 be the variances of X_1 and X_2, respectively.

- If X_1 and X_2 and independent,

$$Var(X_1 + X_2) = \sigma_1^2 + \sigma_2^2 \qquad Var(X_1 - X_2) = \sigma_1^2 + \sigma_2^2$$

- If X_1 and X_2 are dependent.

$$Var(X_1+X_2) = \sigma_1^2+\sigma_2^2+2\text{Cov}(X_1, X_2) \qquad Var(X_1-X_2) = \sigma_1^2+\sigma_2^2-2\text{Cov}(X_1, X_2)$$

- If $X_1, ..., X_n$ are random variables with variances $\sigma_1^2, ..., \sigma_n^2$

$$Var(a_1X_1 + ... + a_nX_n) = a_1^2\sigma_1^2 + ... + a_n^2\sigma_n^2 + \sum_i \sum_{j\neq i} a_i a_j \text{Cov}(X_i, Y_i)$$

Corollary. Let $X_1, ..., X_n$ be iid with common variance σ^2. Then

$$Var\left(\sum_{i=1}^n X_i\right) = n\sigma^2 \qquad Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n}$$

Corollary. Suppose $X_1, ..., X_n$ are iid Bern(p). Then

$$Var(\hat{p}) = \frac{1(11-p)}{n}$$

4.6 Quantifying Dependence

4.6.1 Pearson's Correlation Coefficient

Two random variables X and Y are *positively dependent* if larger values of X are associated with larger values of Y and are *negatively dependent* if larger values of X are associated with smaller values of Y.

However, covariance is not scale-free as it depends on the units. The *correlation coefficient* of X and Y solves this problem.

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

Properties

- For constants $a, b, c,$ and d

$$\text{Corr}(aX + b, cY + d) = sign(ac) \times \text{Corr}(X, Y)$$

- $-1 \leq$ Corr(X, Y) ≤ 1
- If X and Y are independent, Corr$(X, Y) = 0$
- Corr$(X, Y) = \pm 1$ iff $Y = aX + b$ for constants $a \neq 0$ and b. This is thus a measure of linear dependence.

Example. Let X_1, X_2 be iid $N(0, 1)$ and let $Y = 4X_1 + X_2$. Find Cov(X_1, Y)

$$\begin{aligned}\text{Cov}(X_1, Y) &= E(X_1 Y) - E(X_1)E(Y) = E(X_1 Y) \\ &= E[X_1(4X_1 + X_2)] = 4E(X_1^2) + E(X_1 X_2) = 4E(X_1^2) \\ &= 4(Var(X_1) + E(X_1)^2) = 4\end{aligned}$$

Note: Corr$(X, Y) = 0$ does not mean X and Y are independent.

4.6.2 Sample Covariance and Correlation

If $(X_1, Y_1), \ldots, (X_n, Y_n)$ are samples from bivariate distribution of (X, Y), the *sample covariance* and the *sample correlation coefficient* are

$$S_{X,Y} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \qquad r = r_{X,Y} = \frac{S_{X,Y}}{S_X S_Y}$$

5 Chapter 5: Approximation Results

5.1 Law of Large Numbers

The Law of Large Numbers Let $X_1, ..., X_n$ bee iid and let g be a function such that $-\infty < E[g(X_1)] < \infty$. =. Then for any $\epsilon > 0$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - E[g(X_1)]\right| > \epsilon\right) \to 0 \text{ as } n \to \infty$$

This means that $\frac{1}{n}\sum_{i=1}^{n} g(X_i)$ converges in probability to $E[g(X_1)]]$. If $-\infty < E(X_1) < \infty$, then $\bar{X}$ converges in probability to $E(X_1)$.

We call $\bar{X}$ a *consistent estimator* of $E(X_1)$. Since $\hat{p}$ is also a sample mean, we have $\hat{p}$ converges in probability to p.

Limitations of LLN:

- as the sample size increases, sample averages approximate the population mean $E(X)$ more closely
- LLN provides no guidance regarding the quality of the estimation.

5.2 Convolutions

The *convolution* of two independent independent random variables refers to the distribution of their sum.

For example, let X and Y be independent random variables.

- If $X \sim Bin(n_1, p)$ and $Y \sim Bin(n_2, p)$, then $X + Y \sim Bin(n_1 + n_2, p)$
- If $X \sim Poisson(\lambda_1)$ and $Y \sim Poisson(\lambda_2)$, then $X + Y \sim Poisson(\lambda_1 + \lambda_2)$
- If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Corollary. Let $X_1, ..., X_n$ be independent random variables with $X_i \sim N(\mu_i, \sigma_i^2)$ and let $Y = a_1X_2 + ... + a_nY_n$ for constants then $a_1, ..., a_n$. Then

$$Y = N(\mu_Y, \sigma_Y^2), \qquad \text{where } \mu_Y = a_1\mu_1 + ... + a_n\mu_n \text{ and } \sigma_Y^2 = a_1^2\sigma_1^2 + ... + a_n^2\sigma_n^2$$

Corollary. Let $X_1, ..., X_n$ be iid $N(\mu, \sigma^2)$ and let $\bar{X} = \frac{1}{n}X_1 + ... + \frac{1}{n}X_n$ be the sample mean.

$$\bar{X} \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}}^2) = N\left(\mu, \frac{\sigma^2}{n}\right)$$

Example. Suppose we want to estimate the mean of a normal population whose variance is known to be 4. What sample size should be used to ensure that $\bar{X}$ lies within

0.5 units of the population mean with probability 0.9?

$$P(-0.5 < \bar{X} - \mu < 0.5) = 0.9, \qquad \bar{X} \sim N(\mu, \frac{\sigma^2}{n}), \text{ so standard deviation is} \frac{\sigma}{\sqrt{n}}$$

$$= P\left(-\frac{0.5}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{0.5}{\sigma/\sqrt{n}}\right) = P\left(-\frac{0.5}{\sigma/\sqrt{n}} < z < \frac{0.5}{\sigma/\sqrt{n}}\right) \text{ where } z \sim N(0,1)$$

5.3 Central Limit Theorem (CLT)

Central Limit Theorem (CLT): Let $X_1, ..., X_n$ be iid with finite mean μ and finite variance σ^2. Then for large enough n,

- $\bar{X}$ has approximately a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

$$\bar{X} \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

- $T = X_1 + ... + X_n$ has approximately a normal distribution with mean $n\mu$ and variance $n\sigma^2$.

$$T = X_1 + ... + X_n \dot{\sim} N(n\mu, n\sigma^2)$$

DeMoivre-Laplace Theorem: If $T \sim Bin(n,p)$, then for large enough n,

$$T \dot{\sim} N(np, np(1-p))$$

Continuity Correction: Let $T \sim Bin(n,p)$ and let $Y \sim N(np, np(1-p))$

$$P(T \leq k) \approx P(Y \leq k + 0.5) \qquad P(T = k) \approx P(k - 0.5 < Y < k + 0.5)$$

6 Chapter 6: Estimation

The goal of estimation is to estimate unknown population parameters.

Point estimation estimates an unknown parameter with a single value.

Notation:

- θ is the unknown population of interest.
- $X_1, ..., X_n$ is the sample before values or data.
- $x_1, ..., x_n$ are the observed sample values or data.
- $\hat{\theta}$ is a quantity used to estimate θ.
- $\hat{\theta}(X_1, ..., X_n)$ is called an *estimator* and are used for random variables.
- $\hat{\theta}(x_1, ..., x_n)$ is called an *estimate* for fixed values.

An estimator $\hat{\theta}$ of θ s unbiased is *unbiased* for θ if

$$E_\theta(\hat{\theta}) = \theta$$

The *bias* of $\hat{\theta}$ is

$$\text{bias}(\hat{\theta}) = E_\theta(\hat{\theta}) - \theta$$

The *standard error* of an estimator $\hat{\theta}$ is

$$\sigma_{\hat{\theta}} = \sqrt{Var_\theta(\hat{\theta})} = \sqrt{Var(\theta)}$$

It is important to note that by the central limit theorem, Var($\bar{X}$)=$\frac{\sigma^2}{n}$.

The *estimated standard error* is denoted by $S_{\hat{\theta}}$.

6.1 Mean Squared Error

The *men squared error* (MSE) of an estimator $\hat{\theta}$ for θ is

$$MSE(\hat{\theta}) = E_\theta[(\hat{\theta} - \theta)^2] = Var_\theta(\hat{\theta}) + [bias(\hat{\theta})]^2$$

For unbiased estimators, the preferred estimator is the one with the smallest variance.

6.2 Models of Moments Estimation

Method of moments estimation is a model based estimation technique, where we assume the population follows a known distribution.

The kth *theoretical moment:*

$$\mu_k = E(X^k)$$

The kth *sample moment:*

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

6.3 Maximum Likelihood Estimation

Maximum likelihood estimation is a model based estimatioon technique that maximizes the probability of the observed data.

Suppose $X_1, ..., X_n$ are iid random variables with PDF $f(x_i|\theta)$ or PMF $p(X_i|\theta)$. The likelihood function is

$$\text{lik}(\theta) = \prod_{i=1}^{n} f(x_i|\theta) \text{ or } (\theta) = \prod_{i=1}^{n} p(x_i|\theta)$$

The value of θ that maximizes lik(θ) is the maximum likelihood estimator (MLE) $\hat{\theta}$. Thus:

$$\log[\text{lik}(\theta)] = \log\left[\prod_{i=1}^{n} f(x_i|\theta)\right] = \sum_{i=1}^{n} \log[f(x_i|\theta)]$$

7 Chapter 7: Introduction to Confidence Intervals

We can use the sample mean $\bar{X}$ as a point estimate for the population mean μ. By the LLN, $\bar{X}$ approximates more accurately as the sample size increases. CLT allows us to assess the probability that $\bar{X}$ will be within a certain distance from μ. However, using only a point estimate is not as informative as we want it to be.

Confidence Interval is an interval for which we can assert, with a given degree of confidence or certainty, that it includes the true value of the parameter being estimated.

7.1 Z Confidence Intervals

Confidence intervals that use percentiles from the standard normal distribution.

$$z_a = 100(1-\alpha)\text{-th percentile}$$

Let $X_1, ..., X_n$ be iid random variables with parameter θ. By CLT, many estimators $\hat{\theta}$ of θ will be approximately normally distributed.

This results in a 95% confidence interval for θ:

$$\hat{\theta} - 1.96 S_{\hat{\theta}} \leq \theta \leq \hat{\theta} + 1.96 S_{\hat{\theta}}$$

The general $(1-\alpha)100\%$ confidence interval for θ is

$$\hat{\theta} - z_{\alpha/2} S_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2} S_{\hat{\theta}}$$

7.2 T Confidence Intervals

T Confidence intervals are intervals that use percentiles from the *T distribution*. T_v is a random variable from the T distribution with v degrees of freedom (*df*). As the degrees of freedom increases, the T density tends toward the standard normal density.

In R,

- `dt(x, v)` gives the PDF of T_v for `x`.
- `pt(x, v)` gives the CDF of T_v for `x`.
- `qt(s, v)` gives the `s`100 percentile of T_v.
- `nt(n, v)` generates a random sample of n T_V random variables.

Many estimators of $\hat{\theta}$ of θ satisfy

$$\frac{\hat{\theta} - \theta}{S_{\hat{\theta}}} \sim T_v$$

This gives a $(1-\alpha)100\%$ confidence interval for θ

$$(\hat{\theta} - t_{v,\alpha/2} S_{\hat{\theta}}, \hat{\theta} + t_{v,\alpha/2} S_{\hat{\theta}})$$

7.3 Confidence Intervals for Proportions

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \dot{\sim} N(0,1)$$

Therefore,

$$1-\alpha = P\left(-z_{\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le z_{\alpha/2}\right) = P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

The $(1-\alpha)100\%$ *confidence interval* for p is

$$\left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

The confidence interval works well as long as

$$n\hat{p} \ge 8 \text{ and } n(1-\hat{p}) \ge 8 \qquad \text{Success and failure both greater than 8.}$$

The *precision* in the estimation of p is quantified by the margin of error or the length of the confidence interval.

$$MOE = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

To make the confidence interval shorter, we either have to decrease confidence level or increase sample size. Therefore, for desired length of confidence of interval L, we choose n so that

$$n \ge \frac{4z_{\alpha/2}^2\hat{p}_{pr}(1-\hat{p}_{pr})}{L^2}$$

where $\hat{p}_{pr}$ a preliminary estimate. We use 0.5 if we don't have any estimate.

7.4 Confidence Interval for the Mean

Let $X_1, ..., X_n$ be a simple random sample from a population with mean μ and variance σ^2.

7.4.1 Z Confidence Interval for the Mean

If the population variance σ^2 is known, then by the CLT,

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1)$$

A $(1-\alpha)100\%$ confidence interval for μ is therefore: (We have to know σ^2!!)

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

7.4.2 T Confidence Interval for the Mean

If we make the additional assumption that $X_1, ..., X_n$ are iid $N(\mu, \sigma^2)$, then

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim T_{n-1}$$

A $(1-\alpha)100\%$ confidence interval for μ is

$$\left(\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right)$$

Note that if the population is *normal*, this confidence interval holds for all n and we can use a Q-Q plot to check it. If the population is not normal, this confidence interval will still work well if $n \geq 30$. In R, we get $t_{n-1,\alpha/2}$ using `qt(1-α/2,n-1)`.

7.4.3 Precision

For means, the margin of error is

$$MOE = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

To find a desired sample size, we have

$$n \geq \left(2z_{\alpha/2}\frac{S_{pr}}{L}\right)^2$$

7.5 Confidence Intervals for the Variance

Suppose we want a confidence interval for the true variance σ^2. Then if the the population has a normal distribution

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

7.5.1 The χ^2 distribution

$X \sim \chi^2_v$ has the distribution with v degrees of freedom

A $(1-\alpha)100\%$ confidence interval for σ^2 is

$$\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}$$

Percentiles for the χ^2 distribution can be found in R with `qchisq(p, v)` gives $\chi_{v,1-p}$. This calculation can thus be used to calculate the confidence interval for standard deviations.

8 Chapter 8 Introduction to Hypothesis Testing

Hypothesis tests allow us to assess evidence provided by the data in favor of some claim about a population parameter. Often, we want to decide if the **observed** value of a statistic is consistent with some **hypothesized** value of a parameter.

A statistical hypothesis involved two opposing hypothesis about population parameter.

- The *null hypothesis* H_0 is a statement about "no effect" and it is the thing we are trying to disprove.
- The *alternative hypothesis* H_a is a statement about an effect that we are trying to prove.
- Favor is given to the null hypothesis H_0. Assume it is true and prove otherwise.

A hypothesis test has two possible conclusions. We either reject H_0 and decide H_a is true, or fail to reject H_0 and claim that either hypothesis could be true.

A **Type I error** occurs when we reject H_0 when H_0 is true while a **Type II error** occurs when we fail to reject H_0 when H_0 is false.

We cannot prevent Type I and Type II errors from happening, but we can control the possibilities that they occur.

- α describes the possibility of Type I error.
- β describes the probability of a Type II error

As α increases, β decreases. It is easiest to control α and we will call α the level of significance of a hypothesis test.

Example. Consider a criminal trial:

- H_0: Defendant is innocent
- H_a: Defendant is guilty
- Reject H_0: Find the defendant guilty
- Do not reject H_0: Find the defendant not guilty (NOT innocent)
- Type I Error: Find an innocent person guilty
- Type II Error: Find a guilty person not guilty

8.1 Hypothesis Testing

Our hypothesis tests are performed with *level of significance* α, with common value 0.05.

Procedure:

1. State the Hypotheses (H_0, H_a)
2. Compute a test statistic based on an estimator of the parameter
3. Reach conclusion that rejects or does not reject H_0 using either rejection rule or p-value
4. State your conclusion in the context of the problem

8.2 Hypothesis Tests for Proportions

Let $X \sim Bin(n, p)$ and let $\hat{p} = \frac{X}{n}$ be the sample proportion. There are three scenarios for stating the hypotheses, given p_0 is the hypothesized/benchmark value:

$$H_0 : p = p_0 \qquad H_0 : p = p_0 \qquad H_0 : p = p_0$$
$$H_a : p > p_0 \qquad H_a : p < p_0 \qquad H_0 : p \neq p_0$$

The test statistic is computed based on the condition that $np_0 \geq 5$ and $n(1 - p_0) \geq 5$:

$$Z_{H_0} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

If H_0 is true, then $Z_{H_0} \dot{\sim} N(0, 1)$. Therefore, we want to determine if the observed value of Z_{H_0} is unusual (by α) for a $N(0, 1)$ variable and reject H_0 if it is.

$H_a : p > p_0$	$H_a : p < p_0$	$H_0 : p \neq p_0$
Reject if $Z_{H_0} \geq Z_\alpha$	Reject if $Z_{H_0} \leq Z_\alpha$	Reject if $\lvert Z_{H_0} \rvert \geq Z_\alpha$

Meanwhile, the p-value conveys the strength of the evidence against H_0 and is the probability of observing what was observed if H_0 is true. Reject H_0 if p-value $\leq \alpha$

$H_a : p > p_0$	$H_a : p < p_0$	$H_0 : p \neq p_0$
p-value $= P(Z \geq Z_{H_0})$	p-value $= P(Z \leq Z_{H_0})$	p-value $= 2P(Z \geq \lvert Z_{H_0} \rvert)$

8.3 Hypothesis Tests for Means

Let $X_1, ..., X_n$ be a simple random sample from a population and let $\bar{X}$ and S^2 be the sample mean and sample variance. For hypothesized mean μ_0:

$$H_0 : \mu = \mu_0 \qquad H_0 : \mu = \mu_0 \qquad H_0 : \mu = \mu_0$$
$$H_a : \mu > \mu_0 \qquad H_a : \mu < \mu_0 \qquad H_0 : \mu \neq \mu_0$$

The test statistic is computed based on the condition that population is normal or $n \geq 30$.

$$T_{H_0} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

If H_0 is true, then $T_{H_0} \sim T_{n-1}$. We want t determine if the observed value of T_{H_0} is unusual for a T_{n-1} random variable. Reject H_0 if T_{H_0} is unusual.

$H_a : \mu > \mu_0$	$H_a : \mu < \mu_0$	$H_0 : \mu \neq \mu_0$
Reject if $T_{H_0} \geq T_{n-1,\alpha}$	Reject if $T_{H_0} \leq -T_{n-1,\alpha}$	Reject if $\lvert T_{H_0}\rvert \geq T_{n-1,\alpha/2}$

If p-value $\leq \alpha$, H_0 should also be rejected.

$H_a : \mu > \mu_0$	$H_a : \mu < \mu_0$	$H_0 : \mu \neq \mu_0$
p-value $= P(T_{n-1} \geq T_{H_0})$	p-value $= P(T_{n-1} \leq T_{H_0})$	p-value $= 2P(T_{n-1} \leq \lvert T_{H_0}\rvert)$

To test a set of values `x` in R given a test mean, we use

```
t.test(x, mu = 25, alternative = "two-sided")
```

where `mu` is the hypothesized value μ_0 and ”two.sided” can be replaced with ”greater” or ”this”

8.4 Hypothesis Tests for Variances

Let $X_1, ..., X_n$ be iid normal with variance σ^2. Let S^2 be the sample variance.

$H_0 : \sigma^2 = \sigma_0^2$	$H_0 : \sigma^2 = \sigma_0^2$	$H_0 : \sigma^2 = \sigma_0^2$
$H_a : \sigma^2 > \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$H_0 : \sigma^2 \neq \sigma_0^2$

The test statistic is computed based on the condition that the population is normal:

$$\chi^2_{H_0} = \frac{(n-1)S^2}{\sigma_0^2}$$

8.5 Confidence Intervals and Hypothesis Tests

Suppose we want to test a hypotheses with significance level α. The confidence interval contains plausible values of μ. If μ_0 is in the confidence interval, we do not reject H_0. If μ_0 is not in the confidence interval, we reject H_0.

9 Chapter 9 Comparing Two Populations

9.1 Comparing Two Means

Let $X_1, ..., X_{n_1}$ be a simple random sample from a population with mean μ_1 and variance σ_1^2 and $X_1, ..., X_{n_2}$ be a simple random random sample from a population with mean μ_2 and variance σ_2^2.

9.2 Case 1: Equal Variance

Assuming $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we can define a *pooled estimator* of common variance σ^2 as

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

If both populations are normal, then

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim T_{n_1+n_2-2}$$

In practice, our inference will hold if the population are not normal but $n \geq 30$ and $n_2 \geq 30$ and if S_1^2 and S_2^2 are close enough such that

$$\frac{\max\{S_1^2, S_2^2\}}{\min\{S_1^2, S_2^2\}} < \begin{cases} 5 \text{ if } n_1, n_2 \approx 7 \\ 3 \text{ if } n_1, n_2 \approx 15 \\ 2 \text{ if } n_1, n_2 \approx 30 \end{cases}$$

A $(1 - \alpha)100$th confidence interval for $\mu_1 - \mu_2$ is

$$(\bar{X}_1 - \bar{X}_2) \pm t_{n_1+n_2-2,\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

If we want a hypothesis test for $\mu_1 - \mu_2$, we first state the hypotheses

$$\begin{array}{lll} H_0 : \mu_1 - \mu_2 = \Delta_0 & H_0 : \mu_1 - \mu_2 = \Delta_0 & H_0 : \mu_1 - \mu_2 = \Delta_0 \\ H_a : \mu_1 - \mu_2 > \Delta_0 & H_a : \mu_1 - \mu_2 < \Delta_0 & H_a : \mu_1 - \mu_2 \neq \Delta_0 \end{array}$$

$$T_{H_0}^{EV} = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

If H_0 is true, $T_{H_0}^{EV} \sim T_{n_1+n_2-2}$ and we can find rejection rules and p-values with the same method.

9.3 Case 2: Unequal Variances

We allow $\sigma_1^2 \neq \sigma_2^2$ in this case. If both populations are normal, then

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \dot{\sim} T_v \text{ where } v = \left\lfloor \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} \right\rfloor$$

This approximation works well if the populations are not normal but $n_1 \geq 30$ and $n_2 \geq 30$.

The $(1-\alpha)100$th confidence interval for $\mu_1 - \mu_2$ is

$$(\bar{X}_1 - \bar{X}_2) \pm t_{v,\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

In this case, the test statistic is

$$T_{H_0}^{SS} = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

In R, for data `x1` and `x2`, the equal variance test could be tested as

```
t.test(x1, x2, mu = 2, alternative= "less", var.equal=T, conf.level=0.99)
```

9.4 Paired Data

Paired data arise from an alternative sampling design used for the comparison of two means. Each data point in the first sample is matched with a unique data point in the second sample. We thus analyze the differences between the observations for each pair of data, D.

In this case, μ_D is the population mean difference, $\bar{D}$ is the sample mean difference, S_D is the sample standard deviation of the sample, and n is the number of pairs.

Assuming the differences are normally distributed or $n \geq 30$, then

$$\frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \sim T_{n-1}$$

The $(1-\alpha)100$th confidence interval for $\mu_D = \mu_1 - \mu_2$ is

$$\bar{D} \pm t_{n-1,\alpha/2}\frac{S_D}{\sqrt{n}}$$

The test statistic of a hypothesis test is thus given by

$$T_{H_0} = \frac{\bar{D} - \Delta_0}{S_D/\sqrt{n}}$$

The paired value test in R is given by the `paired = T` option.

```
t.test(post, pre, mu = 0, paired = T, alternative = "greater")
```

10 Chapter 10: Analysis of Variance (ANOVA)

This answers the core question of how to compare means across more than two populations.

ANOVA uses a single hypothesis test to check whether that means across several populations are equal.

$$H_0 : \mu_1 = ... = \mu_k$$

$$H_a : \text{at least one mean is different}$$

If H_0 is true, the variability between the sample means should be small. The variability between the sample means is called the *variability between groups*. In order to determine if the variability between groups is large or small, we need to compare it to the *variability within each group*. This is why this method is called Analysis of *Variance*.

When $H_0 : \mu_1 = \mu_2 = \mu_3$ is false, the between groups variability will be much greater than the within groups variability

10.1 The Details

Suppose we have independent random samples from k populations.

$$X_{11}, X_{12}, ..., X_{1n_1}$$

$$X_{21}, X_{22}, ..., X_{2n_2}$$

$$\vdots$$

$$X_{k1}, X_{k2}, ..., X_{kn_k}$$

- $\bar{X}_i$ is the sample mean for the ith random sample
- S_i^2 is the sample variance for the ith random sample.
- $\bar{X}$ is the overall sample mean.
- $N = n_1 + ... + n_k$ is the overall sample size

10.1.1 Variability between Groups

$$\text{SSTr} = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2$$

$$\text{MSTr} = \frac{\text{SSTr}}{k-1}$$

SSTr is called the *treatment sum of squares*

MSTr is called the *mean squares for treatment*

10.1.2 Variability within groups

$$\text{SSE} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2 = \sum_{i=1}^{k}(n_i - 1)S_i^2$$

$$MSE = \frac{\text{SSE}}{N-k}$$

SSE is called the *error sum of squares*.

MSE is called the *mean squares for the error*

10.1.3 Test Statistic

The test statistic F_{H_0} is the ratio of the MSTr and MSE:

$$F_{H_0} = \frac{\text{MSTr}}{\text{MSE}}$$

If H_0 is true, F_{H_0} has an *F distribution*, aa positive distribution skewed to the right. This distribution is frequently used then the test statistic is a ratio, and it has two degrees of freedom v_1 and v_2 where

$$v_1 = DF_{SSTR} = k - 1 \quad \text{and} \quad v_2 = DF_{SSE} = N - k$$

We reject H_0 if $F_{H_0} \geq F_{k-1,N-k,\alpha}$, where `qf(1-`$\alpha$`, `$v_1$`, `$v_2$`)` gives $F_{v_1,v_2,\alpha}$.

p-value $= P(F_{k-1,N-k} > F_{H_0})$. Therefore, $P(F_{v_1,v_2} < x) =$ ˘`pf(`x`, `v_1`, `v_2`)`

10.1.4 Summary

$$SST = SSTr + SSE \qquad DF_{SST} = DF_{SSTr} + DF_{SSE} + 1$$

Assumptions

- The k samples are independent.
- The variances of the k populations are equal
- The k populations are normally distributed (or $n_i \geq 30$ for all i).

Suppose `values` represents all data and `group` represents the data's category at that index:

```
fit<-(aov(values ~ as.factor(group))
anova(fit)
```

11 Chapter 11: Simple Linear Regression

Simple linear regression allows us the investigate the relationship between two variables X and Y. We try to describe hw $E(Y|X = X)$ varies as a function of x

11.1 Regression Model

The simple linear regression model is given by

$$Y = \alpha_1 + \beta_1 X + \varepsilon$$

The error variable ε has a $N(0, \sigma_\varepsilon^2)$ distribution and has $\text{Cov}(X, \varepsilon) = 0$. σ_ε^2 is the same for all values of x.

Corollary. $E(Y \mid X = x) = \alpha_1 + \beta_1 x$

Corollary. The distribution of $Y \mid X = x$ is $N(\alpha_1 + \beta_1, \sigma_\varepsilon^2)$

11.2 The Method of Least Squares

We will use *least squares* to estimate the parameters of the linear regression model. We let $(X_1, Y_1), ..., (X_n, Y_n)$ be a random bivariate sample from the population.

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (\hat{\alpha_1} + \hat{\beta}_1 x_i))^2$$

The vertical distances are the estimated errors and are called *residuals*.

$$\varepsilon = Y_i - \bar{Y}_i$$

Minimizing the SSE provides the estimates:

$$\hat{\beta}_1 = \frac{S_{X,Y}}{S_X^2} \text{ and } \hat{\alpha}_1 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

We caan use the SSE to estimate σ_ε^2

$$S_\varepsilon^2 = \frac{\text{SSE}}{n-2}$$

In R, we can obtain the estimates of x and y as following:

```
model <- lm(y ~ x)
model$coefficients
model$residuals
sum(model$residuals^2)/model$df.residual
```

Note the least squares estimates are also the maximum likelihood estimates of the parameters.

The estimated intercept, α_1, is the estimated average value of Y when $X = 0$. This, however, doesn't provide meaning in most statistical context.

The estimated slope, $\hat{\beta}_1$, is the estimated change in the average of $Y \mid X = x$ for a one-unit increase in x.

11.3 Inference for beta

If $Y \mid X = x$ has a normal distribution and $S_{\hat{\beta}_1}$ is the estimated standdardd error of $\hat{\beta}_1$, then

$$\frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim T_{n-2}$$

holds if $Y \mid X = x$ is normal or $n \geq 30$. A $(1 - \alpha)100$th confidence interval for β_1 is

$$\hat{\beta}_1 \pm t_{n-2,\alpha/2} S_{\hat{\beta}_1}$$

In R, given `model` that has been calculated previously, we can get

```
confint(model, level=0.95)
```

The test statistic is given by

$$T_{H_0} = \frac{\hat{\beta}_1 - \beta_{1,0}}{S_{\hat{\beta}_1}}$$

We can also get the value in R with

```
summary(model)
```

11.4 Prediction

The estimated regression line is

$$\hat{Y} = \hat{\alpha}_1 + \hat{\beta}_1 x$$

We use this regression line to predict Y given X, denoted with $\hat{\mu}_{Y|X}(x)$. This value represents the average value of Y for sub-population with $X = x$ and an individual value of Y when $X = x$. However, it is important to note that the errors and intervals associated with the two cases are different.

11.4.1 Confidence Interval for an Average Response

A $(1 - \alpha)100$th confidence interval is

$$\hat{\mu}_{Y|X}(x) \pm t_{n-2,\alpha/2} S_{\varepsilon} \sqrt{\frac{1}{n} + \frac{(x - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$$

11.4.2 Prediction Interval for an Individual Response

A $(1-\alpha)100$th confidence interval is

$$\hat{\mu}_{Y|X}(x) \pm t_{n-2,\alpha/2} S_{\varepsilon} \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{X})^2}{\sum(X_i-\bar{X})^2}}$$

A Common Expressions in R Code

- `dbinom(`y, n, p`)`: Binomial PMF
- `pbinom(`y, n, p`)`: Binomial CDF
- `dhyper(`x, M_1, M_2, n`)`: Hypergeometric PMF
- `phyper(`x, M_1, M_2, n`)`: Hypergeometric CDF
- `dnbinom(`$y - r, r, p$`)`: Negative binomial PMF
- `pnbinom(`$y - r, r, p$`)`: Negative binomial CDF
- `dpois(`x, λ`)`: Poisson PMF
- `ppois(`x, λ`)`: Poisson CDF
- `dnorm(`x, μ, σ`)`: Normal PDF
- `pnorm(`x, μ, σ`)`: Normal PDF
- `qnorm(`s, μ, σ`)`: Normal distribution sth percentile
- `rnorm(`n, μ, σ`)`: Normal distribution random sample of size n
- `dt(`x, v`)`: PDF of T_v
- `pt(`x, v`)`: CDF of T_v
- `qt(`s, v`)`: s100 percentile of T_v
- `qchisq(`p, v`)`: Gives χ^2 distribution's p percentile
- `qf(1-`α`, `v_1`, `v_2`)`: Gives $F_{v_1, v_2, \alpha}$ distribution's percentile
- `cov(x, y)` calculates the covariance of data `x` and `y`
- `cor(x, y)` calculates the correlation of data `x` and `y`
- `t.test(x, mu = `μ_0`, alternative = "two-sided")` gives hypothesis test statistics, where `x` is the data, `mu` is the hypothesis test value μ, and `alternative` is test method that could be `two.sided, greater`, or `less` based on H_a
- `t.test(x1, x2, mu = 2, alternative= "less", var.equal=T, conf.level=0.99)` gives hypothesis test statistics for two data `x1` and `x2`. The option `var.equal=T`

states equal variance assumption, while `conf.level` gives confidence level $(1 - \alpha)$

- `t.test(post, pre, mu = 0, paired = T, alternative = "greater")` gives a paired value hypothesis test, established by the option `paired = T`