2. 기초통계

강의자 : 어지원 박사님

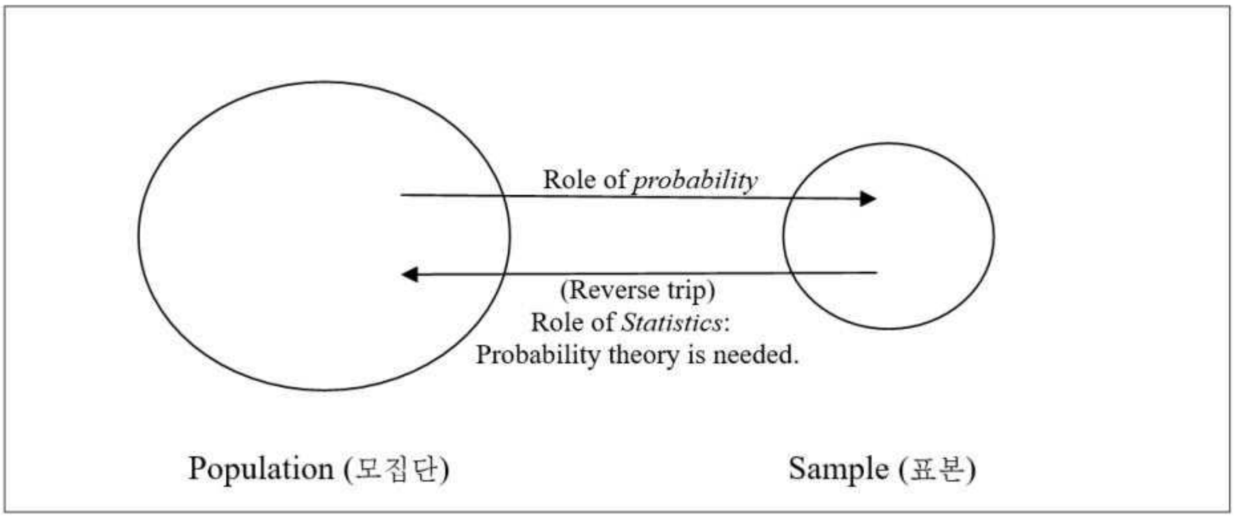
2022-2학기 DFMBA 사전교육

통계

본 강의 자료는 이 회경 교수님의 강의 자료를 참고하여 만들었습니다.

01

Probability & Sample



01

Sample - Descriptive Statistics

Relative Position

- Q1(25%p percentile)
- Median(Q2, 50%p percentile)
- Q3(75%p percentile)
- Box Plot

Center of a Distribution

- Mode
- Mean or Average
- Geometric Mean

01

Sample - Descriptive Statistics

Spread of a Distribution

- Range = largest-smallest
- IQR = Q3-Q1
- Mean Absolute Deviation(MAD)
- Variance & Standard Deviation
- Coefficient of Variation(CV) (Sharpe Ratio)

Shape of a Distribution

- Skewness
- Kurtosis

01

Probability Models

- An experiment, ε, is the process by which an observation is made.
- With each experiment, we define sample space, S, as the set of all possible outcomes of ε.
- An event is defined as a subset of the sample space S.

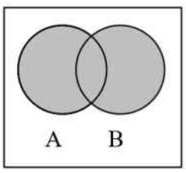
01

Events

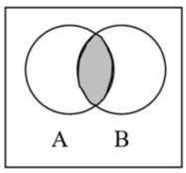
An event is defined as a subset of the sample space S.

- Compound Events

$A \cup B$

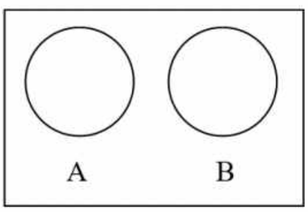


$A \cap B$



- Mutually Exclusive (Disjoint) Events

$A \cap B = \phi$



01

Events

If A and B are mutually exclusive, then

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

since $P(A \cap B) = P(\phi) = 0$.

In general, $P(A \cup B)$ is

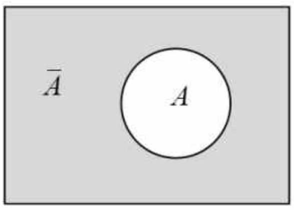
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

01

Events

An event is defined as a subset of the sample space S.

- Complement



$$P(A) + P(\bar{A}) = 1$$

01

Conditional Probability

Definition:

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

From the definition,

$$P(A \cap B) = P(A) \cdot P(B \mid A) \text{ ,or}$$

$$P(A \cap B) = P(B) \cdot P(A \mid B) .$$

01

Statistical Independence

A is *statistically* ***independent*** of B if and only if

$$P(B \mid A) = P(B), \quad \text{or}$$
$$P(A \mid B) = P(A).$$

Alternatively,

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

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

01

Mutual vs. Pairwise Independence

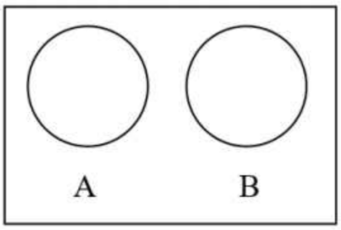
Three events, A, B and C are *mutually independent* (or *independent*) if all pairs are independent and $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$. That is,

$$\left.\begin{aligned} P(A \cap B) &= P(A) \cdot P(B) \\ P(A \cap C) &= P(A) \cdot P(C) \\ P(B \cap C) &= P(B) \cdot P(C) \end{aligned}\right\} \text{"pairwise independent", and}$$

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

01

Independence vs. Mutual Exclusiveness

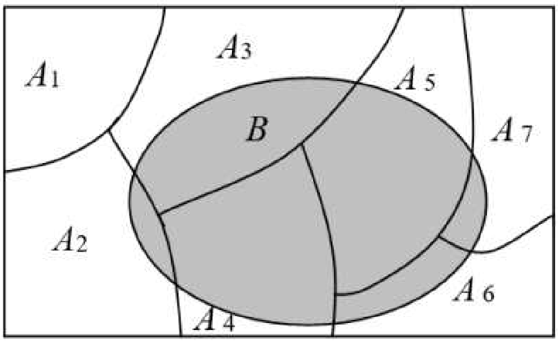


A and B are statistically dependent, because

$P(A \cap B) = P(\phi) = 0$, but $P(A) \neq 0$ and $P(B) \neq 0$. $\Rightarrow$ $P(A \cap B) \neq P(A) \cdot P(B)$.

01

Partition



Total Probability: $$P(B) = P(A_1 \cap B) + \cdots + P(A_K \cap B)$$
$$= P(A_1) \cdot P(B \mid A_1) + \cdots + P(A_K) \cdot P(B \mid A_K)$$

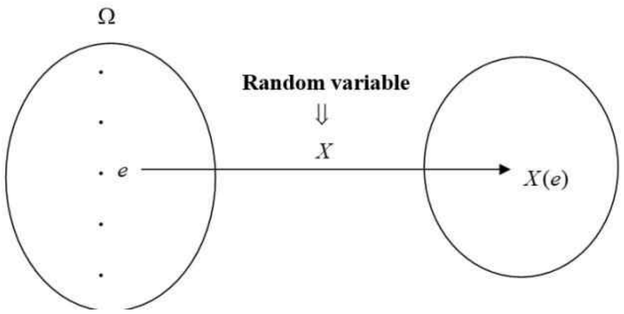
01

Bayes Theorem

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$
$$= \frac{P(A_i) \cdot P(B \mid A_i)}{P(A_1) \cdot P(B \mid A_1) + \cdots + P(A_K) \cdot P(B \mid A_K)}$$

01

Probability Distribution



- Random Variable : A function that assigns one and only one real number to each element e that belongs in the sample space S.

- If each probability for all possible $x \in X$ is known, 'probability distribution of X' is known.

01

Probability Distribution

Mean:

$$E(X) = \begin{cases} \sum_x x\, p(x) & \text{if } X \text{ is discrete} \\ \int x\, f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

(Notation: $E(X) = \mu$)

Variance:

$$\mathrm{var}(X) = E(X-\mu)^2 = \begin{cases} \sum_x (x-\mu)^2\, p(x) & \text{if } X \text{ is discrete} \\ \int (x-\mu)^2 f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

(Notation: $\mathrm{var}(X) = \sigma^2$)

01

Probability Distribution

Note: Let $Y = g(X)$. Then $E(Y)$ is given by either

$$E(Y) = \begin{cases} \sum y_i \, p(y_i) & \text{if } Y \text{ is discrete} \\ \int y \, f_Y(y) \, dy & \text{if } Y \text{ is continuous} \end{cases}$$

or

$$\boxed{E(Y) = E(g(X)) = \begin{cases} \sum_x g(x) \cdot p(x) & \text{if } X \text{ is discrete} \\ \int g(x) f(x) \, dx & \text{if } X \text{ is continuous} \end{cases}}$$

01

Probability Distribution

Theorem.

$$\mathrm{var}(X) = E(X^2) - [E(X)]^2$$

Note. $E(X^2) = \mu^2 + \sigma^2$

Probability Distribution

Example Find the average number of girls in the family of three children.

x	0	1	2	3
$P(X=x)$	0.14	0.39	0.36	0.11

$$\mu \equiv E(X) = \sum_{x=0}^{3} x \cdot p(x) = 0(0.14) + 1(0.39) + 2(0.36) + 3(0.11) = 1.44$$

$$E(X^2) = \sum_{x=0}^{3} x^2 \cdot p(x) = 0(0.14) + 1^2(0.39) + 2^2(0.36) + 3^2(0.11) = 2.82$$

$$\sigma^2 \equiv \mathrm{var}(X) = E(X^2) - \mu^2 = 2.82 - (1.44)^2 = 0.75$$

Probability Distribution

- Discrete Random Variable
  - Bernoulli
  - Binomial
  - Poisson
  - Hypergeometric
  - Geometric
  - Negative Binomial Distribution

01

Bernoulli Distribution

The individual repetitions in the binomial experiment are called Bernoulli trials. That is,

[Binomial distribution with $n = 1$] = [Bernoulli distribution].

Let $W_i = \begin{cases} 1 & \text{if "success" occurs with probability } p \text{ on the } i\text{th trial} \\ 0 & \text{elsewhere} \end{cases}$

Note: *Binomial* random variable X and *Bernoulli* random variable W_i

Consider the sum of W_is,

$$X = W_1 + W_2 + \ldots + W_n,$$

then X is the number of "successes" in n trials. $\Rightarrow X$ is a binomial random variable.

01

Bernoulli Distribution

Mean and variance of W_i: The probability distribution of W_i is[6]

$$W_i = \begin{cases} 1 & \text{with } p \\ 0 & \text{with } q = 1 - p. \end{cases}$$

Hence

$$E(W_i) = 1 \cdot p + 0 \cdot (1-p) = p$$
$$E(W_i^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$
$$\text{var}(W_i) = E(W_i^2) - [E(W_i)]^2 = p - p^2 = p(1-p) \equiv pq$$

Mean: $E(W_i) = p$

Variance: $\text{var}(W_i) = pq$ where $q = 1 - p$

01

Binomial Distribution

Examples of binomial random variables:

X = the number of heads in ten tosses of a coin.

X = the number of girls in a family of three children.

Assumptions:

1. There are n trials (시행) (e.g. n tosses of a coin)

2. In each trial, *two* events ("success" or "failure") occur. The respective probabilities are denoted by p and $q = 1 - p$.

3. The trials are statistically *independent*.

Then, X, the total number of successes in n trials, is called a *binomial* variable.

(Alternatively, we say that X has a *binomial distribution*.)

01

Binomial Distribution

Binomial probability distribution:

The probability mass function (pmf) of a binomial random variable is

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \cdots, n$$

Mean: $E(X) = np$

Variance: $\mathrm{var}(X) = npq$ where $q = 1 - p$

01

Binomial Distribution

(1) The numbers $\binom{n}{x}$ are sometimes called *binomial coefficients* since they appear in the expansion of the binomial expression $(a+b)^n$.

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \quad \text{where } n! = n\cdot(n-1)\cdots 2\cdot 1$$

$$\binom{n}{x} = \binom{n}{n-x}$$

(2) Binomial theorem

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}.$$

01

Binomial Distribution

Note: The mean and variance of a binomial random variable X can be easily obtained from the above results.

$$\begin{aligned} E(X) &= E(W_1 + W_2 + \ldots + W_n) = E(W_1) + \ldots + E(W_n) \\ &= np \\ \mathrm{var}(X) &= \mathrm{var}(W_1 + W_2 + \ldots + W_n) \\ &= \mathrm{var}(W_1) + \ldots + \mathrm{var}(W_n) \quad \text{since } W_i\text{s are IND} \\ &= np(1-p) \equiv npq \end{aligned}$$

01

Poisson Distribution

Unlike the binomial distribution, there is *no* specified number n of possible trials in Poisson distribution. X is the number of *rare* events that occur in a period of *time* or *space* during which an average of λ such events can be expected to occur. It is assumed that the events occur *independently* of one another. Some examples that might follow a Poisson distribution:

1. The number of days in a *given year* in which a 50-point change occurs in the Dow Jones Index. (This is an example of an interval of time.)
2. The number of automobile accidents at a particular intersection during a time period of *one week*.
3. The number of people who apply for a job to your company during a *given day*.
4. The number of defects in a finished product during a *given day*.
5. The number of typos on a *printed page*. (This is an example of an interval of space.)

01

Poisson Distribution

A random variable X is said to have a *Poisson probability distribution* if its pmf is

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \qquad x = 0, 1, 2, \ldots, \quad \lambda > 0.$$

where X = the number of occurrences in an interval,

λ = *mean* number of occurrences in an interval.

Allow at most one success in each subinterval (trial) $\Rightarrow$ Each trial is Bernoulli, and all trials constitute a binomial distribution.

If X has a Poisson distribution with parameter λ, then

$$E(X) = \lambda,$$

$$\text{var}(X) = \lambda.$$

01

Poisson Distribution

Example Because your firm's quality is so high, you expect only 1.2 of your products to be returned, on average, each day for warranty repairs. What are the chances that no products will be returned tomorrow? That one will be returned? How about two? (Let X be the number of products returned.)

Example Customers enter a waiting line "at random" at a rate of 3 per minute. Assuming that the number entering the line in any given time interval has a Poisson distribution, what is the probability that at least one customer enters the line in a given half-minute interval?

01

Hypergeometric Distribution

Binomial $\leftrightarrow$ Sampling *with* replacement. Trials are IND.
Hypergeometric $\leftrightarrow$ Sampling *without* replacement. Trials are not IND.

r **red** balls (i.e., r "**successes**") got mixed with $N-r$ **white** ones (i.e., $N-r$ "**failures**") in a container. Let X be the number of red balls. n balls are chosen ***without*** replacement from the container. Then X follows the hypergeometric distribution whose pmf is

$$P(X=x)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \qquad x=0,1,\cdots,r.$$

01

Negative Binomial Distribution

Assume Bernoulli trials. Let X = the number of trials required until the **r**th success, and $P(\text{success}) = p$. Then X follows the negative binomial distribution whose pmf is

$$P(X = x) = \binom{x-1}{r-1}(1-p)^{x-r} p^r, \qquad x = r, r+1, r+2, \ldots$$

The mean and variance are

$$E(X) = \frac{r}{p}$$

$$\operatorname{var}(X) = \frac{rq}{p^2} = \frac{r(1-p)}{p^2}$$

Note: A geometric distribution is a special case of a negative binomial distribution with $r = 1$.

01

Probability Distribution

- Continuous Random Variable
  - Uniform
  - Normal
  - Exponential
  - Gamma
  - Weibull
  - Lognormal
  - …

Probability Distribution

- Continuous Random Variable

X is said to be a continuous random variable if there exists a function f, called the *probability density function* (확률밀도함수) (pdf) of X, satisfying the following conditions.

(i) $f(x) \geq 0$ for all x

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

(iii) For any *a, b* with $-\infty < a < b < \infty$, we have $P(a \leq X \leq b) = \int_a^b f(x)dx$.

01

Probability Distribution

- Continuous Random Variable

(a) $P(a \leq X \leq b)$ represents the area under $f(x)$.

(b) For any specific value of *x*, say x_0, we have $P(X = x_0) = 0$, since

$$P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0.$$

(c) Due to (b), the following probabilities are all the same if <u>*X* is continuous</u>.

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$$

(d) $f(x)$ does <u>not</u> represent the probability of anything. Only when the function is integrated between the two limits, does it yield a probability.

Example The pdf of X is given by

$$f(x)=\begin{cases}2x & 0<x<1\\ 0 & \text{elsewhere}\end{cases}$$

Find $P(X \le \frac{1}{2})$ and $P(X \le \frac{1}{2} \mid \frac{1}{3} \le X \le \frac{2}{3})$.

Example Let X be the life length of a certain item. The pdf of X is given by

$$f(x)=\begin{cases}\dfrac{a}{x^3} & 2 \le x \le 4\\ 0 & \text{elsewhere}\end{cases}$$

Calculate a for $f(x)$ to be a valid pdf.

Cumulative Distribution Function

Let X be a random variable, discrete or continuous. We define F to be a *cumulative distribution function* (cdf) of the random variable X where

$$F(x)=P(X \le x).$$

(a) If X is a discrete random variable,

$$F(x)=\sum_j p(x_j),$$

where the sum is taken over all indices j satisfying $x_j \le x$.

(b) If X is a continuous random variable with pdf f,

$$F(x)=\int_{-\infty}^{x} f(s)ds.$$

Cumulative Distribution Function

<Obtaining pdf (or pmf) of X from cdf of X>

(a) Let X be a discrete random variable with possible values $x_1, x_2, \cdots,$ and suppose that it is possible to label these values so that $x_1 < x_2 < \cdots$. Let F be the cdf of X. Then

$$p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1}).$$

(b) Let F be the cdf of a continuous random variable with pdf f. Then

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

for all x at which F is differentiable.

Uniform Distribution

The pdf of a continuous uniform distribution is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{elsewhere} \end{cases}.$$

The mean and variance are

$$E(X) = \frac{b+a}{2}$$

$$\text{var}(X) = \frac{(b-a)^2}{12}$$

01

Exponential Distribution

The pdf of the exponential distribution is

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

Properties of the exponential distribution

(1) $E(X)=\dfrac{1}{\lambda}$ and $\mathrm{var}(X)=\dfrac{1}{\lambda^2}$.

(2) $F(x)=P(X\leq x)=1-e^{-\lambda x}$ for $x\geq 0$.

Hence, $P(X>x)=e^{-\lambda x}$.

(3) "The memoryless property"

$$P(X>s+t\mid X>s)=\frac{P(X>s+t)}{P(X>s)}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}$$
$$=P(X>t)$$

01

Exponential Distribution

Note: The geometric distribution also has this memoryless property.

Example Consider the exponential random variable with the pdf

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

Calculate $P(X>\frac{1}{\lambda})$. Is it 0.5?

Assume X follows a Poisson process. X is the number of customers arriving at an ATM in an interval of length 1 (e.g., 1 hour.) λ is the mean number of customers arriving in an interval of length 1. ($P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$.) Let W be the *waiting time* until the ***first*** customer arrives. Then W follows an exponential distribution. Why? Let's first find the cdf of W, $F(w) = P(W \le w)$.

0 * ** * * * 1

$$F(w) = P(W \le w) = 1 - P(W > w) = 1 - P(\text{no customers in } [0, w])$$
$$= 1 - P(X = 0 \text{ with mean } \lambda w) = 1 - \frac{(\lambda w)^0}{0!} e^{-\lambda w} = 1 - e^{-\lambda w}$$

which is the cdf of the exponential random variable W. Alternatively, the pdf of W is

$$f_W(w) = \frac{dF_W(w)}{dw} = \lambda e^{-\lambda w}$$

which is the pdf of an exponential random variable.

Normal Distribution

The pdf of the normal distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0 .$$

The mean and variance are derived as

$$E(X) = \mu$$
$$\text{var}(X) = \sigma^2$$

Notation: $X \sim N(\mu, \sigma^2)$

Theorem: Let $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$ where a and b are constants. Then,

$$Y \sim N(a\mu + b, a^2\sigma^2) .$$

01

Normal Distribution

Corollary: *Standard normal* distribution.

If $X \sim N(\mu, \sigma^2)$ and $Z = \dfrac{X-\mu}{\sigma}$, then

$$Z \sim N(0,1).$$

The pdf of Z is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

The *cdf* of the standard normal random variable, $\Phi(x)$, has been tabulated. Using this standard normal table, one can calculate $P(c \le X \le d)$. That is,

$$P(c \le X \le d) = P\left(\frac{c-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{d-\mu}{\sigma}\right) = P\left(\frac{c-\mu}{\sigma} \le Z \le \frac{d-\mu}{\sigma}\right)$$
$$= \Phi\left(\frac{d-\mu}{\sigma}\right) - \Phi\left(\frac{c-\mu}{\sigma}\right)$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

The values in the table are the areas between zero and the z-score. That is, P(0<Z<z-score)

01

Normal Distribution

Example A client has an investment portfolio whose mean value is equal to \$500 with a standard deviation of \$15. She has asked you to determine the probability that the value of her portfolio is between \$485 and \$530.

Solution: Let X be the investment portfolio. Then

$$P(485 \le X \le 530) = P\left(\frac{485-500}{15} \le \frac{X-\mu}{\sigma} \le \frac{530-500}{15}\right) = P(-1 \le Z \le 2) = 0.8185.$$

01

Normal Distribution

If random variable X ~ Normal and Y ~ Normal and X, Y are independent, linear combination of random variables X and Y, aX+bY also follows Normal Distribution.

* Proof can be done through moment generating function.

01

Moment Generating Function

$$M_X(t) = E(\mathrm{e}^{Xt})$$

01

Moment Generating Function

Example Bernoulli random variable

Let $X = \begin{cases} 1 & \text{with prob } p \\ 0 & \text{with prob } q = 1 - p \end{cases}$

$M_x(t) = E(e^{tX}) = q + pe^t$.

01

Moment Generating Function

Example Binomial random variable

$$X \sim \text{Binomial}(n, p); \quad p(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \cdots, n$$

$$M_x(t) = E(e^{tX}) = (q + pe^t)^n$$

01

Moment Generating Function

Example Standard normal random variable

$$X \sim N(0,1)$$

$$M_x(t) = E(e^{tX}) = e^{\frac{1}{2}t^2}$$

01

Moment Generating Function

Example Normal random variable

$Y \sim N\left(\mu, \sigma^2\right)$

$M_x(t) = E(e^{tX}) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

01

Lognormal Distribution

Suppose X is $N(\mu, \sigma^2)$. Let

$$Y = e^X.$$

Then Y has the *lognormal* distribution. That is,

$$Y \text{ is } \textit{lognormal} \Leftrightarrow \ln Y \text{ is normal.}$$

The pdf of Y is $f(y) = \dfrac{1}{y\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2}$.

$$E(Y) = e^{\mu + \frac{1}{2}\sigma^2}$$

$$\text{var}(Y) = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

Note: Chi-squared (χ^2), t, and F distributions will be discussed later in statistical inference.

01

Chi Square Distribution

- If $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables, then X, defined by $X = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ is said to have a chi-square Distribution with n degrees of freedom ($X \sim \chi^2[n]$)
- If X_1 and X_2 are independent chi-square random variables with n_1 and n_2 degrees of freedom, respectively, then $X_1 + X_2$ is chi-square with $n_1 + n_2$ degrees of freedom.
- If X is a chi-square random variable with n degrees of freedom, then for any α $(0 \le \alpha \le 1)$, the quantity $\chi^2_{\alpha,n}$ is defined to be such that $P\left(X > \chi^2_{\alpha,n}\right) = \alpha$.
- Using Tables
  - ✓ Find $\chi^2_{0.05,8}$, $\chi^2_{0.95,8}$.

01

The t Distribution

- If Z and χ^2_n are independent random variables, with Z having a standard normal distribution and χ^2_n having a chi-square distribution with n degrees of freedom, then the random variable X defined by $X = \dfrac{Z}{\sqrt{\chi^2_n / n}}$ is said to have a t-distribution with n degrees of freedom($X \sim t[n]$).
- For any α $(0 \le \alpha \le 1)$, the quantity $t_{\alpha,n}$ is defined to be such that $P\left(X > t_{\alpha,n}\right) = \alpha$.
- Using Tables

$t_{1-\alpha,n} = -t_{\alpha,n}$

  - ✓ Find $t_{0.05,25}$, $t_{0.05,74}$, $t_{0.95,10}$.

The F Distribution

- If χ_n^2 and χ_m^2 are independent chi-square random variables with n and m degrees of freedom,

respectively, then the random variable X defined by

$X = \dfrac{\chi_n^2 / n}{\chi_m^2 / m}$ is said to have an F-distribution with n and m degrees of freedom. ($X \sim F[n,m]$)

- If $X \sim t[n]$, $X^2 \sim F[1,n]$
- For any $\alpha\ (0 \le \alpha \le 1)$, the quantity $F_{\alpha,n,m}$ is defined to be such that $P\left(X > F_{\alpha,n,m}\right) = \alpha$
- Using Tables

$$\frac{1}{F_{\alpha,n,m}} = F_{1-\alpha,m,n}$$

✓ Find $F_{0.05,5,7}$, $F_{0.95,4,8}$.

Two Random Variables

- Conditional Expectation of Y given X !

$$E(Y \mid X = x) = \sum_y y\ p(y \mid x) \quad \text{for discrete } (X,Y),$$

$$E(Y \mid X = x) = \int y\ f(y \mid x) dy \quad \text{for continuous } (X,Y).$$

Note: In the regression context, $E(Y \mid X)$ is called a regression line (or curve).

02

Two Random Variables

- Conditional Expectation of Y given X !

(Simple) law of iterated expectations:

$$E[E(Y \mid X)] = E(Y)$$

(General) law of iterated expectations: For two random variables X and W ,

$$E\left[E(Y \mid X, W) \mid X\right] = E(Y \mid X) = E\left[E(Y \mid X) \mid X, W\right]$$

"The smaller information set wins."

02

Two Random Variables

- Conditional Expectation of Y given X !

		Y				
		1	2	3	$p(x)$	$E(Y \mid X = x)$
X	1	0	0.26	0.13	0.39	7/3
	2	0.14	0.24	0.12	0.5	1.96
	3	0.11	0	0	0.11	1
	$p(y)$	0.25	0.5	0.25	1	

02

Two Random Variables

- Conditional Expectation of Y given X !

Notes:

1) $E(Y \mid X)$ is a function of X , hence $E(Y \mid X)$ is a random variable. Strictly speaking, $E(Y \mid x)$ is the value of the random variable $E(Y \mid X)$.

2) Since $E(Y \mid X)$ is a random variable, we can think about its expectation, $E[E(Y \mid X)]$, which becomes $E(Y)$.

02

Two Random Variables

- Independence

The random variables X and Y are said to be independent (IND) if and only if

$$p(x, y) = p(x)\,p(y) : \text{discrete}$$

$$f(x, y) = f_X(x)\,f_Y(y) : \text{continuous}$$

Alternatively, X and Y are said to be IND if and only if

$$p(x \mid y) = p(x) \quad \text{or} \quad p(y \mid x) = p(y) : \text{discrete}$$

$$f(x \mid y) = f_X(x) \quad \text{or} \quad f(y \mid x) = f_Y(y) : \text{continuous}$$

Two Random Variables

- Covariance

$$\sigma_{X,Y} \equiv cov(X,Y)$$

The covariance of X and Y is defined as

$$cov(X,Y) = E[X - E(X)][Y - E(Y)]$$

Theorem.

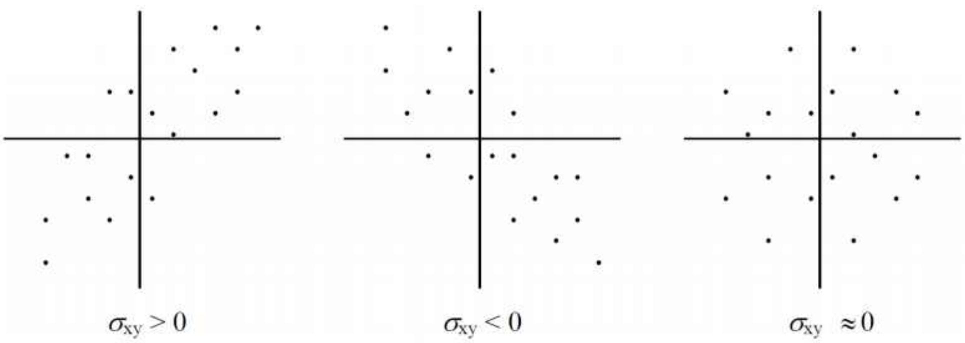
$$cov(X,Y) = E(XY) - E(X)E(Y)$$

Two Random Variables

- Covariance

Note: $cov(X,Y) = 0 \leftrightarrow E(XY) = E(X)E(Y)$.

Note: $cov(X,Y)$ shows how closely X and Y are related.



02

Two Random Variables

- Correlation Coefficient

The covariance statistic depends on the unit of measurement for X and Y . A unit-free measure, the *correlation coefficient* , $\rho_{X,Y} \equiv corr(X,Y)$ is defined as

$$\boxed{\rho_{x,y} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}} \left(\equiv \frac{\sigma_{x,y}}{\sigma_x \sigma_y} \right)}$$

02

Two Random Variables

- Correlation Coefficient

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

2) $Y = aX + b \;\; \leftrightarrow \;\; \rho_{X,Y}^2 = 1$

3) $\rho_{X,Y}$: the degree of *linearity* between X and Y

Let $V = aX + b$ and $W = cY + d$, where a, b, c, and d are constants. Then

$$\rho_{V,W} = \frac{ac}{|ac|} \rho_{x,y} .$$

That is, if $ac > 0$, then $\rho_{V,W} = \rho_{X,Y}$.
If $ac < 0$, then $\rho_{V,W} = -\rho_{X,Y}$.
Also, $\rho_{V,W}^2 = \rho_{X,Y}^2$.

Two Random Variables

- $\text{corr}(X,Y) = \text{corr}\left(\dfrac{X-\mu_x}{\sigma_x}, \dfrac{Y-\mu_y}{\sigma_y}\right)$

- Let X = income (measured in million \$), Y = consumption (measured in million \$), X^* = income (measured in \$), and Y^* = consumption (measured in \$). Then, $\text{corr}(X,Y) = \text{corr}(X^*,Y^*)$.

Independence

Proof:

$$\begin{aligned}
E(XY) &= \iint xy\, f(x,y)\, dxdy \\
&= \iint xy\, f_X(x) f_Y(y)\, dxdy \quad \text{since } X \text{ and } Y \text{ are IND} \\
&= \int \left[\int x\, f_X(x) dx \right] y\, f_Y(y) dy \\
&= \int x\, f_X(x) dx \cdot \int y\, f_Y(y) dy \\
&= E(X)E(Y)
\end{aligned}$$

The converse, however, is *not* true in general. The zero covariance does *not* imply independence in general.

Independence

- $E(XY) = (EX)(EY)$
- $\rho_{xy} = 0$ $(cov(x, y) = 0)$ ($*$ *Converse is not true*)
    - ➢ If X and Y are bivariate normal(jointly normal), then converse is true as well.

- $E(X|Y) = E(Y)$ *and* $E(Y|X) = E(X)$

Linear Combination of Two R.V.

Let a,b,c and d be constants. Then

$$E(aX + bY) = aE(X) + bE(Y)$$
$$\operatorname{var}(aX + bY) = a^2 \operatorname{var}(X) + b^2 \operatorname{var}(Y) + 2ab \operatorname{cov}(X, Y)$$

$$\operatorname{cov}(aX + bY, cX + dY) = ac \operatorname{var}(X) + (ad + bc) \operatorname{cov}(X, Y) + bd \operatorname{var}(Y)$$
$$\operatorname{cov}(aX + b, cY + d) = ac \operatorname{cov}(X, Y)$$

Linear Combination of Two R.V.

Example Suppose that a couple is drawn at random from a large population of working couples. Let X = man's income and Y = woman's income. Suppose the couple's pension contribution (W) is 10% of the man's income, and 20% of the woman's income: $W = (0.1)X + (0.2)Y$. Calculate E(W) and var(W).

(Let $E(X) = 20$, $E(Y) = 16$, $\text{var}(X) = 60$, $\text{var}(Y) = 70$, and $\text{cov}(X, Y) = 49$.)

2021-2학기 MGT503 Business Statistics

Day 2.

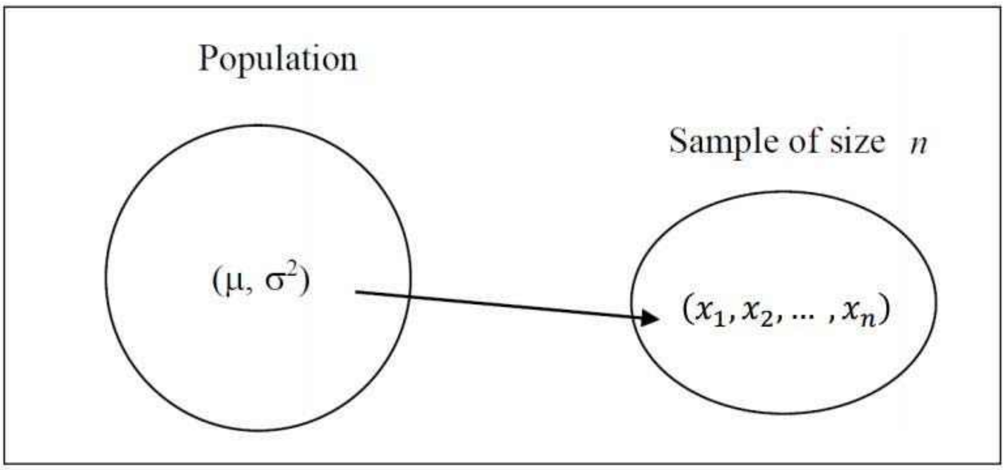
Sampling and Estimation

CONTENTS

03

Sampling

We want to infer, from the sample of size n , what the characteristics of the population (e.g. μ, σ^2) might be.
$\Rightarrow$ "estimation" of μ and σ^2 .



03

Sampling

Definition: Random sample

Let X be a random variable with a certain probability distribution. Let $X_1, X_2, \ldots, X_n$ be n <u>independent</u> random variables each having the same distribution as X . We then call $(X_1, X_2, \ldots, X_n)$ a *random sample* from the random variable X .

Sample mean : $\bar{X} = \dfrac{X_1 + \cdots + X_n}{n} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

Sample variance : $S^2 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$, $n-1$: degrees of freedom (자유도)

03

Sampling

$$E(\bar{X}) = \mu$$

$$\mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$E(S^2) = \sigma^2$$

$$\mathrm{var}(S^2) = \frac{1}{n}\left[E(X_i - \mu)^4 - \frac{n-3}{n-1}\sigma^4 \right]$$

03

Central Limit Theorem

As the sample size increases, the shape of the sample mean $\bar{X}$ approaches the normal distribution regardless of the population distribution.

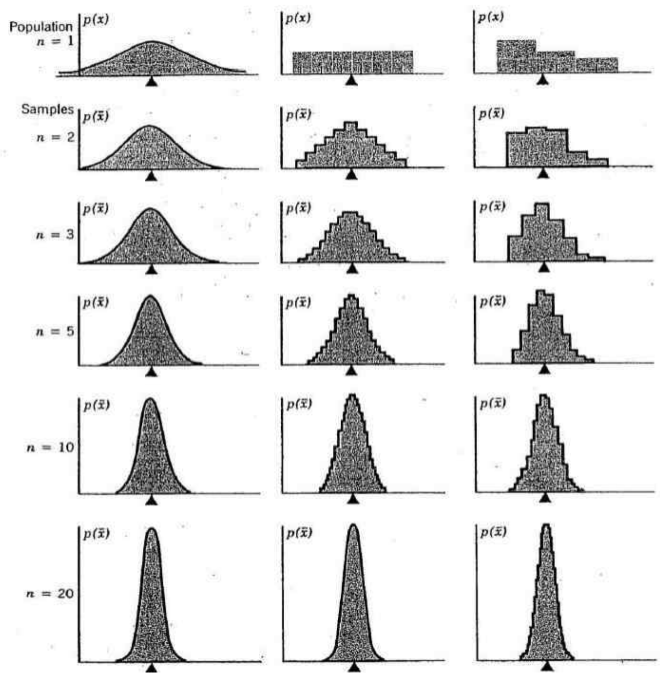
If X_i is independent, and identically distributed (i.i.d.) with $E(X_i) = \mu$ and $var(X_i) = \sigma^2$, then

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\mathrm{var}(\bar{X})}} = \frac{\bar{X} - E(\bar{X})}{\mathrm{se}(\bar{X})} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1) \quad \text{as} \quad n \to \infty .$$

Alternatively, $\sqrt{n}\left(\bar{X} - \mu\right) \xrightarrow{d} N(0, \sigma^2)$ as $n \to \infty$.

03

Central Limit Theorem



03

Central Limit Theorem

If the objective is to use $\bar{X}$ to make an inference about μ, the relevant statement may be

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

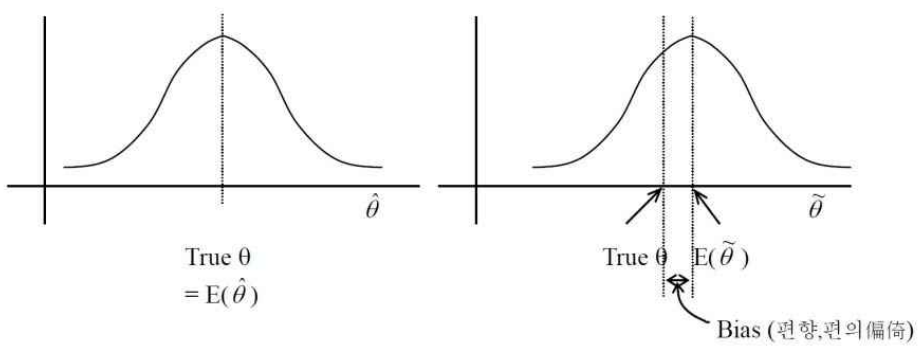
The notation $\overset{a}{\sim}$ stands for "asymptotically distributed“.
That is, $\bar{X}$ is asymptotically normally distributed with mean μ and variance σ^2/n.
(Note that the variance σ^2/n approaches 0 as n→∞.)

03

Point Estimator

- Unbiasedness

$\hat{\theta}$ is an ***unbiased*** estimator of θ if $E(\hat{\theta}) = \theta$.

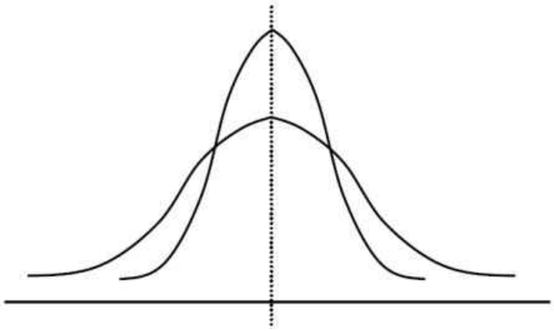


Bias ≡ E($\tilde{\theta}$) – θ

Point Estimator

- Efficiency

Relative efficiency of two unbiased estimators ($\hat{\theta}$ *and* $\tilde{\theta}$)
$= \mathrm{var}(\hat{\theta})/var(\tilde{\theta})$



Point Estimator

- Consistency

Consistent estimator
If the sequence of estimators { $\hat{\theta}_n$ } satisfies

$$\hat{\theta}_n \xrightarrow{p} \theta,$$

$\hat{\theta}_n$, is said to be **consistent** for θ.

Simple Linear Regression

Let $E(Y \mid X) = g(X) = \alpha + \beta X$, then $Y = \alpha + \beta X + \varepsilon$ (which is a model in population form). ε is called the error term (오차항) or the disturbance term (교란항). The model in terms of random sample is

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad i = 1, \ldots, n$$

Assumptions

1. Linearity

$Y = \alpha + \beta X + \varepsilon$. (To be precise, the model is linear in *parameters*.)

Assumptions

2. Exogeneity $E(\varepsilon|X) = 0$

- $E(\varepsilon \mid X) = 0$ implies $E(\varepsilon) = 0$ due to *the law of iterated expectations*:

$$E(\varepsilon) = E[E(\varepsilon \mid X)] = 0$$

- $E(\varepsilon \mid X) = 0$ also implies $\text{cov}(X, \varepsilon) = 0$. That is, the *correlation* between X and ε is 0.

$$\begin{aligned}\text{cov}(X, \varepsilon) &= E(X\varepsilon) - E(X)E(\varepsilon) \\ &= E(X\varepsilon) = 0\end{aligned}$$

since

$$E(X\varepsilon) = E[E(X\varepsilon \mid X)] = E[X \cdot E(\varepsilon \mid X)] = E[X \cdot 0] = 0$$

04

Assumptions

2. Exogeneity $E(\varepsilon|X) = 0$

- X is said to be *exogenous* if $E(\varepsilon \mid X) = 0$. If X is correlated with ε, then X is said to be *endogenous* (내생內生).

Note: *Strict* exogeneity (in matrix form):

$$E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = 0, \text{ where } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \text{ and } \boldsymbol{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}.$$

04

Assumptions

3. Spherical Disturbances

$\mathrm{var}(\varepsilon_i \mid \mathbf{X}) = E(\varepsilon_i^2 \mid \mathbf{X}) = \sigma^2$, for all i : homoscedasticity (동분산同分散)

$\mathrm{cov}(\varepsilon_i, \varepsilon_j \mid \boldsymbol{X}) = E(\varepsilon_i \cdot \varepsilon_j \mid \boldsymbol{X}) = 0$, for $i \neq j$: non-autocorrelation (비非자기상관)

04

Assumptions

3. Spherical Disturbances

Note: If the data $\{(X_i, Y_i): i = 1, 2, \ldots, n\}$ is a *random sample* (following the population model $Y = \alpha + \beta X + \varepsilon$), we automatically have zero covariance between observation *i* and observation *j*, hence $\mathrm{cov}(\varepsilon_i, \varepsilon_j \mid \boldsymbol{X}) = 0$.

- Due to the law of iterated expectations, we may write

$$\mathrm{var}(\varepsilon_i \mid \boldsymbol{X}) = E(\varepsilon_i^2 \mid \boldsymbol{X}) = \sigma^2 \quad \Rightarrow \quad \mathrm{var}(\varepsilon_i) = \sigma^2.$$
$$\mathrm{cov}(\varepsilon_i, \varepsilon_j \mid \boldsymbol{X}) = E(\varepsilon_i \cdot \varepsilon_j \mid \boldsymbol{X}) = 0 \quad \Rightarrow \quad \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0.$$

Assumptions

3. Spherical Disturbances

Note: Variance decomposition formula (or the law of total variance)

For two random variables X and Y,

$$\text{var}(Y) = E[\text{var}(Y \mid X)] + \text{var}[E(Y \mid X)]$$

- Alternatively, using the variance decomposition formula, (in terms of population)

$$\begin{aligned}\text{var}(\varepsilon) &= E[\text{var}(\varepsilon \mid X)] + \text{var}[E(\varepsilon \mid X)] \\ &= E[\sigma^2] + \text{var}[0] = \sigma^2\end{aligned}$$

Assumptions

4. Nonstochastic Regressors

5. Normality

This assumption is needed to derive the distributions of the test statistics such as standard normal, t, χ^2, and F distributions.

Least Squares Estimation

◆ **Least Squares Principle:**

The model is $Y_i = \alpha + \beta X_i + \varepsilon_i$. Let a and b be estimators for α and β, respectively. Then, the fitted (or estimated) Y_i is

$$\hat{Y}_i = a + bX_i,$$

and the residual e_i (잔차殘差) is

$$e_i = Y_i - \hat{Y}_i = Y_i - a - bX_i.$$

The **least squares principle** is

$$\arg\min_{a,b} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - a - bX_i)^2.$$

04

Least Squares Estimation

First-order condition: $\begin{cases} \partial \sum_{i=1}^{n} e_i^2 \Big/ \partial a = 0 \\ \partial \sum_{i=1}^{n} e_i^2 \Big/ \partial b = 0 \end{cases}$

which becomes

$$\begin{cases} \sum_{i=1}^{n} Y_i - a - bX_i = 0 \\ \sum_{i=1}^{n} (Y_i - a - bX_i) X_i = 0 \end{cases} \quad \text{i.e.,} \quad \begin{cases} \sum_{i=1}^{n} e_i = 0 \\ \sum_{i=1}^{n} X_i e_i = 0 \end{cases}$$

or

$$\begin{cases} na + b\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \\ a\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i \end{cases} \quad \text{: “Normal equations”}^*$$

04

Least Squares Estimation

Solving for a and b, we obtain the OLS estimators:

The **Ordinary Least Squares (OLS)** estimators a (for α) and b (for β) are,

$$a = \overline{Y} - b\overline{X}$$

$$b = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

where $x_i = X_i - \overline{X}$ and $y_i = Y_i - \overline{Y}$.[†]

05

Estimator for Coefficient

- Mean and variance of b:

Let

$$b = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \sum_{i=1}^{n} w_i y_i = \beta + \sum_{i=1}^{n} w_i \varepsilon_i{}^{\S} \quad \text{where} \quad w_i = \frac{x_i}{\sum_{i=1}^{n} x_i^2}{}^{**}.$$

Then under the assumption of non-stochastic X,

$$E(b) = \beta + E\left(\sum_{i=1}^{n} w_i \varepsilon_i\right) = \beta + \sum_{i=1}^{n} w_i E(\varepsilon_i) = \beta : \text{ unbiased}$$

$$\text{var}(b) = \text{var}\left(\sum_{i=1}^{n} w_i \varepsilon_i\right) = \sum_{i=1}^{n} w_i^2 \text{ var}(\varepsilon_i) = \sigma^2 \sum_{i=1}^{n} w_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.$$

Estimator for the variance

$$S^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - a - bX_i)^2$$

Degrees of freedom (d.f.) = $n-2$

We can show that S^2 is an *unbiased* estimator of σ^2,††

$$E(S^2) = \sigma^2.$$

- The *estimated* variance of b is, after replacing the unknown variance σ^2 with

estimated S^2,

$$\text{Est. var}(b) = \frac{S^2}{\sum_{i=1}^{n} x_i^2}.$$

The *standard error* (se) of b is defined as

$$\text{se}(b) = \sqrt{\text{var}(b)}.$$

Measure of 'Goodness of fit'

$$\left.\begin{aligned}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 &= \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n} e_i^2\\ \sum_{i=1}^{n} y_i^2 &= \sum_{i=1}^{n} \hat{y}_i^2 \qquad + \sum_{i=1}^{n} e_i^2\\ &= b^2\sum_{i=1}^{n} x_i^2 \quad + \sum_{i=1}^{n} e_i^2\end{aligned}\right\} \quad (1)$$

TSS = ESS + USS

where TSS = *total* sum of squares,

ESS = *explained* sum of squares (i.e., explained by the regression), and

USS = *unexplained* sum of squares (or sometimes called RSS = *residual* sum of squares). Eq.(1) is sometimes called the analysis of variance (ANOVA) formula for least squares regression. Dividing both sides by TSS,

$$1 = \frac{\text{ESS}}{\text{TSS}} + \frac{\text{USS}}{\text{TSS}}.$$

Measure of 'Goodness of fit'

$$R^2 = \frac{\text{ESS}}{\text{TSS}}$$

R^2 = the proportion of variation in Y explained by X.

For instance, in the example of $wage = -7.56 + 2.00(educ)$, $R^2 = 0.1902$ means that 19.02 % of the variation in Y (= *wage*) is explained by X (= *educ*).

Measure of 'Goodness of fit'

Notes:

(1) Decomposition (TSS = ESS + USS) is possible when

(a) the regression model is linear,

(b) the model includes a constant term, and

(c) OLS is used.

(2) Without the constant term, $R^2 = 1 - \frac{\text{USS}}{\text{TSS}}$ can be negative.

(3) $R^2 = r_{Y,X}^2 = r_{Y,\hat{Y}}^2$,

where $r_{Y,X}$ = sample correlation coefficient between Y and X,

$r_{Y,\hat{Y}}$ = sample correlation coefficient between Y and $\hat{Y}$.

05

Measure of 'Goodness of fit'

(4) Relationship between b (the estimated slope) and $r_{Y,X}$:

$$b = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}} \cdot \frac{\sqrt{\sum_{i=1}^{n} y_i^2}}{\sqrt{\sum_{i=1}^{n} x_i^2}} = r_{Y,X} \cdot \frac{S_Y}{S_X}$$

05

Hypothesis Testing

For hypothesis testing, we need the distributional assumption for ε_i. We assume ε_i is *normal*. Under the normality assumption, ε_i may be stated as

$$\varepsilon_i \sim i.i.d.N(0, \sigma^2).$$

Note that $b = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2 = \beta + \sum_{i=1}^{n} w_i \varepsilon_i$. Since ε_i is independently normal, it follows that b is also *normal* with mean β and variance $\sigma^2 \Big/ \sum_{i=1}^{n} x_i^2$.

$$\boxed{b \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}\right).}$$

05

Hypothesis Testing

After the standardization,

$$\frac{b-E(b)}{\sqrt{\text{var}(b)}} = \frac{b-\beta}{\sqrt{\sigma^2 / \sum_{i=1}^{n} x_i^2}} \sim N(0,1).$$

Consider the following hypothesis test,

$$H_0 : \beta = \beta^0 \quad \text{vs.} \quad H_a : \beta \neq \beta^0.$$

If σ^2 is *known*, the *standard normal* distribution can be used. If σ^2 is *unknown*, t distribution may be used instead. We will focus on the testing procedure using t distribution, as σ^2 is unknown in general.

05

Hypothesis Testing

Confidence interval for β:

$$b \pm t_{\alpha/2} \cdot \text{se}(b)$$

where α = *significance level* (유의有意수준), e.g., $\alpha = 0.05$.

Standard error (se) of b: $\text{se}(b) = \sqrt{\text{var}(b)} = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}}$, which is estimated as $\sqrt{\frac{S^2}{\sum_{i=1}^{n} x_i^2}}$.

Hypothesis Testing

$H_0 : \beta = \beta^0$

$$t = \frac{b - E(b)}{\sqrt{\mathrm{var}(b)}} = \frac{b - \beta^0}{\mathrm{se}(b)} = \frac{b - \beta^0}{S\Big/\sqrt{\sum_{i=1}^{n} x_i^2}} \sim t_{(n-2)}$$

In particular, for $H_0 : \beta = 0$, the t statistic is simply

$$t = \frac{b}{\mathrm{se}(b)} = \frac{b}{S\Big/\sqrt{\sum_{i=1}^{n} x_i^2}} \sim t_{(n-2)}.$$

$$t = \frac{b}{\mathrm{se}(b)} = \frac{b}{S\Big/\sqrt{\sum_{i=1}^{n} x_i^2}} = \frac{b\sqrt{\sum_{i=1}^{n} x_i^2}}{S}$$

can be compared to the following F statistic.

ANOVA

F-statistic is defined as

$$F = \frac{\mathrm{ESS}/1}{\mathrm{USS}/(n-2)} \left(= \frac{b^2 \sum_{i=1}^{n} x_i^2}{S^2} \right) \sim F_{(1,n-2)}$$

Decision rule: Reject $H_0 : \beta = 0$ if

$$F = \frac{\mathrm{ESS}/1}{\mathrm{USS}/(n-2)} \left(= \frac{b^2 \sum_{i=1}^{n} x_i^2}{S^2} \right) > F_{(1,n-2),\alpha=0.05}, \text{ (if } \alpha = 0.05\text{)}.$$

The ANOVA table for testing $H_0 : \beta = 0$ is in Appendix 1D.

ANOVA

- t and F distributions

When the null hypothesis $H_0 : \beta = 0$ is true, then

$$t_{(m)}^2 = F_{(1,m)}$$

$$t^2 = \left(\frac{b}{\text{se}(b)}\right)^2 = \left(\frac{b}{S/\sqrt{\Sigma x_i^2}}\right)^2 = \left(\frac{b\sqrt{\Sigma x_i^2}}{S}\right)^2 = \frac{b^2 \Sigma x_i^2}{S^2} = F$$

ANOVA

- F and R^2 relationship

F statistic for $H_0 : \beta = 0$ (in $Y_i = \alpha + \beta X_i + \varepsilon_i$) can be rewritten in terms of R^2.

$$F = \frac{\text{ESS}/1}{\text{USS}/(n-2)} = \frac{\dfrac{\text{ESS}}{\text{TSS}}\Big/1}{\dfrac{\text{USS}}{\text{TSS}}\Big/(n-2)} = \frac{R^2/1}{(1-R^2)/(n-2)} = \frac{R^2}{1-R^2}(n-2) \qquad (2)$$

This F-R^2 relationship shows that $H_0 : \beta = 0$ can also be tested using the R^2 value.