Econometrics

Chapter 2: Review on Statistics

In Choi

Sogang University

Fall 2009

Random variables

- Random variable (r.v.) : A function that assumes numerical values associated with the random outcome of an experiment.

$$X : \Omega \rightarrow R$$

Random variables

Example

Tossing a coin twice

$$\Omega = \{HH, HT, TH, TT\}.$$

Define X as the number of heads out of two tosses :

$\omega = HH$	$X = 2$
$\omega = HT$	$X = 1$
$\omega = TH$	$X = 1$
$\omega = TT$	$X = 0$

where $P(X = 2) = 1/4, P(X = 1) = P(\omega = HT) + P(\omega = TH) = 1/2, P(X = 0) = 1/4.$

Random variables

Example

Define X = number of heads − number of tails:

$\omega = HH$	$X = 2 - 0 = 2$
$\omega = HT$	$X = 1 - 1 = 0$
$\omega = TH$	$X = 1 - 1 = 0$
$\omega = TT$	$X = 0 - 2 = -2$

where $P(X = 2) = P(HH) = 1/4, P(X = 0) = P(HT) + P(TH) = 1/2, P(X = -2) = P(TT) = 1/4.$

Random variables

Example

These can be written as

x	$P(X = x)$
2	$\frac{1}{4}$
0	$\frac{1}{2}$
-2	$\frac{1}{4}$

- This table is called the probability distribution of X.
- Notation

$$\begin{aligned} X &: \quad Random\ variable \\ x &: \quad Values\ \ X\ \ takes \end{aligned}$$

Random variables

- Discrete r.v. : A r.v. that takes finite or countably infinite numbers of distinct values, such as

$$\{x_1, x_2, \ldots, x_n\} \text{ or } \{x_1, x_2, x_3, \ldots\}.$$

(A set A is countably infinite if there is a one-to-one correspondence between A and $\{1, 2, 3, \ldots\}$.)

- Continuous r.v. : A r.v. that can take any value in an interval.

Probability distributions for discrete random variables

Example

Bernoulli distribution (A special case of the binomial distribution)
Binomial distribution
Poisson distribution: An approximation of the binomial distribution

Probability distributions for discrete random variables

- The probability distribution for a discrete r.v.: A graph, table or formula that specifies the probability associated with each possible values of the r.v.

Example

Tossing two coins

$$X = \# \text{ of heads } - \# \text{ of tails}$$

x	$p(x)$
-2	$\frac{1}{4}$
0	$\frac{1}{2}$
2	$\frac{1}{4}$

Probability distributions for discrete random variables

- Requirements for $p(x)$ (probability density function; pdf)
  1. $p(x) \geq 0$ for all values of X
  2. $\sum p(x) = 1$, where the summation is over all possible values of X.
- Summary measures for probability distribution
  Population mean : μ
  Population variance :σ^2

Probability distributions for discrete random variables

- $\mu = E(X) = \sum xp(x)$
  $E(\cdot)$ is called the expectation operator.
  The parameter μ denotes the "center" of the probability distribution.

Example

Tossing two coins
Let $X = \#$ of heads. The pdf of X is

x	$p(x)$
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

.

Thus, $E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1.$

Probability distributions for discrete random variables

- $\sigma^2 = E(X-\mu)^2 = \sum(x-\mu)^2 p(x)$ where $\mu = E(X)$.
  The variance measures the degree of dispersion of the given distribution.
  The standard deviation of the r.v. X is $\sigma = \sqrt{\sigma^2}$.

Probability distributions for discrete random variables

Example

Tossing two coins

Let $X = \#$ of heads. The pdf of X is

x	$p(x)$
2	$\frac{1}{4}$
1	$\frac{1}{2}$
0	$\frac{1}{4}$

From this, we obtain

$$\mu = 1, \quad \sigma^2 = (2-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{1}{2} + (0-1)^2 \times \frac{1}{4} = 1/2$$

Instead of $E(X-\mu)^2$, we may use the notation $Var(X)$.

- Rules for $E(\cdot)$ and $Var(\cdot)$
  1. $E(X+Y) = E(X) + E(Y)$
  2. $E(aX) = aE(X)$, where a is a constant.
  3. $Var(X+Y) = Var(X) + Var(Y) + 2E(X-\mu_x)(Y-\mu_Y)$, where $\mu_x = E(X)$ and $\mu_Y = E(Y)$.
  4. $Var(aX) = a^2 Var(X)$, where a is a constant.

Probability distributions for continuous random variables

- When $P(a < X < b) = \int_a^b f(x)dx$, $f(x)$ is called the probability density function of X.
  Because $P(-\infty < X < \infty) = 1$, $\int_{-\infty}^{\infty} f(x)dx = 1$.
  Because $P(X = a) = P(X = b) = 0$, we have

$$\begin{aligned} P(a &< X < b) \\ &= P(a \leq X < b) \\ &= P(a < X \leq b) \\ &= P(a \leq X \leq b). \end{aligned}$$

  The pdf function $f(x)$ can be estimated by using data. The crudest estimator is the histogram. (density estimation)

Probability distributions for continuous random variables

- The function $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$ is called the cumulative distribution function of the r.v.X.
  1. $0 \leq F(x) \leq 1$
  2. $F(x) \leq F(x')$, when $x \leq x'$ (non-decreasing)
  3. $F(-\infty) = 0, F(\infty) = 1$.

Probability distributions for continuous random variables

- For a continuous r.v. X with pdf $f(x)$,
  1. $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)$
  2. $$\begin{aligned} \sigma^2 &= E(X-\mu)^2 \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\ &= \int_{-\infty}^{\infty} x^2 f(x)dx - 2\mu \int_{-\infty}^{\infty} xf(x)dx + \mu^2 \int_{-\infty}^{\infty} f(x)dx \\ &= \int_{-\infty}^{\infty} x^2 f(x)dx - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Multivariate pdf

- Let X and Y be a discrete r.v.'s. The function

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

is called the joint pdf of X and Y.

Multivariate pdf

Example

Tossing two fair coins with 1 and 0 in each side

X = sum of the two sides

Y = |difference of the two sides|

The joint pdf of X and Y is

		X 0	1	2
Y	0	1/4	0	1/4
	1	0	1/2	0

- Expectation of the function of X and Y is defined by

$$E(g(X,Y)) = \sum_x \sum_y g(x,y) f(x,y).$$

Multivariate pdf

Example

$$E(XY) = 0 \cdot 1/4 + 0 \cdot 1/4 + 1 \cdot 1/2 = 1/2$$

$$E(X + Y^2) = (0 + 0^2) \cdot 1/4 + (2 + 0^2) \cdot 1/4 + (1 + 1^2) \cdot 1/2 = 3/2.$$

Multivariate pdf

- Marginal pdfs

$$f_X(x) = P(X = x) = \sum_y f(x, y)$$

$$f_Y(y) = P(Y = y) = \sum_x f(x, y)$$

Example

Tossing two coins

$$\begin{aligned} f_X(0) &= 1/4,\ f_X(1) = 1/2,\ f_X(2) = 1/4, \\ f_Y(0) &= 1/2,\ f_Y(1) = 1/2. \end{aligned}$$

Multivariate pdf

- When X and Y are continuous r.v.'s, $f(x, y)$ is called the joint pdf of X and Y, if

$$P(X \leq a, Y \leq b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y) dx dy.$$

- Marginal pdf's are:

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

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

- Expected value of $g(x, y)$ is

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy.$$

Multivariate pdf

- Covariance:

$$\begin{aligned} Cov(X, Y) &= E(X - \mu_X)(Y - \mu_Y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mu_x)(y - \mu_y)f(x, y)dxdy, \end{aligned}$$

where $f(x, y)$ is the joint pdf of X and Y and $\mu_X = E(X), \mu_Y = E(Y)$.

Multivariate pdf

- $X$and Y (discrete or continuous) are independent, if

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

- When X and Y are independent,.

1. $Cov(X,Y) = 0$, because

$$\begin{aligned} Cov(X,Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}(x-\mu_x)f_X(x)dx\right](y-\mu_y)f_Y(y)dy \\ &= 0. \end{aligned}$$

Note that $\mu_x = \int_{-\infty}^{\infty} xf_X(x)dx$ and that $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

2. $E(XY) = E(X)E(Y)$, because

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy \\ &= \int_{-\infty}^{\infty} yf_Y(y)dy \int_{-\infty}^{\infty} xf_X(x)dx = E(X)E(Y). \end{aligned}$$

Multivariate pdf

- Correlation coefficient

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

1. $-1 \leq \rho \leq 1$.
2. ρ contains information about the linear relationship between X and Y.

Conditional pdf and conditional expectation

- Let (X, Y) be a discrete bivariate random vector with joint pdf $f(x, y)$, and marginal pdf's $f_X(x)$ and $f_Y(y)$. Assume $f_X(x) > 0$. The conditional pdf of Y given $X = x$ is defined by

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

  This denotes the probability that Y takes value y given that X takes value x.
  This is a pdf because (i) $f_{Y|X}(y \mid x) \geq 0$ and (ii) $\sum_y f_{Y|X}(y \mid x) = 1$.

Conditional pdf and conditional expectation

Example

Define the joint pdf of (X, Y) by

$$\begin{aligned} f(0,10) &= f(0,20) = \frac{2}{18}, f(1,10) = f(1,30) = \frac{3}{18}, \\ f(1,20) &= \frac{4}{18}, f(2,30) = \frac{4}{18}. \end{aligned}$$

The marginal pdf's are $f_X(0) = f(0,10) + f(0,20) = \frac{4}{18}$; $f_X(1) = f(1,10) + f(1,30) + f(1,20) = \frac{10}{18}$; $f_X(2) = f(2,30) = \frac{4}{18}$. The probability of Y taking 10 given that X takes 0 is $f_{Y|X}(10 \mid 0) = \frac{f(0,10)}{f_X(0)} = \frac{2/18}{4/18} = \frac{1}{2}$.

Conditional pdf and conditional expectation

- Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x, y)$, and marginal pdf's $f_X(x)$ and $f_Y(y)$. Assume $f_X(x) > 0$. The conditional pdf of Y given $X = x$ is defined by

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

  This is also a pdf.

Conditional pdf and conditional expectation

- The conditional expected value of $g(Y)$ given that $X = x$ is denoted by $E(g(Y) \mid x)$ and is given by

$$E(g(Y) \mid x) = \sum_{y} g(y) f_{Y|X}(y \mid x) \text{ and } E(g(Y) \mid x) = \int_{-\infty}^{\infty} g(y) f_{Y|X}(y$$

Example

Conditional mean

$$E(Y \mid x) = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x) dy.$$

Conditional variance

$$Var(Y \mid x) = E(Y^2 \mid x) - (E(Y \mid x))^2 .$$

Normal (Gaussian) distribution

- A r.v. X which has the pdf

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

  is called a normal r.v. We have $E(X) = \mu$ and $Var(X) = \sigma^2$. When X is a normal r.v. with mean μ and variance σ^2, write

$$X \sim N(\mu, \sigma^2) \text{ or } X \equiv N(\mu, \sigma^2)$$

- The pdf of a normal r.v. looks like a bell curve symmetric around μ.
- A normal r.v. with $\mu = 0, \sigma^2 = 1$ is called a standard normal r.v., and its distribution is called a standard normal distribution.

Normal (Gaussian) distribution

- When $X \sim N(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma}$ has a standard normal distribution.

$$E(Z) = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}(\mu-\mu) = 0$$

$$Var(Z) = E\left[\frac{X-\mu}{\sigma} - 0\right]^2 = \frac{1}{\sigma^2}E(X-\mu)^2 = \frac{\sigma^2}{\sigma^2} = 1$$

- When $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$ and X and Y are independent,

$$aX + bY \sim N\left(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2\right)$$

where a and b are constant.

Normal (Gaussian) distribution

- Reading the standard normal table:
  Suppose that $Z \sim N(0,1)$.
  (i) $P(0 < Z < 0.27) = 0.1064$
  (ii) $P(0 < Z < 1.22) = 0.3888$
  (iii) $P(Z > 0.27) = P(0 \leq Z < \infty) - P(Z < 0.27) =$
  $0.5 - 0.1064 = 0.3936$
  (iv) $P(Z \geq 1.22) = 0.5 - 0.3888 = 0.1112$
  (v) $P(-1 \leq Z \leq 1) = P(-1 \leq Z < 0) + P(0 \leq Z \leq 1)$
  $= P(0 < Z \leq 1) + P(0 \leq Z \leq 1) = 2 \times 0.3413 = 0.68.$
  (vi) $P(Z < -2) = P(Z \geq 2) = 0.5 - P(0 \leq Z \leq 2) =$
  $0.5 - 0.4772 = 0.0228$
  $(vii)P(|Z| > 2) = P(Z \geq 2) + P(Z \leq -2) = 0.0228 + 0.0028 =$
  $0.0456.$

Normal (Gaussian) distribution

- Calculating probabilities for a normal r.v.
  When $X \sim N(1, 2^2)$,
  (i) $P(X < 3) = P(\frac{X-1}{2} < \frac{3-1}{2}) = P(Z < 1) = P(-\infty < Z < 0) + P(0 \leq Z < 1)$
  $= 0.5 + 0.3413 = 0.8413.$
  (ii) $P(-3 < X < 5) = P(-2 < Z < 2) = 2 \times P(0 < Z < 2)$
  $= 2 \times 0.4772 = 0.9544$
  (iii) $P(X < -3) = P(Z < -2) = P(Z > 2) = 0.5 - P(0 < Z < 2) = 0.5 - 0.4772 = 0.0228.$

Normal (Gaussian) distribution

- When $X \sim N(\mu, \sigma^2)$,

$$P(\mu - \sigma < X < \mu + \sigma) = P(-1 < Z < 1) \approx 0.68$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = P(-2 < Z < 2) \approx 0.95$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) \approx 1$$

Normal (Gaussian) distribution

Example

Assume that a light bulb lasts X hours where X ˜ $N\left(1000, 50^2\right)$. Find the probability that the light bulb lasts more than1100 hours.

$$\begin{aligned} P(X &> 1100) = P(\frac{X-1000}{50} > \frac{1000-1000}{50}) \\ &= P(Z > 2) = 0.5 - P(0 < Z < 2) \\ &= 0.5 - 0.4772 = 0.0228. \end{aligned}$$

Normal (Gaussian) distribution

Example

Find z_0 such that $P(-z_0 < Z < z_0) = 0.95$

$$P(|Z| > z_0) = 1 - .95 = 0.05$$

Then

$$P(Z > z_0) = 0.05/2 = 0.025$$

or

$$0.5 - P(0 < Z < z_0) = 0.025.$$

These imply $P(0 < Z < z_0) = 0.475$, giving $z_0 = 1.96$.

Sampling distributions

- Parameter : A numerical measure describing a population
  - $X \sim N(\mu, \sigma^2) => \mu$ and σ^2 are parameters
  - $X \sim B(n, p) =>$ n and p are parameters.
- Sample statistic : A numerical descriptive measure of a sample. Calculated from the observations in the sample.

Example

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

- Sample statistics are r.v.'s, and their probability distributions are called sampling distributions.

Example

Suppose that $X \sim B(1, 1/2)$

x	$p(x)$
0	1/2
1	1/2

A random sample of size 2 ($n = 2$) is selected from the population. What is the sampling distribution of the sample mean $\overline{X}$?

Sampling distributions

Example

(x_1, x_2)	$\overline{x}$	$p(x_1, x_2)$
$(0, 0)$	0	1/4
$(0, 1)$	1/2	1/4
$(1, 0)$	1/2	1/4
$(1, 1)$	1	1/4

Thus,

$\overline{x}$	$p(\overline{x})$
0	1/4
1/2	1/2
1	1/4

Sampling distributions

- The sampling distribution of $\overline{X}$

When $E(X_i) = \mu$, $Var(X_i) = \sigma^2$ for all i and X_i are independent, $E(\overline{X}) = \mu$ and $Var(\bar{X}) = \sigma^2/n$.

- Interpretation :
  $E(\overline{X}) = \mu$: The population mean μ, which $\overline{X}$ estimates, is at the center of the distribution of $\bar{X}$. Therefore, the realized values of $\bar{X}$ will be close to μ. $\bar{X}$ is an unbaiased estimator of μ.
  $Var(\overline{X}) = \sigma^2/n$: As n gets larger, the distribution of $\overline{X}$ will be more centered around μ.

Sampling distributions

- Two theorems for the sampling distribution of $\overline{\mathbf{X}}$

Theorem

When $X_i \sim iidN(\mu, \sigma^2)$ for all i, $\overline{X} = \sum_{i=1}^{n} X_i / n \sim N(\mu, \sigma^2 / n)$ for all sample size n.

Note: In the theorem above, we assume normal distribution for X_i.

Sampling distributions

Theorem

When X_i has the same distribution (NOT necessarily normal) with mean μ and variance σ^2 for all i (we write $X_i \sim iid(\mu, \sigma^2)$),

$$\overline{X} = \sum X_i / n \simeq N(\mu, \sigma^2 / n)$$

for large n.

The sign " $\simeq$ " means

$$P(\overline{X} < x) \approx P(N(\mu, \sigma^2/n) < x)$$

when n is large.

This theorem is called the Central Limit Theorem(CLT).

The central limit theorem holds for any distribution of X_i as long as it has finite mean and variance.

The Chi–square distribution

- When $X_i \sim iidN(0,1)$, $\sum_{i=1}^{k} X_i^2$ follows the chi-square distribution with degrees of freedom k. We write $\sum_{i=1}^{k} X_i^2 \sim \chi^2(k)$.
- The distribution is skewed especially when k is small.
- $E(\chi^2(k)) = k$, $Var(\chi^2(k)) = 2k$.
- If Z_1 and Z_2 are two independent chi-square variable with k_1 and k_2 degrees of freedom, respectively, $Z_1 + Z_2 \sim \chi^2(k_1 + k_2)$.

The t-distribution

- When $X \sim N(0,1)$, $Y \sim \chi^2(k)$ and X and Y are independent, $Z = X/\sqrt{Y/k}$ follows the t-distribution with k degrees of freedom.
- The t-distribution is symmetric around 0.
- $E(Z) = 0$, $Var(Z) = \frac{k}{k-2}$.

The t-distribution

- As k gets larger, the t-distribution becomes closer to a standard normal distribution. When k is larger than 30, virtually no different from a standard normal distribution.
- When $X_i \sim iidN(\mu, \sigma^2)$ $(i = 1, ..., n)$, let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Then,

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

  follows the t-distribution with $n - 1$ degrees of freedom.

The F-distribution

- If $X \sim \chi^2(k_1)$, $Y \sim \chi^2(k_2)$, and X and Y are independent, $\frac{X/k_1}{Y/k_2}$ follows the F-distribution with degrees of freedom (k_1, k_2).
- The F-distribution is skewed like the chi-square distribution.

- Inference: Extracting information about population parameters from sample.

Example

Suppose that X_i is the difference in the average temperature of the years 1990 and 2000 for region i and that $X_i \sim iidN(\mu, \sigma^2), i = 1, ..., 50$.
We use $\bar{X} = \sum_{i=1}^{50} X_i/50$ to estimate μ, and $s^2 = \sum_{i=1}^{50} (X_i - \bar{X})^2/49$ to estimate σ^2.
To test the hypothesis that there is no temperature difference, we test the hypothesis $H_0 : \mu = 0$ by using the test statistic $\frac{\bar{X}}{s/\sqrt{50}}$.

Estimation

- Estimation
  1. Point estimation: This provides a single numerical value about a population parameter.
  2. Interval estimation: This provides an interval for a population parameter. The interval is called a confidence interval. The confidence interval assesses the accuracy of point estimation. The confidence interval requires sampling distribution of the estimator for the population parameter.

Large-sample confidence intervals

- When $X_i \sim iid(\mu, \sigma^2)$, we have for large n

$$\bar{X} \simeq N(\mu, \frac{\sigma^2}{n}).$$

Here $\simeq$ means "approximately distributed as".
Alternativley, one may write

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \simeq N(0, 1).$$

Large-sample confidence intervals

- Therefore,

$$P(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \leq z_{\alpha/2}) \simeq 1-\alpha,$$

  which gives

$$P(\bar{X} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}) \simeq 1-\alpha.$$

The interval

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

is called the $100(1-\alpha)\%$ confidence interval for μ. $100(1-\alpha)\%$ is called the confidence level.

Large-sample confidence intervals

- In practice, we use $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$ instead of σ because σ is not known.

Confidence level $(100(1-\alpha))$	α	$\alpha/2$	$z_{\alpha/2}$
90%	0.1	0.05	1.645
95%	0.05	0.025	1.96
99%	0.01	0.005	2.575

Large-sample confidence intervals

Example

A random sample of 90 observations produced a mean $\bar{x} = 25.9$ and a standard deviation $s = 2.7$.

1. 95% CI

$$\begin{aligned} & (25.9 - 1.96 \times \frac{2.7}{\sqrt[2]{90}},\ 25.9 + 1.96 \times \frac{2.7}{\sqrt[2]{90}}) \\ = & (25.9 - 0.56,\ 25.9 + 0.56) \\ = & (25.34,\ 26, 46) \end{aligned}$$

2. 90% CI

$$\begin{aligned} & (25.9 - 1.645 \times \frac{2.7}{\sqrt[2]{90}},\ 25.9 + 1.645 \times \frac{2.7}{\sqrt[2]{90}}) \\ = & (25.9 - 0.47,\ 25.9 + 0.47) \\ = & (25.43,\ 26, 37) \end{aligned}$$

- Interpretation for $100(1-\alpha)\%$ CI for μ :
  "We are $100(1-\alpha)\%$ sure that the CI contain μ."
  "If we formulate $100(1-\alpha)\%$ CI's for μ by using different samples for the same population, approximately $100(1-\alpha)$ CI's will contain μ."

Small-sample confidence intervals

- Assume $X_i \sim iidN(\mu, \sigma^2)$ for all i. Then, $t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ has the t-distribution with $(n-1)$ degrees of freedom. Thus,

$$P(-t_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n}}} \leq t_{\alpha/2}) \simeq 1-\alpha$$

Equivalently,

$$P(\bar{X} - t_{\alpha/2}\sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}\sqrt{\frac{s^2}{n}}) \simeq 1-\alpha.$$

The interval

$$(\bar{X} - t_{\alpha/2}\sqrt{\frac{s^2}{n}},\ \bar{X} + t_{\alpha/2}\sqrt{\frac{s^2}{n}})$$

is called the $100(1-\alpha)\%$ confidence interval for the population mean μ.

Small-sample confidence intervals

Example

The following random sample was selected from a normal distribution: 3, 7, 5, 2, 4.
$\bar{x} = 4.2, s = 1.92$
The 90% CI for the population mean μ is

$$(4.2 - 2.132 \times \frac{1.92}{\sqrt[2]{5}},\ 4.2 + 2.132 \times \frac{1.92}{\sqrt[2]{5}}).$$

The 95% CI for the population mean μ is

$$(4.2 - 2.776 \times \frac{1.92}{\sqrt[2]{5}},\ 4.2 + 2.776 \times \frac{1.92}{\sqrt[2]{5}}).$$

Test of hypothesis using a single sample

- Test of hypothesis is used in decision-making.

Examples

1. Is tobacco smoking related to lung cancer?
2. Is caffeine addictive?
3. Are high interest rates desirable for stabilizing exchange rates?

- Hypothesis: A statement about population parameter
  1. Null hypothesis: The hypothesis that will be accepted unless the data provide convincing evidence against it.
  2. Alternative hypothesis: The hypothesis that will be accepted only if the data provide convincing evidence for its truth.

Test of hypothesis using a single sample

- Test statistic: A random variable which can differentiate the null from the alternative.
- Suppose that $X_i \sim iid\ N(\mu, \sigma^2)$ and that σ is known. The sample mean $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ is close to μ whatever the true value of μ is. Therefore, to test whether or not the null hypothesis

$$H_0 : \mu = \mu^o$$

is true, compare $\bar{X}$ and μ^o. Note that μ^o is chosen by a researcher, so it is known. Let us assume we are interested in the alternative hypothesis

$$H_1 : \mu > \mu^o.$$

- When $\bar{X} - \mu^o$ is "large", reject H_0 against H_1. We use the standardized version of $\bar{X} - \mu^o$,

$$Z = \frac{\bar{X} - \mu^o}{\sqrt{Var(\bar{X})}} = \frac{\bar{X} - \mu^o}{\sigma / \sqrt{n}}.$$

Test of hypothesis using a single sample

- Why standardize by $\sqrt{Var(\bar{X})}$?

1. Z usually has a known distribution (normal, t-distribution). In the example, $Z \sim N(0, 1)$.
2. Z has some desirable properties when $\sqrt{Var(\bar{X})}$ is used as a measure of standardization.

Test of hypothesis using a single sample

- Decision procedure: When a test statistic is larger than a certain value(c_α), reject H_0. Otherwise accept it.
- How can we select c_α? We select c_α such that

$$P(Z > c_\alpha | H_0) = \alpha.$$

  We call α the Type I error. This is the error of rejecting H_0 when it is indeed true.
- The Type I error is set to be 0.1, 0.05 or 0.01 in applications.
- Suppose that Z follows a standard normal distribution and that the null is rejected when $P(Z > c_\alpha | H_0) = \alpha$. Then,

$$\alpha = 0.05 \Rightarrow c_\alpha = 1.645; \ \alpha = 0.1 \Rightarrow c_\alpha = 1.28.$$

Test of hypothesis using a single sample

- If the alternative hypothesis is

$$H_1 : \mu < \mu^o,$$

we reject the null hypothesis when Z takes a value smaller than C_α which is determined by the relation

$$P(Z < c_\alpha | H_0) = \alpha.$$

- If the alternative hypothesis is

$$H_1 : \mu \neq \mu^o,$$

we reject the null hypothesis when $|Z|$ takes a value larger than c_α which is determined by the relation

$$P(|Z| < c_\alpha | H_0) = \alpha.$$

Test of hypothesis using a single sample

- Steps for hypothesis tests
  1. Specify the null and alternative hypotheses.
  2. Select a test statistic.
  3. Select the level of Type I error (significance level; usually, 0.1, 0.05 or 0.01).
  4. When the statistic belongs to the rejection region, reject the null.
- Type II error: The error of not rejecting H_0, when H_1 is true. It is $P(Z > c_\alpha | H_1)$ if the the alternative hypothesis is $H_1 : \mu > \mu^o$. Because the distribution of the test statistic under the alternative is unknown, Type II error cannot be computed and cannot be controlled. Tests are called powerful if their Type II errors are small.

Test of hypothesis using a single sample

- Classification of errors in hypothesis testing:

		True state	
		H_0	H_1
Decision	H_0	OK	Type II error
	H_1	Type I error	OK

- Decision-makings based on hypothesis testing are subject to these errors. Type I error can be controlled by a researcher, but not the Type II error.

Test of hypothesis using a single sample

- Large-sample test hypothesis about a population mean

As a test statistic, use

$$Z = \frac{\bar{X} - \mu^o}{s/\sqrt{n}}.$$

When n is large, we know that

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

1. $H_0 : \mu = \mu^0$ versus $H_1 : \mu < \mu^0$: one-sided statistical test.
2. $H_0 : \mu = \mu^0$ versus $H_1 : \mu > \mu^0$: one-sided statistical test.
3. $H_0 : \mu = \mu^0$ versus $H_1 : \mu \neq \mu^0$: two-sided statistical test.

Test of hypothesis using a single sample

Example

Performance of filling machines for cereal

$$n = 100,\ \bar{x} = 11.85\ oz,\ s = 0.5\ oz.$$

1. $H_0 : \mu = 12$ versus $H_1 : \mu \neq 12$.

$$Z = \frac{\bar{x} - 12}{s/10} = \frac{11.85 - 12}{.5/10} = -3.0.$$

Because $Z < -1.645$, reject H_0.

2. $H_0 : \mu = 12$ versus $H_1 : \mu < 12$. $Z < -1.28 \Rightarrow$ Reject H_0, at the 10% significance level.

3. $H_0 : \mu = 12$ versus $H_1 : \mu > 12$. $Z < 1.28 \Rightarrow$ Do not reject H_0.

Test of hypothesis using a single sample

- P-values: The extent to which the test statistic disagrees with the null hypothesis
- Higher p-values imply more supportive evidence for H_0.

1. $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$.

$$\text{P-value} : P(N(0,1) > z)$$

   where z is an observed value of the test Z.

2. $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$

$$\text{P-value} : P(N(0,1) < z).$$

3. $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

$$\text{P-value} : 2 \times P(N(0,1) > z \mid z > 0) \text{ or } 2 \times P(N(0,1) < z \mid z < 0)$$

Example

H_0: $\mu = 12$ versus H_1: $\mu < 12$. P-value $= P((N(0,1) < -3.0) = 0.0013$.

Test of hypothesis using a single sample

- Small-sample test of hypothesis about a population mean.

Assume that $X_i \sim N(\mu, \sigma^2)$, $i = 1, , ..., n$. The null hypothesis is

$$H_0 : \mu = \mu^o$$

Test statistic:

$$t = \frac{\bar{x} - \mu^o}{s/\sqrt{n}}.$$

Under the normal assumption, the t-test has student's t-distribution with $n - 1$ degrees of freedom.

Test of hypothesis using a single sample

Example

$\bar{x} = 15.42$, $s = 0.16$, $n = 17$.

1. H_0: $\mu = 15.5$ versus H_1: $\mu \neq 15.5$.

$$t = \frac{15.42 - 15.5}{.16/\sqrt[2]{17}}. = -2.06.$$

Because $-2.06 > -2.12 = -t_{\frac{0.05}{2}}$, do not reject the null hypothesis.

2. H_0: $\mu = 15.5$ versus H_1: $\mu < 15.5$.

$$t = -2.06.$$

Because $-2.06 < -1.746$, reject the null at the 5% significance level.