

Sampling Distributions and Central Limit Theorem CLT

* Statistics concerns conclusions and predictions resulting from choice outcomes.

$\Downarrow$
 constitute a subset
 $\Downarrow$
 sample

* By using random samples, draw generalizations about population; from which they come from (evidence-based policy making)

* " " build up certain functions of them, which are used to estimate or make decisions about unknown population parameters

i.e. estimate a population mean μ $\Rightarrow$ obtain random sample of size n

$y_1, y_2, y_3, \dots, y_n \Rightarrow$ estimate μ from sample mean is

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

* Goodness of the estimate depends on behavior of random variables $y_1, y_2, y_3, \dots, y_n$ effect of this behavior on

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

* Random variable $\bar{y}$ is a function of random variables $y_1, y_2, \dots, y_n$ and constant sample size n .

* A statistic is a function of observable random variables in a sample and known constants.

* " " is used to make inferences about unknown population parameters.

* estimate vs. estimation

$\downarrow$
value of function

$\downarrow$
function

* Probability distribution of statistic is called sampling distribution which gives theoretical model of relative frequency histogram of possible values of statistic that we would observe through repeated sampling.

* Possible to assume that observable random variables in a random sample of $y_1, y_2, \dots, y_n$ are independent with a common normal density function.

* If $y_1, y_2, \dots, y_n$ be a random sample of size n from a normal population with mean μ and variance σ^2 , then

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ is called normally distributed

with $E(\bar{y}) = \mu_{\bar{y}} = \mu$

$$\text{Var}(\bar{y}) = \sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}$$

$\bar{y}$ is normally distributed with $\mu_{\bar{y}} = \mu$ and $\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}$

Then,

$$z = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma_{\bar{y}}} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \sqrt{n} \left(\frac{\bar{y} - \mu}{\sigma} \right) \sim N(0, 1)$$

Ex: Show that if $y_{11}, y_{12}, \dots, y_{1n_1}$ and $y_{21}, y_{22}, \dots, y_{2n_2}$ are independent random variables with first n_1 constitute a random sample from an infinite population with mean μ_1 , variance σ_1^2 and other n_2 constitute a random sample from an infinite population with mean μ_2 , variance σ_2^2

a) $E(\bar{y}_1 - \bar{y}_2) = \mu_1 - \mu_2$

let $X = \bar{y}_1 - \bar{y}_2$

$$E(X) = E(\bar{y}_1 - \bar{y}_2) = \frac{1}{n_1} \sum E(y_{1i}) - \frac{1}{n_2} \sum E(y_{2i})$$

$\xrightarrow{\frac{1}{n_1} \sum y_{1i}} \quad \xrightarrow{\frac{1}{n_2} \sum y_{2i}}$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$\xrightarrow{\frac{1}{n_1} \sum y_{1i}} \quad \xrightarrow{\frac{1}{n_2} \sum y_{2i}}$

$$= \frac{n_1}{n_1} \mu_1 - \frac{n_2}{n_2} \mu_2 = \mu_1 - \mu_2$$

b) $\text{Var}(\bar{y}_1 - \bar{y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$$\text{Var}(X) = \text{Var}(\bar{y}_1 - \bar{y}_2) = \sum \frac{1}{n_1^2} \text{Var}(y_{1i}) + \sum \frac{1}{n_2^2} \text{Var}(y_{2i})$$

$$= \frac{1}{n_1^2} \cdot n_1 \cdot \sigma_1^2 + \frac{1}{n_2^2} \cdot n_2 \cdot \sigma_2^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\text{Var}(\bar{y}_1 + \bar{y}_2) = \text{Var}(\bar{y}_1) + \text{Var}(\bar{y}_2)$$

$$\text{Var}(\bar{y}_1) = \text{Var}\left(\frac{\sum y_{1i}}{n_1}\right) \rightarrow \frac{1}{n_1^2} \text{Var}\left(\sum y_{1i}\right)$$

$\xrightarrow{\frac{1}{n_1^2} \sum \sigma_1^2}$

$$\text{Var}(\bar{y}_1) + \text{Var}(\bar{y}_2)$$

$$f(x) = (1-p)^{1-x} \cdot p^x \quad x=0,1$$

$$e(x) = p$$

$$V(x) = p(1-p)$$

Ex) If $y_1, y_2, \dots, y_n$ are independent random variables having identical Bernoulli distributions with parameter θ , then $\bar{y}$ is proportion of successes in n trials denoted by $\hat{\theta}$, verify that

a) $E(\hat{\theta}) = \theta$

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^1 e^{tx} \cdot p(x=x)$$

$$= e^0 \cdot \theta + e^{t\theta} \cdot (1-\theta)$$

$$= 1 + \theta(e^t - 1)$$

b) $\text{Var}(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$

moment gen func $M_y = (1 + \theta(e^t - 1))^n$, $m_y = (1 + \theta(e^{tn} - 1))^n$ n lone binden

$$m'_y = n(1 + \theta(e^{tn} - 1))^{n-1} \cdot \frac{\theta}{n} e^{tn}$$

$$= \theta(1 + \theta(e^{tn} - 1))^{n-1} e^{tn}$$

$m'_y(0) = \theta$

$$m''_y = \theta(1 + \theta(e^{tn} - 1))^{n-1} \cdot \frac{1}{n} \cdot e^{tn} + \theta \cdot e^{tn} (n-1) (1 + \theta(e^{tn} - 1))^{n-2} \cdot \frac{\theta}{n} e^{tn}$$

$m''_y(0) = \frac{\theta}{n} + \frac{\theta^2(n-1)}{n}$

$$\sigma^2 = \frac{\theta}{n} + \frac{\theta^2(n-1)}{n} - \theta^2 = \frac{\theta + n\theta^2 - \theta^2 - n\theta^2}{n} = \frac{\theta(1-\theta)}{n}$$

Ex: Show that the formula of sample variance can be written as

$$s^2 = \frac{n \left(\sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n y_i \right)^2}{n(n-1)}$$

$$\frac{n(n \cdot E(y^2)) - (n E(y))^2}{n(n-1)}$$

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - 2\bar{y} \sum_{i=1}^n y_i + n\bar{y}^2 \right)$$

$$= \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - \underbrace{2n\bar{y}^2}_{-n\bar{y}^2} + n\bar{y}^2 \right) = \frac{\sum y_i^2}{n-1} - \frac{n\bar{y}^2}{n-1} = \frac{n \sum y_i^2 - (\sum y_i)^2}{n(n-1)}$$

$\bar{y} = \frac{\sum y_i}{n}$
 $n \cdot \bar{y} = \sum y_i$

Let $Y_1, Y_2, \dots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2 and

$Z_i = \frac{(Y_i - \mu)}{\sigma}$ are independent s.t. normal random variables where $i = 1, 2, \dots, n$.

$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n \left(\frac{Y_i - \mu}{\sigma} \right)^2 \sim \chi^2$ distributed with n degrees of freedom.
Standard Normal Distribution

If Y has a standard normal distribution, Y^2 has a χ^2 distribution with $\nu = 1$ degrees of freedom.

If $Y_1, Y_2, \dots, Y_n$ are independent random variables having standard normal distribution, then

$$X = \sum_{i=1}^n Y_i^2 \sim \chi^2_{\nu=n}$$

If $Y_1, Y_2, \dots, Y_n$ are independent random variables with $\nu_1, \nu_2, \dots, \nu_n$ degrees of freedom, then

$$X = \sum_{i=1}^n Y_i^2 \sim \chi^2_{\nu_1 + \nu_2 + \dots + \nu_n}$$

If Y_1 and Y_2 are independent random variables and $Y_1 \sim \chi^2_{\nu_1}$ and $\nu_1 + Y_2 = \chi^2_{\nu}$, $\nu > \nu_1$, then $Y_2 \sim \chi^2_{\nu - \nu_1}$.

If $\bar{Y}$ and S^2 are mean and variance of a random sample of size n from a normal population with mean μ and variance σ^2 , then

① $\bar{Y}$ and S^2 are independent

② random variable $\bar{W} = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum (Y_i - \bar{Y})^2 \sim \chi^2_{n-1}$

$$\text{where } S^2 = \frac{1}{n-1} \sum (Y_i - \bar{Y})^2$$

Ex: Find median value of χ^2 corresponding to

a) 9 degree of freedom.

$$\chi^2_{0.50, 9} = 8.341283$$

b) 28 degree of freedom.

$$\chi^2_{0.50, 28} = 27.3263$$

c) 10 degree of freedom.

$$\chi^2_{0.50, 10} = 9.347826$$

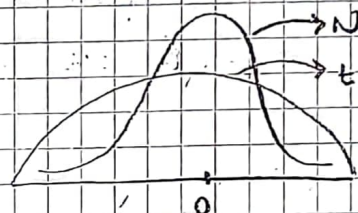
If w and z are independent random variables,

$$w \sim \chi^2, z \sim N(0,1)$$

$T = \frac{z}{\sqrt{w/n}}$ is having t dist. with ν degrees of freedom. (d.f.)

$$T = \frac{z}{\sqrt{w/n}} = \frac{\sqrt{n}(\bar{y} - \mu)/\sigma}{\sqrt{(n-1)s^2/\sigma^2}} = \sqrt{n} \left(\frac{\bar{y} - \mu}{s} \right) \sim t_{n-1}$$

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \text{ for } -\infty < t < \infty$$



t-distribution random variables has same exp. value as st. normal density.

st. normal distribution has variance 1, but variance of t-distributed random variable always exceed 1.

Ex: with 9 d.f., find value of t_1 .

a) shaded area in right is 0.05

$$t_{9,0.05} = 1.833$$

b) total shaded area is 0.05

$$1 - 0.025 = 0.975 \quad t_{9,0.025} = 2.26$$

c) total unshaded area is 0.99

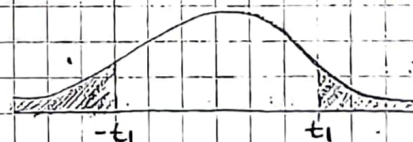
$$t_{9,0.005} = 3.250$$

d) shaded area on left is 0.01

$$t_{9,0.01} = 2.821$$

e) area to left of t_1 is 0.80

$$t_{9,0.20} = 1.383$$



If S_1^2 and S_2^2 are variances of independent random samples of size n_1 and n_2 from normal population with variances σ_1^2 and σ_2^2 ,

$$T = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 - S_1^2}{\sigma_1^2 - S_2^2} \sim F_{n_1, n_2 - 1}$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow T = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\bar{X} - \mu}{\sqrt{w}/\sigma}$$

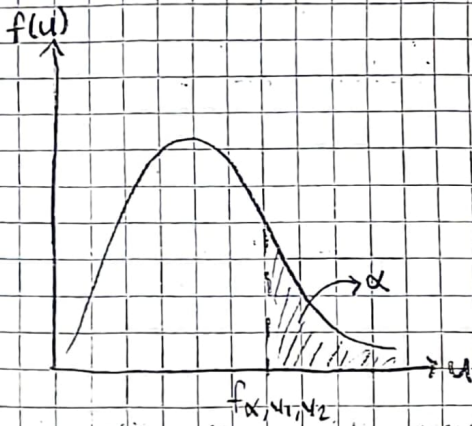
$$T = \frac{Z}{\sqrt{w/\sigma}}$$

If w_1 and w_2 are independent random variables having χ^2 distributions with ν_1, ν_2 of

$$F = \frac{w_1/\nu_1}{w_2/\nu_2} \sim F_{\nu_1, \nu_2}$$

Probability Density of F Distribution

$$f(F) = \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} f^{\frac{\nu_1}{2}-1} \left(1 + \frac{\nu_1}{\nu_2} f\right)^{-\frac{1}{2}(\nu_1 + \nu_2)}, \quad f > 0$$



Ex: $F_{7,9}^{0.05} = 3.08$

$F_{7,12}^{0.10} = 4.64$

$F_{2,14}^{0.005} = 3.74$

Ex: If X has an F distribution with ν_1 and ν_2 df show that $Y = \frac{1}{X}$ has an F distribution with ν_2 and ν_1 df.

$$\frac{1}{F} = \frac{\nu_1/\nu_2}{\nu_2/\nu_1} \sim F_{\nu_2, \nu_1}$$

Central Limit Theorem

When b becomes larger, we have more information on population and population parameters

→ i.e. $\bar{Y}$ becomes close to μ .

By using Chebyshev's Theorem,

for any positive constant c , probability that $\bar{Y}$ will take on a value between $\mu - c$ and $\mu + c$ is at least

$$1 - \frac{\sigma^2}{nc^2} \quad \left. \vphantom{1 - \frac{\sigma^2}{nc^2}} \right\} \text{this result is called as law of large numbers.$$

as $n \rightarrow \infty$,
this probability $\rightarrow 1$.

Central Limit Theorem concerns limiting distribution of standardized mean of n random variables $n \rightarrow \infty$.

If $Y_1, Y_2, \dots, Y_n$ constitute a random sample from an infinite population with mean μ and variance σ^2 , moment-generating function $M_Y(t)$, then the limiting (approximating) distribution is

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \text{ as } n \rightarrow \infty \text{ is standard normal distribution}$$

Therefore, CLT is sometimes interpreted as implying that distribution of $\bar{Y}$ approaches to a normal distribution as $n \rightarrow \infty$

→ Incorrect since $\text{var}(\bar{Y}) \rightarrow 0$ when $n \rightarrow \infty$, on the other hand, CLT justifies approximating distribution of $\bar{Y}$ with a normal distribution having mean μ and σ^2/n when n is large.

large small properties (asymmetric properties)

In practice, this approximation is used when $n \geq 30$ regardless of actual shape of population/density sampled.

For smaller values of n , approximation is questionable.

If $\bar{Y}$ is mean of a random sample of size n from a normal population with mean μ and variance σ^2 , its sampling distribution is a normal distribution with mean μ and variance σ^2/n .

Order Statistics

Consider a random sample of size n from an infinite population with a continuous density.

Suppose that we arrange values of $x_1, x_2, \dots, x_n$ according to size. If we look upon the smallest of x 's as value of random variable Y_1 , the next largest as value of Y_2 , next largest as Y_3 , ..., the largest value of Y_n . We refer these Y 's as order statistics.

$Y_1 \Rightarrow 1^{\text{st}}$ order statistics

$Y_2 \Rightarrow 2^{\text{nd}}$ order statistics

$Y_3 \Rightarrow 3^{\text{rd}}$ order statistics

$\vdots$

$Y_n \Rightarrow n^{\text{th}}$ order statistics

we are limiting these discussion to infinite populations with continuous densities, there are zero probability that any of two x 's will be alike.

Consider the case $n=2$,

$y_1 = x_1$ and $y_2 = x_2$ when $x_1 < x_2$

$y_1 = x_2$ and $y_2 = x_1$ when $x_2 < x_1$.

And $n=3$,

$y_1 = x_1$, $y_2 = x_2$ and $y_3 = x_3$ when $x_1 < x_2 < x_3$

$y_1 = x_1$, $y_2 = x_3$ and $y_3 = x_2$ when $x_1 < x_3 < x_2$

$y_1 = x_3$, $y_2 = x_2$ and $y_3 = x_1$ when $x_3 < x_2 < x_1$

Since we cannot write all the possibilities everytime, we derive a formula for probability density of r^{th} order such that $r = 1, 2, \dots, n$

For random samples of size n from an infinite population that has value $f(x)$ at x probability density of r^{th} order statistics, Y_r is given by

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!} \left[\int_{-\infty}^{y_r} f(x) dx \right]^{r-1} f(y_r) \left[\int_{y_r}^{\infty} f(x) dx \right]^{n-r}$$

for $-\infty < y_r < \infty$.

$$g_1(y_1) = n f(y_1) \left[\int_{y_1}^{\infty} f(x) dx \right]^{n-1} \quad \text{for } -\infty < y_1 < \infty$$

→ sampling distribution of y_1 , the smallest value in a random sample of size n .

$$g_n(y_n) = n f(y_n) \left[\int_{-\infty}^{y_n} f(x) dx \right]^{n-1} \quad -\infty < y_n < \infty.$$

→ sampling distribution of y_n , the largest value in a random sample of size n .

$t=2000, \dots, 2018$

per capita household income in Turkish provinces, $n=1, \dots, 81$.

(take 2018 as base year for ranking.)

$y_{1,2018}$	$y_{1,2018}$
$\vdots$	$\vdots$
$y_{n,2018}$	$y_{n,2018}$

In a random sample of size $n=2m+1$, sample median $\tilde{x}$ is y_{m+1} whose sampling distribution is

$$h(\tilde{x}) = \frac{(2m+1)!}{m!m!} \left[\int_{-\infty}^{\tilde{x}} f(x) dx \right]^m f(\tilde{x}) \left[\int_{\tilde{x}}^{\infty} f(x) dx \right]^m \quad \text{for } -\infty < \tilde{x} < \infty.$$

For random sample size $n=2m$, median is defined as $\frac{1}{2}(y_m + y_{m+1})$

we have an interesting result about sampling distribution of median, which holds population density is continuous and nonzero at population median, μ such that

$$\int_{-\infty}^{\mu} f(x) dx = \frac{1}{2}$$

For large n , sampling distribution of median for random sample of size $2m+1$ is approximately normal with mean μ and variance

$$\frac{1}{8 [f(\mu)]^2}$$

For random samples of size $2n+1$ from a normal population, we have $\mu = \bar{x}$

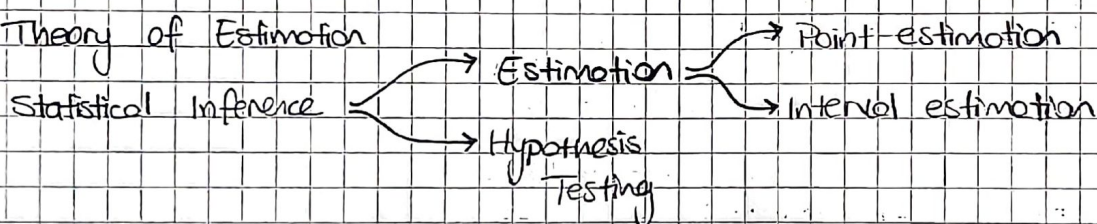
$$\text{so } f(\bar{x}) = f(\mu) = \frac{1}{\sigma \sqrt{2\pi}}$$

If we compare this with variance of mean which for random samples of size $2n+1$ from an infinite population is

$$\frac{\sigma^2}{2n+1}$$

✓ We find that for large samples from normal populations, the mean is more reliable than median; mean is subject to smaller chance functions than median.

Theory of Estimation



Purpose of point estimation is to use sample data and any a priori (non-sample) information we have about probability distribution in question to produce a value that our best estimate of unknown parameter value.

→ concerned about reliability of estimate.
(Reliability measures)

The Bias and Mean Square Error of Point Estimators

1. Bias

$\hat{\theta}$ is an unbiased estimator of θ if

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

midterm!
show that what is an unbiased estimator.

If $E(\hat{\theta}) \neq \theta$, then $\hat{\theta}$ is a biased estimator for θ .

$$E(\hat{\theta}) - \theta = \text{bias} \rightarrow \text{amount of bias}$$

Unbiasedness means that an estimator on average will yield the parameter value.

If underlying experiment repeated infinitely many times by drawing sample of size n average value of estimate $\hat{\theta}$ from all these samples will be θ .

Ex: Consider sample mean $\hat{\theta} = \sum_{i=1}^n x_{i/n}$. Is that an unbiased estimator?

$$E(\hat{\theta}) = \sum_{i=1}^n E(x_{i/n}) = \sum_{i=1}^n \theta/n = n(\theta/n) = \theta$$

so it is unbiased.

Ex: If $x_1, x_2, \dots, x_n$ constitute a random sample from population given by

$$f(x) = \begin{cases} e^{-(x-\delta)} & \text{for } x > \delta \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\bar{x}$ is biased estimator of δ .

Mean population is

$$\mu = \int_{-\infty}^{\infty} x \cdot e^{-(x-\delta)} dx = 1 + \delta$$

$$E(\bar{x}) = 1 + \delta \neq \delta \Rightarrow \bar{x} \text{ is a biased estimator.}$$

$$E(\bar{x}) - \delta = 1 + \delta - \delta = 1 \Rightarrow \text{amount of bias.}$$

$$E(\bar{x} - 1) = \delta \Rightarrow \bar{x} - 1 \text{ is an unbiased estimator of } \delta.$$

Risk vs. Precision

Both bias and variance must be taken into account when evaluating an estimator.

One way to do this is to consider an estimator's mean square error

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

$(\hat{\theta} - \theta)^2 \Rightarrow$ measures loss from using $\hat{\theta}$ as estimator of θ

Taking mathematical expectation of this loss, yields average loss or risk of using $\hat{\theta}$ instead of θ (to estimate θ).

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

$$= E\{[\hat{\theta} - E(\hat{\theta})] + [E(\hat{\theta}) - \theta]\}^2$$

$$= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 + 2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]$$

$$E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] = [E(\hat{\theta}) - \theta]E[\hat{\theta} - E(\hat{\theta})]$$

$$= [E(\hat{\theta}) - \theta][E(\hat{\theta}) - E(\hat{\theta})] = 0$$

$$MSE(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]^2 = \text{Var}(\hat{\theta}) + [\text{Bias}(\hat{\theta})]^2$$

↘ midterm

Use of MSE → allows us to compare estimators.

$$\hat{\theta} = \sum x_i/n$$

$$\hat{\theta} = \sum x_i/(n-1)$$

✓ A biased estimator may be preferred over unbiased one if yields smaller MSE.

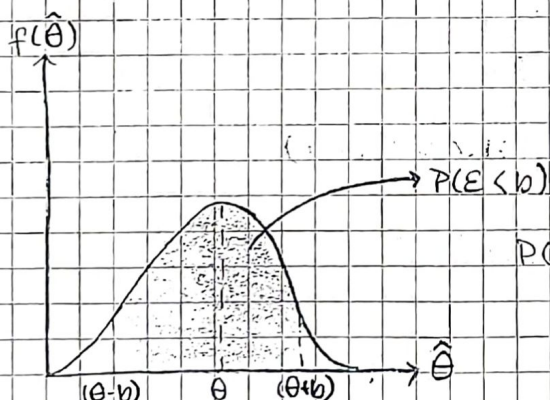
Evaluating Goodness of Point Estimator

Error of estimation, E , is distance between an estimator and its true parameter value.

$$E = |\hat{\theta} - \theta|$$

✳ Since $\hat{\theta}$ is a random variable, E is also a random quantity.

We cannot say how large or small it will be, we can make probability statements about it.



$$P(|\hat{\theta} - \theta| < b) = P[-b < (\hat{\theta} - \theta) < b]$$

$$= P[\theta - b < \hat{\theta} < \theta + b]$$

$b \Rightarrow$ probabilistic bound on error of estimation

$$b = 0.0001, 0.001, 0.1, \dots$$

$P(E < b)$ provides a measure of goodness of a single estimate.

→ identifies fraction of times that estimator $\hat{\theta}$ falls within b units of θ , target parameter θ , in repeated sampling

Suppose we know probability density of $\hat{\theta}$ and want to find value of b so that

$$P(E < b) = 0.95$$

$$\int_{\theta-b}^{\theta+b} f(\hat{\theta}) d\hat{\theta} = 0.95$$

→ efficient olup olmadığını göstermek için önce unbiased olup olmadığa, sonra Cramer-Rao.

However, whether we know probability density of θ or not, if $\hat{\theta}$ is unbiased, we can find an approximate bound on E by expressing b as multiple of statistical error of $\hat{\theta}$.

For instance, for $k \geq 1$, let $b = k\sigma_{\hat{\theta}}$

From Tchebysheff's Theorem, that E will be less than $k\sigma_{\hat{\theta}}$ with probability at least $1 - 1/k^2$.

For $k=2$, E will be less than $b = 2\sigma_{\hat{\theta}}$ with probability at least 0.75

✓ 4 property; unbiasedness, efficiency, consistency, sufficiency.

Efficiency

Another way to take into account both bias and variance is compare estimators that are unbiased. We check whether a given unbiased estimator has the smallest possible variance.

∴ Minimum variance unbiased estimator (MVUE) - best unbiased

An estimator $\hat{\theta}$ is an efficient estimator of θ if

$$E(\hat{\theta}) = \theta \text{ and } \text{var}(\hat{\theta}) \leq \text{var}(\tilde{\theta})$$

↓
necessary condition

where $\tilde{\theta}$ is any other unbiased estimator of θ .

Cramer-Rao Inequality provides a sufficient but not necessary condition for an estimator to be efficient.

↳ sufficient condition

$$\text{var}(\hat{\theta}) \geq \frac{1}{nE\left[\left(\frac{\partial \ln f(x)}{\partial \theta}\right)^2\right]} \rightarrow \text{CRLB (Cramer-Rao Lower Bound)}$$

This result states that if an unbiased estimator $\hat{\theta}$ has a variance equal to

$$\frac{1}{nE\left[\left(\frac{\partial \ln f(x)}{\partial \theta}\right)^2\right]}, \text{ then it is efficient since no lower variance is possible for an unbiased estimator.}$$

→ this condition is sufficient but not necessary since it is possible that no unbiased estimator has variance as small as CRLB.

at least 1 question from sufficiency in midterm.

Ex: Show that $\bar{x}$ is a minimum variance unbiased (efficient) estimator of mean μ of a normal population.

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

taking ln does not change the result because it is a monotonic func.

$$\begin{aligned} \ln f(x) &= -\ln \sigma \sqrt{2\pi} - \frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \\ &= \frac{\partial \ln f(x)}{\partial \mu} = \frac{1}{\sigma} \left(\frac{x-\mu}{\sigma} \right) \end{aligned}$$

$$E \left[\left(\frac{\partial \ln f(x)}{\partial \mu} \right)^2 \right] = \frac{1}{\sigma^2} E \left[\left(\frac{x-\mu}{\sigma} \right)^2 \right] = \frac{1}{\sigma^2} \cdot 1 = \frac{1}{\sigma^2}$$

$$n E \left[\left(\frac{\partial \ln f(x)}{\partial \mu} \right)^2 \right] = n \frac{1}{\sigma^2} = \frac{\sigma^2}{n}$$

Since $\bar{x}$ is unbiased and its variance is $\text{var}(\bar{x}) = \frac{\sigma^2}{n}$, it follows that $\bar{x}$ is minimum variance unbiased (efficient) estimator of μ .

Asymptotic Efficiency

Assume that sample size n is large, $n \rightarrow \infty$

Suppose we have two estimators $\hat{\theta}$ and $\tilde{\theta}$ of a parameter θ such that $\sqrt{n}(\hat{\theta} - \theta) \sim N(0, \sigma_1^2)$ and $\sqrt{n}(\tilde{\theta} - \theta) \sim N(0, \sigma_2^2)$.

If $\sigma_2^2 \geq \sigma_1^2$, then $\hat{\theta}$ is asymptotically efficient relative to $\tilde{\theta}$.

Consistency

It ensures that estimation rule will produce an estimate that is close to true parameter with high probability of sample size large enough.

Let $\hat{\theta}$ be an estimator of θ , based on sample size n . $\hat{\theta}$ is consistent estimator if

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| < c) = 1 \quad \rightarrow c \text{ is an arbitrarily small positive number.}$$

Probability that value of $\hat{\theta}$ falls in interval $[\theta - c, \theta + c]$ can be made arbitrarily close to 1 given large enough sample size, no matter how small c is.

If this is true, then sequence of random variable $\hat{\theta}$ is said to converge in probability to constant θ .

θ is said to be probability limit of sequence $\hat{\theta}$,

$$\text{plim } \hat{\theta} = \theta$$

Thus, estimator is consistent for θ if

$$\text{plim } \hat{\theta} = \theta$$

If $\hat{\theta}$ is unbiased, estimator for parameter θ ,

$$\text{var}(\hat{\theta}) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

then $\hat{\theta}$ is a consistent estimator of θ .

Ex: Show that for a random sample from a normal population, sample variance S^2 is a consistent estimator of σ^2 .

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right] = n-1 \quad E(S^2) = \frac{\sigma^2(n-1)}{n-1} = \sigma^2$$

$$\text{var}\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1) \quad \text{var}(S^2) = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{(n-1)}$$

$$\text{var}(S^2) \rightarrow 0 \text{ when } n \rightarrow \infty$$

S^2 is a consistent estimator of variance of a normal population.

Study probability distribution of an estimator as sample size becomes increasingly large. Estimators whose distribution is unknown in small samples can have particular known distributions in large samples, i.e. CLT

Let $x_1, x_2, \dots, x_n$ be independently, identically distributed (iid) random variables with

$$E(x_i) = \beta; \text{Var}(x_i) = \sigma^2$$

Let $\hat{\beta}_n = \sum_{i=1}^n x_i / n$, then

$\sqrt{n}(\hat{\beta}_n - \beta) \rightarrow$ converges in distribution to $N(0, \sigma^2)$

↓
has a limiting (asymptotic) distribution.

$$\sqrt{n}(\hat{\beta}_n - \beta) \sim N(0, \sigma^2)$$

we can make statements when exact small sample distribution of estimator cannot be determined.

If a limiting distribution exists, we can use this limiting distribution as an approximation to true but unknown distribution of sample if sample is sufficiently large.

Sufficiency

An estimator $\hat{\theta}$ is said to be sufficient if it utilizes all information in a sample relevant to estimation of θ .

$$f(x_1, x_2, \dots, x_n | \hat{\theta}) = f(x_1, x_2, \dots, x_n) / g(\hat{\theta}) \rightarrow \text{from the def. of cond. prob.}$$

The statistic $\hat{\theta}$ is a sufficient estimator of parameter θ if and only if for each value of $\hat{\theta}$, conditional probability distribution or density of random sample $x_1, x_2, \dots, x_n$ given is independent of θ .

Ex: If $x_1, x_2, \dots, x_n$ constitute a random sample of size n from a Bernoulli population, show that

$\hat{\theta} = \frac{x_1 + x_2 + \dots + x_n}{n}$ is a sufficient estimator of parameter θ .
 $\xrightarrow{\quad} n\hat{\theta} = \sum x_i$

$f(x_i, \theta) = \theta^{x_i} (1-\theta)^{1-x_i}$ Bernoulli

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum x_i} (1-\theta)^{n-\sum x_i}$$

$$= \theta^x (1-\theta)^{n-x} = \theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}}$$

Since $X = x_1 + x_2 + \dots + x_n$ is a binomial random variable with parameter θ and n its distribution is

$$b(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$g(\hat{\theta}) = \binom{n}{n\hat{\theta}} \theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}} \text{ for } \hat{\theta} = \theta, \frac{1}{n}, \dots, 1$$

$$f(x_1, x_2, \dots, x_n | \hat{\theta}) = \frac{f(x_1, x_2, \dots, x_n)}{g(\hat{\theta})} = \frac{\theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}}}{\binom{n}{n\hat{\theta}} \theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}}} = \frac{1}{\binom{n}{n\hat{\theta}}}$$

$$= \frac{1}{\binom{n}{x}} = \frac{1}{\binom{n}{x_1 + x_2 + \dots + x_n}}$$

$\rightarrow$ this does not depend on θ and $\hat{\theta} = \frac{x}{n}$ is a sufficient estimator.

Factorization Theorem

Statistic $\hat{\theta}$ is a sufficient estimator of parameter θ if and only if joint probability distribution or density of random sample can be factored so that

$$f(x_1, x_2, \dots, x_n; \theta) = g(\hat{\theta}, \theta) h(x_1, x_2, \dots, x_n)$$

where $g(\hat{\theta}, \theta)$ depends on $\hat{\theta}$ and θ (which includes population information) and $h(x_1, x_2, \dots, x_n)$ does not depend on θ .

Ex: Show that $\bar{x}$ is a sufficient estimator of μ mean of normal population with known variance.

$$f(x_1, x_2, \dots, x_n, \mu) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n e^{-1/2 \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2}$$

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

$$f(x_1, x_2, \dots, x_n, \mu) = \underbrace{\left\{ \frac{1}{\sigma \sqrt{2\pi}} e^{-1/2 \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2} \right\}}_{\text{depends on } \bar{x} \text{ and } \mu} \cdot \underbrace{\left\{ \frac{1}{\sqrt{n}} \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^{n-1} e^{-1/2 \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2} \right\}}_{\text{not involve target parameter } \mu \text{ since variance is known}}$$

depends on $\bar{x}$ and μ BUT not involve target parameter μ since variance is known.

Therefore, $\bar{x}$ is a sufficient estimator of μ .

If $\hat{\theta}$ is a sufficient estimator of θ , then any single-valued function $y = u(\hat{\theta})$, not involving θ is also a sufficient estimator of θ and $u(\theta)$.

Thus, provided $y = u(\hat{\theta})$ can be solved to give single valued inverse $\hat{\theta} = w(y)$ (from factorization theorem and

$$f(x_1, x_2, \dots, x_n, \theta) = g[w(y), \theta] h(x_1, x_2, \dots, x_n)$$

where $g[w(y), \theta]$ depends on y and θ but

$h(x_1, x_2, \dots, x_n)$ only sample information.

Rao-Blackwell Theorem and Minimum Variance Unbiased Estimators (MVUE)

Sufficient statistics play a central role in finding good estimators for parameters. If $\hat{\theta}$ is an unbiased estimator for θ , and if U is a statistics that is sufficient for θ , then there is a function of U that is also an unbiased estimator for θ and has no larger variance greater than that of $\hat{\theta}$.

→ If we seek unbiased estimators with small variances, we can restrict our search to estimators that are functions of sufficient statistics.

✓ If we can obtain sufficiency, then it is also efficient and unbiased.

Rao-Blackwell Theorem

Let $\hat{\theta}$ be an unbiased estimator of θ such that $\text{Var}(\hat{\theta}) < \infty$.

If U is a sufficient statistics for θ , define $\hat{\theta}^* = E(\hat{\theta} | U)$. Then, for all θ ,

$$E(\hat{\theta}^*) = \theta, \quad \text{since it is sufficient, it is unbiased.} \quad (E(\hat{\theta}) = \theta)$$

$$\text{Var}(\hat{\theta}^*) \leq \text{Var}(\hat{\theta})$$

Rao-Blackwell Theorem implies that an unbiased estimator for θ with a small variance is or can be made to be a function of a sufficient statistics.

If we start with an unbiased estimator for a parameter θ , and the sufficient statistics obtained through factorization criterion, application of Rao-Blackwell Theorem leads to a MVUE for parameter.

✓ Rao-Blackwell Theorem is build up to Factorization Theorem, :

If U is a sufficient statistics, then that best summarizes nature of sufficiency (uses all samples) data,

and some function of U such as $h(U)$ can be found such that $E[h(U)] = \theta$.

It follows that $h(U)$ is MVUE for θ .

Ex: Consider a Bernoulli random variable with unknown p . Use factorization criterion to find a sufficient statistics that best summarize the data. Give a MVUE for p .

$$P(Y_i = y_i) = p^{y_i} (1-p)^{1-y_i} \quad y_i = 0, 1 \quad \text{Bernoulli}$$

$$\begin{aligned} L(p) &= L(y_1, y_2, \dots, y_n | p) = P(y_1, y_2, \dots, y_n | p) = \\ &= p^{y_1} (1-p)^{1-y_1} \cdot p^{y_2} (1-p)^{1-y_2} \cdot \dots \cdot p^{y_n} (1-p)^{1-y_n} \\ &= \underbrace{p^{\sum y_i} (1-p)^{n - \sum y_i}}_{g(\sum y_i, p)} \cdot 1 \rightarrow h(y_1, y_2, \dots, y_n) \end{aligned}$$

According to factorization theorem,

$$U = \sum_{i=1}^n Y_i \text{ is a sufficient statistics}$$

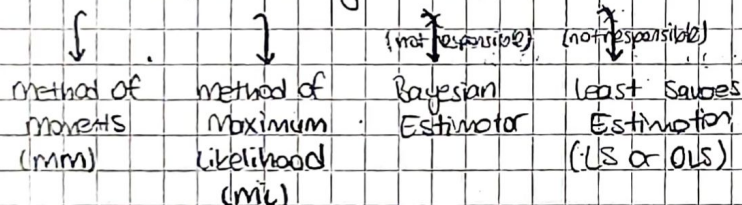
This statistics best summarizes data or information on parameter p .

$$E(U) = n \cdot p \quad \text{or} \quad E(U/n) = p \rightarrow U/n = \bar{Y} \text{ is an unbiased estimator for } p.$$

Since this estimator is a function of sufficient statistics,

$$\sum_{i=1}^n Y_i \text{ estimator } \hat{p} = \bar{Y} \text{ is MVUE for } p.$$

Methods for Finding Point Estimators



Estimation and Estimator

Let Y be a random variable with pdf $f(y|\theta)$ where θ is a parameter we want to estimate.

Let $y_1, y_2, \dots, y_n$ be a random sample from this population. Then, an estimator of θ is a function or rule of form $\hat{\theta} = \hat{\theta}(y_1, \dots, y_n) \Rightarrow \hat{\theta}$; estimator is a random variable since it is a function of random variables.

estimate $\rightarrow$ number, or a function
 $\rightarrow$ not same as estimation

estimator $\rightarrow$ a rule
 $(\hat{\theta} = \sum y_i/n)$

When values of random variable are inserted in this equation, estimate is obtained. $\rightarrow$ it is not random.

Method of moments

r^{th} moment of a random variable Y about origin is

$$\mu_r' = E(Y^r)$$

If pdf of Y is $f(y|\theta')$, where $\theta' = (\theta_1, \dots, \theta_k)$ is a vector of unknown parameters.

In general, μ_r' will be a function of θ , $\mu_r' = \mu_r'(\theta)$.

Idea of MM is to use a random sample of data $y_1, \dots, y_n$ to compute sample moments.

$$\hat{\mu}_r' = \sum_{i=1}^n y_i^r/n, \quad r=1, \dots, k$$

! Method of moment estimator, but, some difficulty but. (midterm)

then to equate sample and true moments

$$\hat{\mu}_r' = \mu_r'(\theta)$$

and solve resulting system of k equations for unknown parameters if possible.

Resulting estimator, $\hat{\theta}$ is method of moments estimator.

Ex: Let $y_1, \dots, y_n$ be a random sample from a $N(\theta, \sigma^2)$ population. Find MM estimators of mean and variance.

$$E(Y) = \theta = \mu_1', \quad \sigma^2 = E(Y^2) - [E(Y)]^2 = \mu_2' - (\mu_1')^2$$

Equating sample to population moments,

$$\left. \begin{aligned} \hat{\mu}_1' &= \sum_{i=1}^n y_i/n = \bar{y} = \hat{\theta} \\ \hat{\mu}_2' &= \sum_{i=1}^n y_i^2/n = \hat{\sigma}^2 + \hat{\theta}^2 \end{aligned} \right\} \text{ so, } \hat{\sigma}^2 = \sum_{i=1}^n y_i^2/n - \bar{y}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

There are difficulties of method of moments,

- ① These estimators need not to be unique.
- ② Method can be difficult to apply in complicated problems.
- ③ It also depends on random variables having moments, which is not always the case.

Method of Maximum Likelihood (ml or mle)

Let say Y is a Bernoulli random variable with parameter

$$P \begin{cases} \rightarrow 1/4 \\ \rightarrow 3/4 \end{cases}$$

Suppose, we have random sample of $n=3$, $y_1=1$, $y_2=1$, $y_3=0$

How do we use data to estimate p , $\hat{p}$?

Method of ml choose that value of unknown parameter p , that maximizes probability (likelihood) of randomly drawing sample that was actually obtained.

Since $f(y|p) = p^y(1-p)^{1-y}$ for $y=0$ or $y=1$, calculate probability of our random sample from joint pdf of y_1, y_2, y_3 .

$$f(y_1=1, y_2=1, y_3=0) = f(1, 1, 0) = \prod_{i=1}^3 p^{y_i} (1-p)^{1-y_i} = p \cdot p \cdot (1-p) = p^2(1-p)$$

→ It is a function of p given sample observations.

↓
It is called likelihood function → $l(p|y)$

Likelihood function is mathematically identical to joint pdf of random sample.

→ It is function of unknown parameters instead of random variables. All sample information is contained in likelihood function.

$$l(1/4|y) = 0.0625 \text{ and } l(3/4|y) = 0.1406$$

Likelihood of obtaining sample who actually have max. by choosing.

$$\boxed{\hat{p} = 3/4} \text{ rather than } 1/4.$$

p can take any value in parameter space,

$$\Omega = \{p | 0 \leq p \leq 1\}$$

$\max l(p|y) = p^2(1-p)$ by numerical techniques.

$\max l(p|y)$ by using calculus techniques.

A local max. of continuous function occurs where its slope is zero and 2nd derivative is negative.

$$\frac{\partial \ell(p|y)}{\partial p} = 2p - 3p^2$$

$$\frac{\partial^2 \ell(p|y)}{\partial p^2} = 2 - 6p$$

$$2 - 4 = -2$$

$$-4 < 0$$

$$-2 < 0$$

we cannot choose sample.

$$2p - 3p^2 = 0 \Rightarrow \hat{p}$$

$$0$$

$$2/3$$

plug in.

$\hat{p}=0$ is ruled out it does not max. likelihood

$$\boxed{\hat{p} = 2/3}$$

To sum up,

① likelihood function $\ell(\theta|y)$ is algebraically identical to joint pdf $f(y|\theta)$ of random sample $y_1, \dots, y_n$ where θ is unknown vector of parameters with a positive value in Ω .

② Difference between $\ell(\theta|y)$ and $f(y|\theta)$ is that the former is a function of θ given values of random sample. the latter is a function of y given values of θ .

③ ML estimate of θ is that value in Ω , $\hat{\theta}$, that max. $\ell(\theta|y)$

④ Value of $\hat{\theta}$ is in general a function of y , $\hat{\theta} = \hat{\theta}(y)$ so that random variable $\hat{\theta} = \hat{\theta}(Y)$ where $Y = Y_1, \dots, Y_n$ is ML estimator of θ .

⑤ In practice, to max. natural logarithm of likelihood function is used, since log-likelihood $\lambda = \ln \ell(\theta|y)$ reaches its maximum like $\ell(\theta|y)$.

→ log-likelihood function is a monotonic function.

Ex: let $y_1, y_2, \dots, y_n$ be a random sample of Bernoulli random variables with

$$f(y_i|p) = p^{y_i} (1-p)^{1-y_i}, \quad i=1, \dots, n, \quad 0 \leq p \leq 1$$

$$\begin{aligned} \ell = \ln \ell(p|y) &= \ln \left(\prod_{i=1}^n f(y_i|p) \right) = \ln [p^{\sum y_i} (1-p)^{\sum (1-y_i)}] \\ &= \sum y_i \ln(p) + \sum (1-y_i) \ln(1-p) \end{aligned}$$

$$\frac{\partial \ell}{\partial p} = \sum y_i \frac{1}{p} - \sum (1-y_i) \frac{1}{1-p} = 0 \Rightarrow \sum y_i \frac{1}{p} - \sum (1-y_i) \frac{1}{1-p} = 0$$

$$\frac{\partial^2 \ell}{\partial p^2} = -\sum y_i \frac{1}{p^2} - \sum (1-y_i) \frac{1}{(1-p)^2}$$

$$\hat{p} = \frac{\sum y_i}{n}$$

sample mean

$$\frac{\partial^2 \ell}{\partial p^2} < 0 \text{ for } 0 < p < 1.$$

25.03.19
 Ex: Let $y_1, \dots, y_n$ be a random sample from a $N(\beta, \sigma^2)$.
 Find μ estimators of mean and variance. $\hat{\beta}, \hat{\sigma}^2$?

$$f = \ln(\theta(y)) = \ln \left\{ \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2 \frac{(y_i - \beta)^2}{\sigma^2}} \right\}$$

$$= \ln \left\{ (2\pi)^{-n/2} (\sigma^2)^{-n/2} e^{-1/2 \sum_{i=1}^n \frac{(y_i - \beta)^2}{\sigma^2}} \right\}$$

$$f = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \beta)^2}{\sigma^2}$$

$$\frac{\partial f}{\partial \beta} = \frac{1}{\sigma^2} \left(\sum_{i=1}^n y_i - n\beta \right)$$

$$\frac{\partial f}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \beta)^2 = 0$$

$$\frac{\partial^2 f}{\partial \beta^2} = \frac{n}{\sigma^2} < 0$$

$$\frac{\partial^2 f}{(\partial \sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \beta)^2 < 0$$

$$\frac{\partial^2 f}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4} \left(\sum_{i=1}^n y_i - n\beta \right)$$

$$\frac{1}{\sigma^2} \left(\sum_{i=1}^n y_i - n\hat{\beta} \right) = 0$$

$$\frac{-n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (y_i - \hat{\beta})^2 = 0$$

$$\hat{\beta} = \sum_{i=1}^n \frac{y_i}{n} = \bar{y}$$

$$\frac{n}{2\hat{\sigma}^2} = \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (y_i - \hat{\beta})^2$$

$$\hat{\sigma}^2 = \frac{(\sum y_i - n\hat{\beta})^2}{n}$$

Matrix of 2nd derivatives should be negative definite

$$\begin{bmatrix} \frac{\partial^2 f}{\partial \beta^2} & \frac{\partial^2 f}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 f}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 f}{(\partial \sigma^2)^2} \end{bmatrix} = \begin{bmatrix} \frac{-n}{\hat{\sigma}^2} & \sim \\ \sim & \frac{-n}{2\hat{\sigma}^4} \end{bmatrix}$$

A symmetric matrix, A , is negative definite if and only if $x^T A x < 0$ for all $x \neq 0$.

Bayesian Estimation

In classical approach, parameter θ is thought to be unknown but fixed quantity.

A random sample $y_1, \dots, y_n$ is drawn from a population indexed by θ and based on observed values in sample.
 $\rightarrow$ knowledge on θ is observed.

In Bayesian approach, θ is considered to be quantity whose variation can be described by probability distribution called prior distribution.

$\rightarrow$ subjective distribution based on experimenter's belief

formulated before data is seen.

Then, a sample is taken from a population indexed by θ and prior distribution is updated with this sample information.

$\therefore$ updated prior $\Rightarrow$ is called posterior distribution.

This updating is done by use of Baye's Theorem.

$\rightarrow$ If $B_1, B_2, \dots, B_k$ constitute a portion of sample spaces and $P(B_i) \neq 0$ for $i=1, 2, \dots, k$, then

$$P(B_i | A) = \frac{P(B_i) \cdot P(A | B_i)}{\sum_{i=1}^k P(B_i) \cdot P(A | B_i)}$$

Prior distribution by $\pi(\theta)$ and sampling distribution by $f(y|\theta)$.

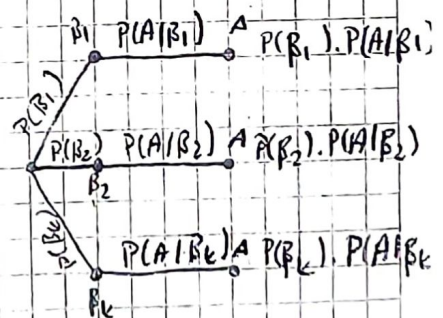
Posterior distribution is cond. dist. of θ given sample y .

$$\pi(\theta | y) = \frac{f(y|\theta) \cdot \pi(\theta)}{g(y)} \rightarrow f(y|\theta) \cdot \pi(\theta) = f(y, \theta)$$

marginal dist. of y

$$\rightarrow f(y) = \int f(y|\theta) \cdot \pi(\theta) \cdot d\theta$$

conditional upon observing sample



Interval Estimation

An alternative way to point estimation is interval estimation. An interval estimate of θ is an interval from $\theta_1 < \theta < \theta_2$.

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha \rightarrow \text{degree of confidence}$$

some special probability
lower and upper confidence limits

Estimation of Means

A mean of random sample is used to estimate mean of normal population with known variance.

Sampling distribution of $\bar{X}$ for random sample of size n from a normal population with mean μ and σ^2 has a normal distribution.

$$\mu_{\bar{X}} = \mu \text{ and } \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

$$P(|z| < z_{\alpha/2}) = 1 - \alpha, \quad z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$z_{\alpha/2}, \quad P(|\bar{X} - \mu| < z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$$P(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Ex. If a random sample of size $n=20$ from a normal population with $\sigma^2=225$ has a mean $\bar{X}=64.3$ construct 95% of confidence interval for population mean μ

$$P(64.3 - 1.96 \cdot \frac{15}{\sqrt{20}} < \mu < 64.3 + 1.96 \cdot \frac{15}{\sqrt{20}}) = 0.95$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$P(57.7 < \mu < 70.3) = 0.95$$

$$z_{0.025} = 1.96$$

$$P(|z| < z_{\alpha/2}) = 1 - \alpha$$

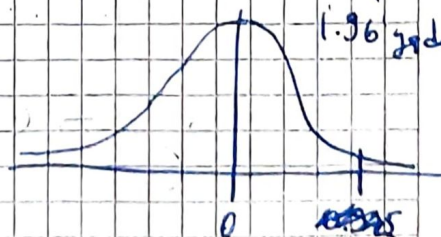
$$P(|\bar{X} - \mu| < z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

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

$$1 - 0.025 = 0.975$$

look up table

1.96 z value



By help of Central Limit Theorem, we can construct confidence intervals for non-normal populations if n is sufficiently large. ($n \geq 30$)

Use sample standard deviation as approximate to population standard deviation.

$$P\left(\bar{X} - z_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Suppose σ is unknown and $n < 30$,

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1} \quad P(-t_{\alpha/2, n-1} < T < t_{\alpha/2, n-1}) = 1 - \alpha$$

$$P\left(\bar{X} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Ex: A random sample of six cars from a particular model year had mean $\bar{x} = 19.4833$ and $s^2 = 0.98$ fuel cons. find 90% confidence interval for population mean fuel cons. for cars of this model year assuming population distribution is normal.

$$P\left(19.4833 - 2.015 \frac{(0.98)}{\sqrt{6}} < \mu < 19.4833 + 2.015 \frac{(0.98)}{\sqrt{6}}\right) = 0.90$$

$$P(18.67 < \mu < 20.18) = 0.90$$

Confidence Intervals for Differences between Means

For independent random samples from normal population,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

σ_1 and σ_2 should be close or equal to each other.

$$P\left[(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] = 1 - \alpha$$

④ By Central Limit Theorem, $n_1 \geq 30$ and $n_2 \geq 30$, we can use same interval for non-normal populations.

To construct confidence interval for differences between means ($n_1 \geq 30, n_2 \geq 30$) but σ_1^2 and σ_2^2 are unknown.

$$P\left[(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right] = 1 - \alpha$$

⑤ σ_1^2 and σ_2^2 are unknown, small samples.

Pooled estimator for variance
$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Ex: In a study of R&D activities on profit of banks random sample of 6 banks having a higher level of R&D expenditures showed mean annual percentage increase in profit as 9.9+2 and standard deviation of sample 7.470 lower higher level of R&D, $\bar{x}_2 = 2.098$ and $\sigma^2 = 10.834$

Assuming that 2 population distributions are normal with same variance. Find a 90% confidence interval.

• They are independent, normal distributed and $\sigma_1^2 = \sigma_2^2$

$$\sigma^2 p = \frac{(6-1)(7.470)^2 + (9-1)(10.834)^2}{6+9-2} = 93.693$$

$$s_p = \sqrt{93.693} = 9.680$$

$$\alpha = 0.10, \alpha/2 = 0.05$$

$$t_{0.05, 13} = 1.771$$

$$P[(9.972 - 2.098) - (1.771) \cdot (9.680) \sqrt{\frac{1}{6} + \frac{1}{9}} < \mu_1 - \mu_2 < (9.972 - 2.098) + (1.771) \cdot (9.680) \sqrt{\frac{1}{6} + \frac{1}{9}}] = 0.90$$

$$P[-1.61 < \mu_1 - \mu_2 < 16.809] = 0.90$$

negative profits

higher profits

variance decrease, efficiency in production.

static confidence interval, not dynamic

Estimation of Variances

Given a random sample of size n from a normal population, we can obtain $100(1-\alpha)\%$ confidence interval for σ^2 .

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

$$P\left[\chi^2_{1-\alpha/2, n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2, n-1}\right] = 1-\alpha$$

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}\right] = 1-\alpha$$

? Ex: In a university $S_2 = 120$ of an exam for a random sample of size 101 students. From a normal population construct 95% confidence interval for population variance, σ^2 .

$$S_2 = 120$$

$$\alpha = 0.05$$

$$\chi^2_{0.025, 100} = 129.56$$

$$n = 101$$

$$\alpha/2 = 0.025$$

$$\chi^2_{0.975, 100} = 74.22$$

$$P\left[\frac{(101-1) \cdot 120}{129.56} < \sigma^2 < \frac{100 \cdot 100}{74.22}\right] = 0.95 \Rightarrow P[92.62 < \sigma^2 < 161.68] = 0.95$$

Estimation of Ratio of Two Variances

If S_1^2 and S_2^2 are variances of independent random samples of size n_1 and n_2 from "normal populations"

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F_{n_1-1, n_2-1}$$

$$f_{1-\alpha/2, n_1-1, n_2-1} = \frac{1}{f_{\alpha/2, n_1-1, n_2-1}}$$

$$P[f_{1-\alpha/2, n_1-1, n_2-1} < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2, n_1-1, n_2-1}] = 1 - \alpha$$

Ex: According to results of a study to compare variability of productivity of two car factory, by using 10 workers in factory 1, standard deviation random sample 0.5, for factor 2 standard deviation 0.7 for sample of 8 workers.

a) Find a 98% confidence interval for σ_1^2 / σ_2^2 .

b) State your assumptions in constructing this interval.

$$\alpha = 0.02, \alpha_{1,2} = 0.01$$

$$n_1 = 10$$

$$n_2 = 8$$

$$s_1 = 0.5$$

$$s_2 = 0.7$$

$$s_1^2 = 0.25$$

$$s_2^2 = 0.49$$

$$f_{0.001, 9, 7} = 6.72$$

$$P\left[\frac{0.25}{0.49} \cdot \frac{1}{6.72} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{0.25}{0.49} \cdot 5.61\right] = 0.98$$

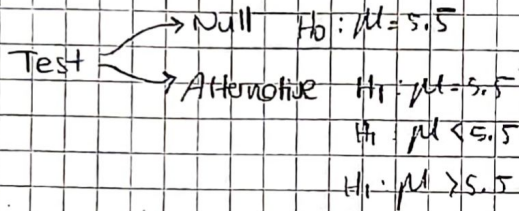
$$f_{0.001, 7, 9} = 5.61$$

$$P[0.076 < \frac{\sigma_1^2}{\sigma_2^2} < 2.862] = 0.98$$

Hypothesis Testing

Not say I know the truth rather than I know a truth...

A statistical hypothesis is a statement about numerical value of population parameter.



Test procedure partitions possible values of test statistics into two subsets:

- non-rejection region for H_0
- rejection region for H_0 (critical)

* Rejection of H_0 when it is true → Type I error
 probability of committing type I error is α .

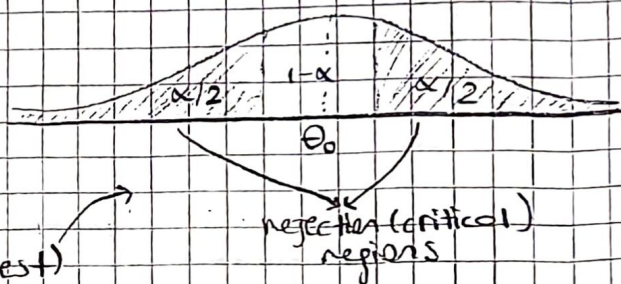
* Non-rejection of H_0 when it is false → Type II error
 probability of committing type II error is β .

Size of critical (rejection) region is probability α of committing type I error. → level of significance

Not Reject	H_0 True correct decision prob. = $1 - \alpha$	H_0 False Type II error prob. = β
Reject	Type I error prob. = α	correct decision prob. = $1 - \beta$ power of test

To perform a hypothesis testing:

- i. state hypothesis
- ii. state decision rule $\rightarrow$ test statistics
- iii. compute value of test statistics
- iv. decide $\rightarrow$ reject H_0
 $\rightarrow$ not reject H_0
- v. interpret your results



$$H_0: \theta = \theta_0$$

$$H_1 \rightarrow H_1: \theta \neq \theta_0 \text{ (two-tailed test)}$$

$$H_1 \rightarrow H_1: \theta < \theta_0 \text{ or } \theta > \theta_0 \rightarrow \text{one-tailed test}$$



Test of Means $n \rightarrow \geq 30$
 $\rightarrow < 30$

$\sigma^2 \rightarrow$ known
 $\rightarrow$ unknown

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0 \text{ or } \mu > \mu_0 \text{ or } \mu < \mu_0$$

by using random sample of size n from a normal population with known variance, σ^2

$$\bar{z} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

H_0 is rejected if

$$\begin{cases} |\bar{z}| > z_{\alpha/2} & \text{for } H_1: \mu \neq \mu_0 \\ \bar{z} \geq z_{\alpha} & \text{for } H_1: \mu > \mu_0 \\ \bar{z} \leq -z_{\alpha} & \text{for } H_1: \mu < \mu_0 \end{cases}$$

Simple Linear Regression Model (Ordinary Least Squares Estimation)

29.06.19
ECON206

OLS

①

Basic Assumptions of Model

$$C_t = \beta_0 + \beta_1 Y_t^d + u_t$$

consumption model

$$\Theta_t = \beta_0 + \beta_1 K_t + \beta_2 L_t + u_t$$

$$Y_t = \beta_0 + \beta_1 X_t + u_t \quad t=1, \dots, T$$

β_0 and $\beta_1 \Rightarrow$ regression parameters, coefficients

Y_t and $X_t \Rightarrow$ are observable variables.

dependent variable

independent variable

$u_t \Rightarrow$ residual term, not observable

① Expected or mean value of u_t is zero.

$$E(u_t) = 0$$

This assumption is violated in various ways.

i. left out variable: Assume you left out an important variable, z_t .

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 z_t + u_t, \quad t=1, \dots, T \Rightarrow \text{correct model} \quad E(u_t) = 0$$

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \quad t=1, \dots, T \Rightarrow \text{incorrect model}$$

$$u_t = \beta_2 z_t + v_t \quad E(u_t) = E(\beta_2 z_t) + E(v_t) = \beta_2 E(z_t) + E(v_t)$$

$$E(u_t) = \beta_2 E(z_t) + 0$$

mesela, üretim fonksiyonuna laboratu koymazsak doğru olan ve regression modeli doğru olur.

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \quad t=1, \dots, T$$

intercept term $\rightarrow$ slope term

Y instead of Y^d , for ex.

ii. Usage of incorrect variable, instead of correct variable w_t , x_t is used incorrectly.

$$Y_t = \alpha_0 + \alpha_1 w_t + v_t, \quad t=1, \dots, T \Rightarrow \text{correct model}$$

$$Y_t = \beta_0 + \beta_1 x_t + u_t, \quad t=1, \dots, T \Rightarrow \text{incorrect model}$$

$$C_t = \beta_0 + \beta_1 \underbrace{Y_t^d}_{Y_t} + u_t$$

$$E(u_t) = (\beta_0 - \alpha_0) + (\beta_1 x_t - \alpha_1 w_t) \neq 0$$

income disposable income autonomous consumption

$$(Y_t) = \beta_0 + \beta_1(X_t) + u_t \rightarrow C_t = 5 + 0.7X_t$$

(think as consumption function)

	Y_t	X_t
1980	80	100
	40	110
	50	120
2017		

at 1980

$$C_t = 5 + 0.7(100)$$

$$Y_{1980} = 80$$

$$\hat{Y}_{1980} = 75$$

$\hat{u}_{1980} = 45$ (note term)

$$u_t = Y_t - \beta_0 - \beta_1 X_t$$

random variable

model in term model

$$\min. \sum \hat{u}_t^2$$

homoscedasticity

② Variance of u_t is a constant and equal to σ^2 and does not vary systematically with t .

$$E[u_t - E(u_t)]^2 = E(u_t)^2 = \sigma^2, \forall t$$

If variance of u_t was not same for all observations, all observations would not equally be reliable. heteroscedasticity → if it is violated.

③ Value of t^{th} disturbance is independent of any other disturbance term, like s^{th} .

$$\text{cov}(u_t, u_s) = E[u_t - E(u_t)][u_s - E(u_s)] = E[u_t, u_s] = 0, \text{ for } t \neq s.$$

If this assumption is violated, then u_t together with X_t becomes a member of systematic part.

$$Y_t^* = \beta_0 + \beta_1 X_t + \beta_2 u_t - 1^{tu_t}$$

④ Autocorrelation

X_t → fixed, nonstochastic
 X_t → a random variable

⑤ If we assume x_t is a random variable, we need to assume that u_t are independent of all T values in regressor X .

$$\text{cov}(u_t, x_s) = E(u_t x_s) - E(u_t) E(x_s) = E(u_t x_s) = 0$$

⑥ Intercept term and x_t must be linearly interrelated.

$$\begin{bmatrix} x_t \\ 5 \\ 10 \\ \vdots \end{bmatrix} \quad \begin{bmatrix} u_t \\ 10 \\ 20 \\ \vdots \end{bmatrix}$$

multicollinearity
VIF

there is a linear relationship between this two.

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t + u_t$$

Estimation Problem

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad t=1, \dots, T$$

Since y_t is a random variable information about its behavior is contained in its probability distribution.

Since y_t and u_t is linearly related $\Rightarrow$ mean, variance, covariance of y_t is related with distribution of u_t .

$$\begin{aligned} E(y_t) &= E[\beta_0 + \beta_1 x_t + u_t] = E(\beta_0) + E(\beta_1 x_t) + E(u_t) \\ &= \beta_0 + \beta_1 E(x_t) + E(u_t) \end{aligned}$$

since by assumption 1, $E(u_t) = 0$ $E(x_t) = x_t$

$$E(y_t) = \beta_0 + \beta_1 x_t$$

$$\text{var}(y_t) = E[y_t - E(y_t)]^2$$

$$\text{var}(y_t) = E(y_t - \beta_0 - \beta_1 x_t)^2 = E(u_t)^2 = \sigma^2 \text{ by assumption 2}$$

$$\begin{aligned}\text{cov}(Y_t, Y_s) &= E\{[Y_t - E(Y_t)][Y_s - E(Y_s)]\} \\ &= E\left[\underbrace{Y_t - \beta_0 - \beta_1 X_t}_{u_t} \underbrace{[Y_s - \beta_0 - \beta_1 X_s]}_{u_s}\right] \\ &= E(u_t u_s) = 0 \text{ by assumption 3.}\end{aligned}$$

(4)

Problem of estimating parameters of model (β_0 and β_1)

may be viewed as estimating parameters of probability distribution of dependent variable Y_t .

Since $E(Y_t) = \beta_0 + \beta_1 X_t$ } if we estimate β_0 and β_1 ,
we estimate mean of Y_t .

Since $\text{var}(Y_t) = E(u_t)^2 = \sigma^2$ } estimating variance Y_t is
equal to estimating variance u_t , σ^2 .

Suppose mean of Y is estimated as

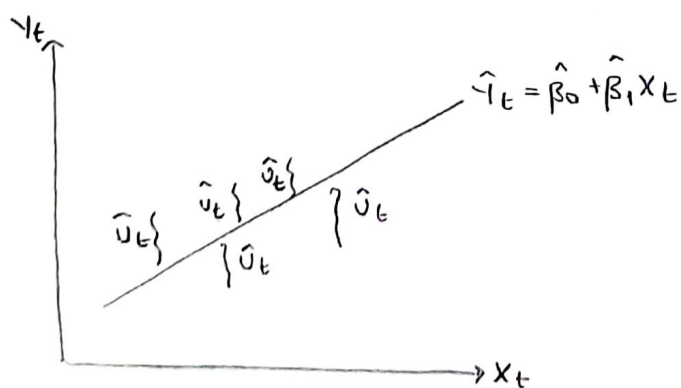
$$\text{fitted values} \leftarrow \hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$$

↘ OLS estimators of β_0 and β_1 .

$\hat{\beta}_0$ and $\hat{\beta}_1$ are random variables, $\hat{Y}_t$ is also a random variable.

There will be a discrepancy between actual values of Y_t and fitted values of $\hat{Y}_t$. These discrepancies are called residuals, $\hat{u}_t$.

$$\hat{u}_t = Y_t - \hat{Y}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t$$



Method of LS proposes to find values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that will make sum of squares of these residuals. (these distances) from line $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$ as small as possible.

$$C_t = \beta_0 + \beta_1 Y_t^d + u_t, \quad t = 1980, \dots, 2017$$

$$\begin{bmatrix} C_t \\ 40 \\ 50 \\ 60 \\ \vdots \end{bmatrix} \quad \begin{bmatrix} Y_t^d \\ 100 \\ 110 \\ 120 \\ \vdots \end{bmatrix} \quad \begin{bmatrix} \hat{C}_t \\ 75 \\ 82.7 \\ 89 \\ \vdots \end{bmatrix} \quad \begin{bmatrix} \hat{u}_t \\ -35 \\ -32.7 \\ -29 \\ \vdots \end{bmatrix}$$

$\hat{C}_t = 5 + 0.7 Y_t^d$

$\min \sum \hat{u}_t^2$

Sum of squared residuals (SSR, S)

$$\min S = \sum_{t=1}^T \hat{u}_t^2 = \sum_{t=1}^T (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t)^2 \quad \text{with respect to } \hat{\beta}_0 \text{ and } \hat{\beta}_1.$$

$$1a. \quad \partial S / \partial \hat{\beta}_0 = -2 \sum_{t=1}^T (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t) = 0 \quad \left. \vphantom{\sum_{t=1}^T} \right\} \text{normal equations}$$

$$1b. \quad \partial S / \partial \hat{\beta}_1 = -2 \sum_{t=1}^T X_t (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t) = 0$$

$$\sum_{t=1}^T Y_t - \sum_{t=1}^T \hat{\beta}_0 - \sum_{t=1}^T \hat{\beta}_1 X_t = 0$$

$$2a. \quad \sum_{t=1}^T -\hat{\beta}_0 T - \hat{\beta}_1 \sum_{t=1}^T X_t = 0$$

$$2b. \quad -\hat{\beta}_0 \sum_{t=1}^T X_t - \hat{\beta}_1 \sum_{t=1}^T X_t^2 + \sum_{t=1}^T X_t Y_t = 0.$$

From 2a.

$$\hat{\beta}_0 T = \sum_{t=1}^T Y_t - \sum_{t=1}^T \hat{\beta}_1 X_t$$

$$3. \quad \boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}}$$

LS estimator of $\hat{\beta}_0$.

Substitute (3) into (2b)

$$\sum_t X_t Y_t - (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum_t X_t - \hat{\beta}_1 \sum_t X_t^2 = 0$$

$$\sum_t X_t Y_t - \bar{Y} \sum_t X_t - \hat{\beta}_1 \bar{X} \sum_t X_t - \hat{\beta}_1 \sum_t X_t^2 = 0$$

$$\sum_t X_t Y_t - \bar{Y} \sum_t X_t - \hat{\beta}_1 (\bar{X} \sum_t X_t - \sum_t X_t^2) = 0$$

$$\hat{\beta}_1 = (\sum_t X_t Y_t - T \bar{X} \bar{Y}) / (\sum_t X_t^2 - T \bar{X}^2)$$

LS estimator of $\hat{\beta}_1$ → raw data form

From 1a and 1b

$$\sum_t \hat{u}_t = 0, \quad \sum_t x_t \hat{u}_t = 0$$

$$E(u_t) = 0$$

$$\hat{\beta}_1 = \sum (x_t - \bar{x})(y_t - \bar{y}) / \sum (x_t - \bar{x})^2$$

$$x_t = x_t - \bar{x}, \quad y_t = y_t - \bar{y}$$

$$\boxed{\hat{\beta}_1 = \sum x_t y_t / \sum x_t^2}$$

LS estimator of $\hat{\beta}_1$ in non-deviation form.

$$\sum_t \hat{y}_t \hat{u}_t - \hat{\beta}_0 \sum_t \hat{u}_t + \hat{\beta}_1 \sum_t x_t \hat{u}_t = 0$$

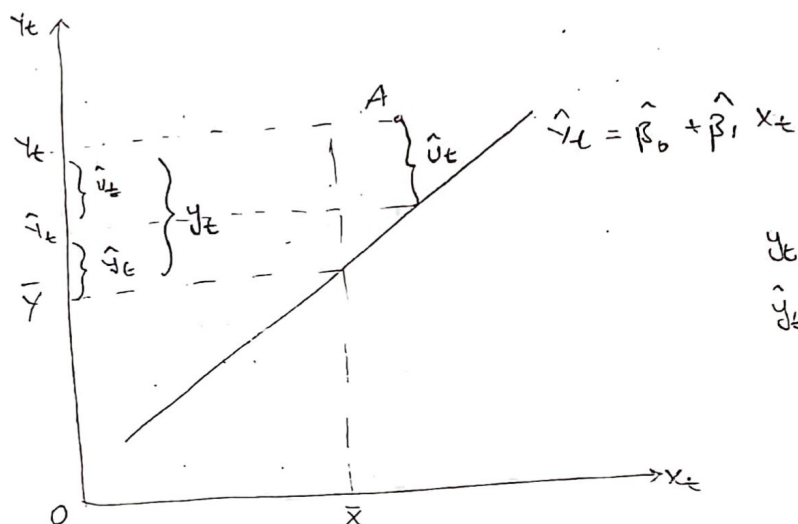
$$\sum_t \hat{y}_t^2 = \sum_t \hat{y}_t \underbrace{(y_t - \hat{u}_t)}_{y_t} = \sum_t \hat{y}_t y_t - \sum_t \hat{y}_t u_t$$

$$\sum_t \hat{y}_t^2 = \sum_t \hat{y}_t y_t$$

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \quad t=1, \dots, T \implies \hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$$

Goodness of Fit

How good is the fit of $\hat{Y}_t$ to scatter of observations on Y_t and X_t ?



$$y_t = Y_t - \bar{Y}$$

$$\hat{y}_t = \hat{Y}_t - \bar{Y}$$

we may define three different distances.

i. Distance from Y_t to origin

$$Y_t - 0 = Y_t$$

ii. Distance of Y_t from its sample mean

$$Y_t - \bar{Y} = y_t$$

iii. Distance of Y_t from fitted regression line

$$Y_t - \hat{Y}_t = \hat{u}_t$$

To form a measure of goodness of fit, variation of Y_t around its sample mean.

$$\sum_t (Y_t - \bar{Y})^2 = \sum_t y_t^2$$

part explained by $\hat{Y}_t$, ($\hat{Y}_t - \bar{Y} = \hat{y}_t$)

part explained by $\hat{u}_t$

$$\begin{aligned}\sum_t y_t^2 &= \sum_t (\hat{y}_t + \hat{u}_t)^2 \\ &= \sum_t \hat{y}_t^2 + \sum_t \hat{u}_t^2 + 2 \underbrace{\sum_t \hat{y}_t \hat{u}_t}_0\end{aligned}$$

→ if it is not zero, system became biased. they will not depend each other, there should be a linear relationship.

$$= \sum_t y_t^2 + \sum_t \hat{u}_t^2 \longrightarrow \text{let's divide both sides by } \sum_t y_t^2.$$

$$R^2 = \frac{\sum_t \hat{y}_t^2}{\sum_t y_t^2} = 1 - \left(\frac{\sum_t \hat{u}_t^2}{\sum_t y_t^2} \right)$$

we wish to make it small as possible.
(note termi kucuk olmak.)

coefficient of determination

$$0 \leq R^2 \leq 1$$

* If $R^2=0$, then observation points are scattered around regression line in a purely random manner.

(not possible, there is always a small relationship between two variable.)

* If $R^2=1$, then all observations are on regression line.
(theoretically possible but not possible in reality.)

both theoretically and statistically.

Ex: $\hat{C}_t = 5 + 0.7 Y_t^d$, $R^2 = 0.89$

$\hat{C}_t = 6 + 0.6 Y_t$, $R^2 = 0.75$

* Low values of R^2 , poor fit of regression.

* High values of R^2 , good fit of regression.

$$R^2 = \frac{[\sum x_t y_t - (\sum x_t)(\sum y_t)]^2}{[\sum x_t^2 - (\sum x_t)^2][\sum y_t^2 - (\sum y_t)^2]}$$

$$R^2 = \left(\sum_t \hat{y}_t^2 / \sum_t y_t^2 \right) \underbrace{\left(\sum_t \hat{y}_t^2 / \sum_t \hat{y}_t^2 \right)}_1$$

$$= (\sum_t y_t \hat{y}_t) / (\sum_t \hat{y}_t^2) (\sum_t y_t^2)$$

$$= \left[\sum_t (\hat{y}_t - \bar{\hat{y}})(y_t - \bar{y}) \right]^2 / \left[\sum_t (\hat{y}_t - \bar{\hat{y}})^2 \right] \left[\sum_t (y_t - \bar{y})^2 \right]$$

This is square of simple correlation coefficient between y_t and $\hat{y}_t$, $r_{y\hat{y}}^2$

$$0 \leq \sqrt{r_{y\hat{y}}^2} \leq 1$$

→ aynı değişkenin kendisi ve fitted value'ları arasındaki correlation (-) olmaz çünkü o kadar ters yönde hareket edemezler.

→ cannot be negative since y_t and $\hat{y}_t$ cannot move in opposite directions.

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$$

$$y_t \rightarrow x_t \quad r > 0$$

$$\hat{y}_t \rightarrow y_t \quad r > 0$$

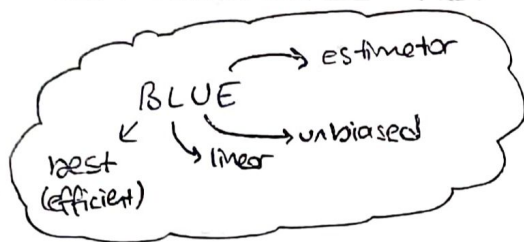
$$-1 \leq \sqrt{r_{xy}^2} \leq 1$$

→ x ve y disposable

arasındaki negatif olabilir, mesela income o'ken saving - olabilir.
 $\hat{s}_t = -7 - 0.3 y_t^d$

Finite Sample Properties of $\hat{\beta}_0$ and $\hat{\beta}_1$.

(4)



Define $a_{t1} = x_t / \sum_t x_t^2$

$$\hat{\beta}_1 = \sum_t x_t y_t / \sum_t x_t^2 = \sum_t a_{t1} y_t$$

$$\hat{\beta}_1 = \sum_t a_{t1} (y_t - \bar{y})$$

$$\hat{\beta}_1 = \sum_t a_{t1} y_t - \bar{y} \sum_t a_{t1}$$

$\hat{\beta}_1 = \sum_t a_{t1} y_t$ is linear with y_t so $\hat{\beta}_1$ is a linear estimator.

For $\hat{\beta}_0$,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_0 = \sum_t y_t / T - \hat{\beta}_1 (\sum_t x_t / T) = \sum_t y_t T^{-1} - \hat{\beta}_1 \sum_t x_t T^{-1}$$

$$\hat{\beta}_0 = \sum_t y_t T^{-1} - \sum_t a_{t1} y_t (\sum_t x_t T^{-1})$$

$$= \sum_t y_t T^{-1} - \sum_t a_{t1} y_t \bar{x}$$

$$= \sum_t y_t T^{-1} - \bar{x} \sum_t a_{t1} y_t = \sum_t \underbrace{(T^{-1} - \bar{x} a_{t1})}_{a_{t0}} y_t = \sum_t a_{t0} y_t$$

Shown that both $\hat{\beta}_0$ and $\hat{\beta}_1$ belong to set of linear estimators.

! HW: $y_t = \beta_0 + \beta_1 x_t + u_t$

For any linear unbiased estimators of β_0 and β_1 . Show that OLS estimators of β_0 and β_1 are efficient.

$$\text{var}(\tilde{\beta}_i)$$

$\tilde{\beta}_i$ are best, linear, unbiased estimators of β_i (BLUE)

→ This result is known as Gauss-Markov Theorem → OLS estimators are robust.

Estimating σ^2

If disturbances u_t were observable a quadratic unbiased estimator, σ^2 would be obtained as

$$\hat{u}_t = y_t - \hat{y}_t \rightarrow \text{observable.}$$

$$s^2 = \sum_t (u_t - \bar{u})^2 / T - 1$$

Since u_t 's are not observable, we need sample counterparts of them, $\hat{u}_t$.

$$\sum_t (\hat{u}_t - \hat{\bar{u}})^2 = \sum_t \hat{u}_t^2 \quad \text{since } \hat{\bar{u}} = T^{-1} \sum_t \hat{u}_t = 0$$

$$\hat{\sigma}^2 = \sum_t \hat{u}_t^2 / (T - 2) \rightarrow \boxed{\hat{\sigma}^2 = \sum_t \hat{u}_t^2 / (T - k - 1)}$$

number of explanatory variables.

For unbiasedness, $\tilde{\beta}_i = \sum_t c_{ti} y_t$ be any linear estimator of two regression coefficients. ⑤

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

$$\tilde{\beta}_i = \sum_t c_{ti} (\beta_0 + \beta_1 x_t + u_t) = \sum_t c_{ti} \beta_0 + \sum_t c_{ti} \beta_1 x_t + \sum_t c_{ti} u_t$$

$$E(\tilde{\beta}_i) = \beta_0 \sum_t c_{ti} + \beta_1 \sum_t c_{ti} x_t \longrightarrow E(\tilde{\beta}_0) = \beta_0 \underbrace{\sum_t c_{t0}}_0 + \beta_1 \underbrace{\sum_t c_{t0} x_t}_0$$

$$E(\tilde{\beta}_0) = \beta_0$$

$$E(\tilde{\beta}_1) = \beta_0 \sum_t c_{t1} + \beta_1 \sum_t c_{t1} x_t$$

$$E(\tilde{\beta}_1) = \beta_1$$

Therefore, these two conditions make $\tilde{\beta}_0$ and $\tilde{\beta}_1$ unbiased.

Note that $\sum_t a_{t1} = \sum x_t / \sum x_t^2 = 0$

$$\sum_t a_{t1} x_t = \sum_t a_{t1} x_t = \sum_t x_t^2 / \sum_t x_t^2 = 1$$

Since $\sum_t a_{t1} = \sum_t x_t / \sum_t x_t^2 \longrightarrow \sum_t a_{t1} x_t = \sum_t x_t^2 / \sum_t x_t^2$

$$\sum_t a_{t0} = 1 - \bar{x} \underbrace{\sum_t a_{t1}}_0 = 1 \quad \text{since } a_{t0} = T^{-1} - \bar{x} a_{t1}$$

$$\begin{aligned} \sum_t a_{t0} x_t &= \sum_t x_t (T^{-1} - \bar{x} a_{t1}) \\ &= \bar{x} - \bar{x} \underbrace{\sum_t a_{t1} x_t}_1 = 0 \end{aligned}$$

at i, it satisfies conditions. Thus, LS estimators of β_0 and β_1 are linear and unbiased estimator.

$$E(\hat{\beta}_i) = \beta_i$$

Let's prove $\hat{\sigma}^2$ is an unbiased estimator of σ^2 .

(7)

$$\bar{\hat{u}} = \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}$$

Subtract this from $\hat{u}_t = y_t - \hat{y}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t$

$\hat{u}_t = y_t - \hat{\beta}_1 x_t \rightarrow$ substituting this for $y_t = \beta_1 x_t + u_t - \bar{u}$

$$\hat{u}_t = -(\hat{\beta}_1 - \beta_1) x_t + u_t - \bar{u}$$

$$\sum_t \hat{u}_t^2 = (\hat{\beta}_1 - \beta_1)^2 \sum_t x_t^2 + \sum_t (u_t - \bar{u})^2 = 2 \sum_t x_t (\hat{\beta}_1 - \beta_1) (u_t - \bar{u})$$

$$E[(\hat{\beta}_1 - \beta_1)^2] \sum_t x_t^2 = \sigma_{\hat{\beta}_1}^2 \sum_t x_t^2 = \sigma^2$$

$$E[E_t(u_t - \bar{u})^2] = (T-1)E(s^2) = (T-1)\sigma^2$$

Since s^2 is an unbiased estimator of σ^2 ,

$$2E\left[\sum_t x_t (\hat{\beta}_1 - \beta_1) (u_t - \bar{u})\right] = 2 \sum_t x_t E[(\hat{\beta}_1 - \beta_1) u_t] \text{ since } E[(\hat{\beta}_1 - \beta_1) \bar{u}] \sum_t x_t = 0$$

$$\text{But } E[(\hat{\beta}_1 - \beta_1) u_t] = \sum_s a_{s1} E(u_s u_t) = \sigma^2 a_{t1}$$

$$2E \sum_t x_t (\hat{\beta}_1 - \beta_1) (u_t - \bar{u}) = 2\sigma^2 \sum_t a_{t1} x_t = 2\sigma^2$$

$$E(\hat{\sigma}^2) = \sigma^2 [1 + (T-1) - 2] / T - 2$$

$$= \sigma^2 [(T-2) / T - 2]$$

$$E(\hat{\sigma}^2) = \sigma^2 \quad \hat{\sigma}^2 \text{ is unbiased estimator of } \sigma^2.$$

By using $\hat{\sigma}^2$, we can obtain estimators for $\sigma_{\beta_0}^2$ and $\sigma_{\beta_1}^2$

$$\hat{\sigma}_{\beta_0}^2 = \hat{\sigma}^2 \frac{\sum_t x_t^2 / T}{\sum_t x_t^2} \rightarrow \frac{\sum_t \hat{u}_t^2 / (T-2)}{\sum_t x_t^2}$$

$$= \left(\frac{\sum_t \hat{u}_t^2}{T-2} \right) \left(\frac{\sum_t x_t^2}{\sum_t x_t^2} \right) / \left(\frac{\sum_t x_t^2}{T-2} \right)$$

$$\hat{\sigma}_{\beta_1}^2 = \sigma^2 / \sum_t x_t^2 = \frac{\sum_t \hat{u}_t^2 / (T-2)}{\sum_t x_t^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum_t x_t y_t}{\sum_t x_t^2}$$

} point estimators

Interval Estimation

$$P \left[\hat{\beta}_i - \hat{\sigma}_{\hat{\beta}_i} t_{T-2, \alpha/2} \leq \beta \leq \hat{\beta}_i + \hat{\sigma}_{\hat{\beta}_i} t_{T-2, \alpha/2} \right] = 1 - \alpha \quad \left\{ \begin{array}{l} \text{interval estimator} \\ \text{for } \beta \end{array} \right.$$

$$\frac{\sum_t \hat{u}_t^2 / \sigma^2}{(T-2)} = \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{T-2}^2$$

$$P \left[\chi_{T-2, \alpha/2}^2 \leq (T-2) \hat{\sigma}^2 / \sigma^2 \leq \chi_{T-2, 1-\alpha/2}^2 \right] = 1 - \alpha$$

$$P \left[(T-2) \sigma^2 / \chi_{T-2, 1-\alpha/2}^2 \leq \sigma^2 \leq (T-2) \hat{\sigma}^2 / \chi_{T-2, \alpha/2}^2 \right] = 1 - \alpha$$

} interval estimator for σ^2

Hypothesis Testing

9

$$Y_t = \beta_0 + \beta_1 X_t + U_t, \quad t=1, \dots, T$$

$$H_0: \beta_i = \beta_i^0$$

$$H_1: \beta_i \neq \beta_i^0$$

or

$$\beta_i < \beta_i^0$$

or

$$\beta_i > \beta_i^0$$

$$\beta_1 = 0.5$$

$$t = \frac{\hat{\beta}_i - \beta_i^0}{\hat{\sigma}_{\hat{\beta}_i}} \sim t_{T-2}$$

$$\textcircled{1} H_1: \beta_i \neq \beta_i^0$$

H_0 will not be rejected if

$$(\hat{\beta}_i - \beta_i^0) / \hat{\sigma}_{\hat{\beta}_i} \leq t_{T-2, \alpha/2}$$

$$\textcircled{2} H_1: \beta_i < \beta_i^0$$

H_0 will not be rejected if

$$(\hat{\beta}_i - \beta_i^0) / \hat{\sigma}_{\hat{\beta}_i} \geq -t_{T-2, \alpha}$$

$$\textcircled{3} H_1: \beta_i > \beta_i^0$$

H_0 will not be rejected if

$$(\hat{\beta}_i - \beta_i^0) / \hat{\sigma}_{\hat{\beta}_i} \leq t_{T-2, \alpha}$$

Individual Tests of Significance $\otimes \rightarrow$ final (5 points)

$$H_0: \beta_i = 0$$

$$H_1: \beta_i > 0$$

or

$$\beta_i < 0$$

or

$$\beta_i \neq 0$$

$$\textcircled{1} H_1: \beta_i \neq 0$$

Test statistics is

$$|\hat{\beta}_i| / \hat{\sigma}_{\hat{\beta}_i} \leq t_{T-2, \alpha/2}, \text{ then } H_0 \text{ will not be rejected.}$$

NO NOT FORGET!

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum X_t Y_t}{\sum X_t^2}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{U}_t^2}{T-2}$$

$$Y_t = \beta_0 + \beta_1 X_t + U_t$$

$$H_0: \beta_0 = 0$$

$$H_1: \beta_0 \neq 0$$

not reject H_0 .

$Y_t = \beta_1 X_t + U_t$, assuming $H_1: \beta_1 = 0$ is rejected.

Testing Linear Restrictions on Regression Coefficients

$$Y_t = \beta_0 + \beta_1 K_t + \beta_2 L_t + u_t$$

$$H_0: \beta_1 + \beta_2 = 1 \rightarrow \text{constant returns to scale}$$

$$\beta_1 + \beta_2 > 1$$

$$\beta_1 + \beta_2 < 1$$

an microeconomic example with Cobb-Douglas prod. function with L, K .

$$H_0: w_0 \beta_1 + w_1 \beta_2 = q \quad \left\{ \begin{array}{l} \text{general restriction} \\ H_1: w_0 \beta_1 + w_1 \beta_2 \neq q \end{array} \right.$$

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + u_t \rightarrow \text{a production function}$$

$$S_u \rightarrow SSR_u \rightarrow \text{unrestricted sum of squared residuals.}$$

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + u_t \rightarrow SSR_e$$

$$F = (S_R - S_u) / S_u \sim F_{p, T-2}$$

↘ number of restrictions

$$H_0: \beta_0 = \beta_1 = \beta_2 = 0 \quad \left\{ \begin{array}{l} \text{joint test significance} \\ H_1: \beta_0 \neq \beta_1 \neq \beta_2 \neq 0 \end{array} \right.$$

$$Q = [(S_R - S_u) / S_u] (T-2/p) \sim F_{p, T-2}$$

$$Y_t = \beta_0 + \beta_1 X_t + u_t \rightarrow H_0: \beta_1 = 0$$

$$H_0: \beta_0 = \beta_2 = 0 \quad \left\{ \begin{array}{l} \text{joint test significance} \\ H_1: \beta_0 \neq \beta_2 \neq 0 \end{array} \right.$$

$$Y_t = \beta_1 X_t + u_t$$

↓
restricted model $\rightarrow S_e$

$$Q = [R^2 / (1 - R^2)] (T-2) \sim F_{1, T-2}$$

Prediction

13.05.19

ECON206

①

y_p determines observed values of dependent variable to be predicted.

$$y_p = \beta_0 + \beta_1 x_p + u_p$$

$$\begin{array}{c} 1990 \\ \vdots \\ 2017 \end{array} \begin{bmatrix} y_t \\ \vdots \\ y_t \end{bmatrix} \quad \begin{bmatrix} x_t \\ \vdots \\ x_t \end{bmatrix} \quad \hat{y} = 50 + 0.5x \rightarrow x_{2018} \rightarrow 50 \quad (E^e \text{ gibi.})$$

$$\begin{array}{ccc} \hat{y}_{2018} & \rightarrow & y_{2018} \\ \downarrow & & \downarrow \\ 75 & \leftrightarrow & 77 \\ & \hat{u}_{2018} & \end{array}$$

$E(u_p) = 0, \text{ var}(u_p) = \text{var}(u_t) = \sigma^2$ $E(u_t u_p) = 0$ $\rightarrow$ rasyonel beklenti modelinde
geçmişten öğrenilir ve sistematik
hata olmaz. (ut ve up arasında
lineer ilişki var.)

Regression coefficients remain same for points to be predicted.

Predictor of $\hat{y}_p$ would be obtained by extending estimator of $E(y)$.

$$\hat{y}_p = \hat{\beta}_0 + \hat{\beta}_1 x_p$$

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$$

$$\hat{\beta}_0 = \sum_t a_{t0} y_t, \quad \hat{\beta}_1 = \sum_t a_{t1} y_t$$

$$\hat{y}_p = \sum_t a_{t0} y_t + x_p \sum_t a_{t1} y_t = \sum_t a_{tp} y_t$$

this predict also linear.

$$a_{tp} = a_{t0} + x_p a_{t1}$$

$\hat{y}_p$ is a linear predictor of y_p .

$$\sum_t a_{tp} = \sum_t a_{t0} + x_p \sum_t a_{t1} = 1$$

$$\sum_t a_{tp} x_t = \sum_t a_{t0} x_t + x_p \sum_t a_{t1} x_t = x_p$$

$$\left. \begin{array}{l} y_t = \beta_0 + \beta_1 x_t + u_t \\ \hat{y}_p = \beta_0 \sum_t a_{tp} + \beta_1 \sum_t a_{tp} x_t + \sum_t a_{tp} u_t \end{array} \right\}$$

$$\hat{y}_p = \beta_0 + \beta_1 x_p + \sum_t a_{tp} u_t \longrightarrow \hat{y}_p = E(y_p) + \sum_t a_{tp} u_t$$

$$E(\hat{y}_p) = E(E(y_p) + E(\sum_t a_{tp} u_t))$$

$E(\hat{y}_p) = E(y_p)$ so, $\hat{y}_p$ is unbiased predictor of y_p .

$$\begin{aligned}\sigma_{\hat{y}_p}^2 &= E[(y_p - E(\hat{y}_p))^2] \\ &= E(\sum_t a_{tp} u_t)^2 \\ &= E(\sum_t a_{tp}^2) + \frac{E(u_t)^2}{\sigma^2}\end{aligned}$$

$$= \sigma^2 \sum_t a_{tp}^2$$

$$a_{t0} = T^{-1} - \bar{x} a_{t1}$$

$$\sum_t a_{tp}^2 = (\sum_t a_{t0} + x_p \sum_t a_{t1})^2$$

$$\sigma_{\hat{y}_p}^2 = \sigma^2 \sum_t (T^{-1} - \bar{x} a_{t1} + x_p a_{t1})^2$$

$$x_p = x_p - \bar{x}$$

$$\sigma_{\hat{y}_p}^2 = \sigma^2 \sum_t (T^{-1} + x_p a_{t1})^2$$

$$= \sigma^2 (T^{-1} + x_p^2 \sum_t a_{t1}^2 + 2T^{-1} x_p \sum_t a_{t1})$$

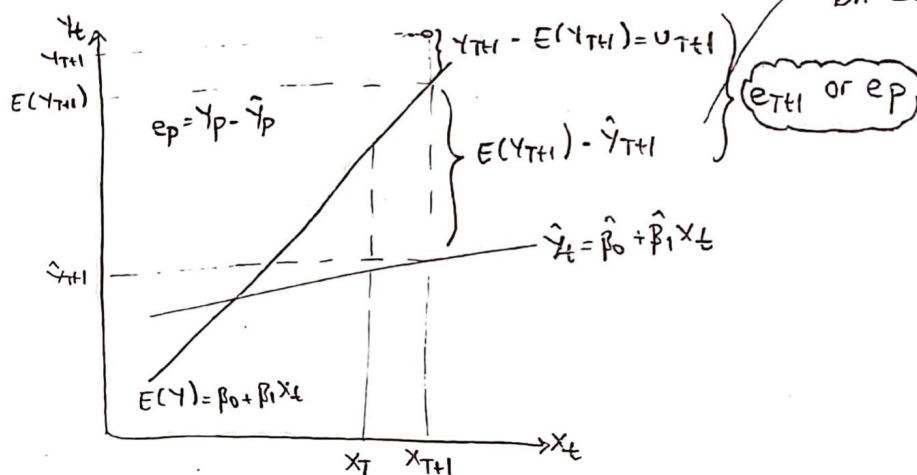
$$\text{since } \sigma^2 \sum_t a_{t1}^2 = \sigma^2 / \sum_t x_t^2$$

$$\sigma_{\hat{y}_p}^2 = \sigma^2 [T^{-1} + (x_p^2 / \sum_t x_t^2)]$$

$\hat{y}_p$ has a minimum variance and its BLUE predictor.

Prediction error is

③



$$e_p = y_p - \hat{y}_p$$

$$e_p = y_p - E(y_p) + E(y_p) - \hat{y}_p$$

$$= u_p + [E(y_p) - \hat{y}_p]$$

disturbance bias in predicting y_p

* If our classical assumptions hold

$$E(e_p) = 0$$

↳ hypothesis testing:

$H_0: E(e_p) = 0$ } to assess predictive performance of model.
 $H_1: E(e_p) \neq 0$

If H_0 is rejected $\xrightarrow{\text{due to 2}}$ model is misspecified. $\Rightarrow E(u_p) \neq 0$
 $\Rightarrow E(y_p) \neq E(\hat{y}_p) \Rightarrow \hat{y}_p$ is a biased estimator of y_p .

$$H_0: E(e_p) = 0$$

$$H_1: E(e_p) \neq 0$$

$$\sigma_{ep}^2 = \sigma^2 (1 + T^{-1} + (x_p^2 / \sum_t x_t^2)) \rightarrow \text{multiple choice}$$

$$u_p \sim N(0, \sigma^2)$$

$$e_p \sim N(0, \sigma_{ep}^2)$$

$$e_p / \hat{\sigma}_{ep} \sim t_{T-2, \alpha/2}$$

$$|e_p / \hat{\sigma}_{ep}| > t_{T-2, \alpha/2}$$

→ it is rejected.

Confidence Interval for y_p

$$P[\hat{y}_p - \hat{\sigma}_{ep} t_{T-2, \alpha/2} \leq y_p \leq \hat{y}_p + \hat{\sigma}_{ep} t_{T-2, \alpha/2}] = 1 - \alpha$$

homework: model without intercept term,

$$y_t = \beta x_t + u_t$$

① Obtain OLS estimator of $\beta, \hat{\beta}$.

$$y_t - \beta x_t = u_t$$

$$\min \sum_t u_t^2$$

② Show that $\hat{\beta}$ is BLUE.

→ interceptsiz modelde R^2 ekisi çıktıyı için,

$$R_r^2 = 1 - (\sum_t \hat{u}_t^2 / \sum_t y_t^2) \rightarrow 0 \leq R^2 \leq 1$$

(özellik)

Problem of Functional Form

- (5) $Y_t = \beta_0 + \beta_1 X_t + u_t$ $R^2 = 0.703$
- (1) $\ln Y_t = \beta_0 + \beta_1 X_t + u_t$ $R^2 = 0.856$ } semi-log form
- (2) $Y_t = \alpha_0 + \alpha_1 \ln X_t + v_t$ $R^2 = 0.813$
- (3) $\ln Y_t = \delta_0 + \delta_1 \ln X_t + w_t$ } loglinear (double-log form)
 $R^2 = 0.897$

Slope terms ($\beta_1, \alpha_1, \delta_1$) cannot be interpreted as dY/dX .

from 1 : $\beta_1 = d \ln Y_t / d X_t = (1/Y_t) (d Y_t / d X_t) \Rightarrow d Y_t / d X_t = \beta_1 Y_t$

from 2 : $\alpha_1 = d Y_t / d \ln X_t = X_t (d Y_t / d X_t) \Rightarrow d Y_t / d X_t = \alpha_1 / X_t$

→ constant point elasticity

from 3 : $\delta_1 = d \ln Y_t / d \ln X_t = (d Y_t / d X_t) (X_t / Y_t) \Rightarrow d Y_t / d X_t = \delta_1 (Y_t / X_t)$

* Tenevlerde kalınca
esnekliği görüyoruz.
Mesela para esnekliğini
bubbliriz.

These types of model are called intrinsically linear models. Take antilogs of (3)

$$Y_t = \hat{\delta}_0^* X_t^{\delta_1} w_t^*$$

$$\hat{\delta}_0^* = e^{\hat{\delta}_0} \quad w_t^* = e^{w_t}$$

$\hat{\delta}_1$ is still BLUE but $\hat{\delta}_0^*$ has no longer finite sample properties.

$\hat{\delta}_0^* = e^{\hat{\delta}_0}$ is nonlinear function of $\hat{\delta}_0$, it is no longer a linear or unbiased estimator since $E(\hat{\delta}_0^*) \neq e^{E(\hat{\delta}_0)}$

(4) $Y_t = \beta_0 + \beta_1 (1/X_t) + u_t$ $R^2 = 0.666$
hyperbolic or reciprocal form.

$$d Y_t / d X_t = -\beta_1 (1/X_t^2)$$

* Eğer ki model oluşturdysak
ve hepsinin de β_0 ve β_1 'i
significant ise, R^2 'si büyük
olan seçeriz!

① specification of dependent variable should be same in all models^⑥,
compared.

(2-4-5, R^2 kullanarak karşılaştırılabilir) $\rightarrow Y_t, \ln Y_t$ ayırımı.

$\rightarrow \underline{\underline{R^2 \text{ büyük olan al.}}}$

② slope terms should be significant.

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0.$$

③ if you have model without intercept term use R^2 .