Confidence

two things affect our "confidence" in an estimate

1) how spread out the samples are

2) how many sample there are

Confidence Interval

Data $\{x_1 \cdots x_N\}$

Sample mean $\hat{\mu} = \bar{x} = \frac{1}{N} \sum x_i$

sample variance $\hat{\sigma} = \frac{1}{N-1} \sum (x_i - \bar{x})^2$

95% confidence Interval $\left[\hat{\mu} - 1.96 \cdot \frac{\hat{\sigma}}{\sqrt{N}}, \ \hat{\mu} + 1.96 \frac{\hat{\sigma}}{\sqrt{N}}\right]$

分布の幅

幅の大きさ

Sample mean

$$\hat{\mu} = \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

functions of RVs are RVs

ex) X_1 X_2 : result of coin toss

$Y = X_1 + X_2$

$\{0,1\} \in X_1$, $\{0,1\} \in X_2$ より

$\rightarrow$ Y is a random valuable

Sums of Normal

$$X_1 \sim N(\mu_1, \sigma_1^2)$$
$$X_2 \sim N(\mu_2, \sigma_2^2)$$
$$Y = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$\downarrow$

we can extend this to N IID

$$X \sim N(\mu, \sigma_2)$$
$$X_1 + X_2 + \cdots + X_n \sim N(N\mu, N\sigma^2)$$

Mean of the Estimate

$$E(\hat{\mu}) = E\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)$$
$$= E\left(\frac{1}{N}(x_1 + x_2 + \cdots + x_N)\right)$$
$$= \frac{1}{N}E(x_1) + \frac{1}{N}E(x_2) + \cdots + \frac{1}{N}E(x_N)$$
$$= \frac{1}{N}\mu + \frac{1}{N}\mu + \cdots \frac{1}{N}\mu$$
$$= \mu$$

Variance of Estimate

$$
\begin{aligned}
\mathrm{var}(\hat{\mu}) &= \mathrm{var}\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) \\
&= E\left[\left(\frac{1}{N}\sum_{i=1}^{N} x_i - \mu\right)^2\right] \\
&= \frac{1}{N^2} E\left[\left(\sum_{i=1}^{N} x_i - N\mu\right)^2\right] \\
&= \frac{1}{N^2} \mathrm{Var}\left(\sum_{i=1}^{N} x_i\right) \\
&= \frac{1}{N^2} N \sigma^2 \\
&= \frac{\sigma^2}{N}
\end{aligned}
$$

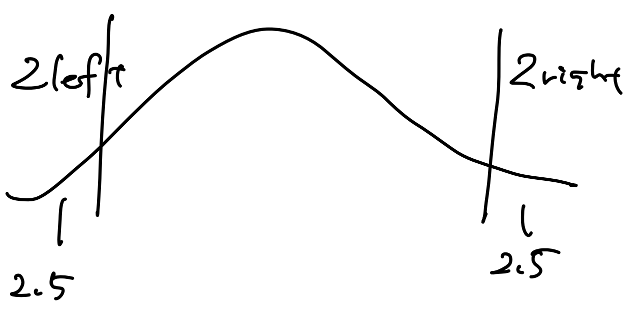
$$\sigma^2 = V[X] = E[(X-\mu)^2]$$

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$$

when σ^2 is larger, variance of estimate is larger

" N " " smaller

パーセント点関数



$Z \sim N(0,1)$

$$0.025 = \int_{-\infty}^{z_{left}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \underline{\Phi(z_{left})}$$

cdf

$$z_{left} = \Phi^{-1}(0.025) = -1.96$$

Standardization

$$x \sim N(\mu, \sigma_0^2) \rightarrow z = \frac{x-\mu}{\sigma_0} \sim N(0,1)$$

Confidence Interval

$$-1.96 \le z \le 1.96$$

$$-1.96 \le \frac{\hat{\mu}-\mu}{\sigma_0} \le 1.96$$

$$-1.96\,\sigma_0 \le \hat{\mu}-\mu \le 1.96\,\sigma_0$$

$$-1.96\,\sigma_0 - \hat{\mu} \le -\mu \le 1.96\,\sigma_0 - \hat{\mu}$$

$$\hat{\mu} - 1.96\,\sigma_0 \le \mu \le \hat{\mu} + 1.96\,\sigma_0$$

we can estimate

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

$$95\% - CI = [\hat{\mu} - 1.96\frac{\sigma}{\sqrt{N}}, \hat{\mu} + 1.96\frac{\sigma}{\sqrt{N}}]$$

$$\gamma - CI = [\hat{\mu} + \Phi^{-1}(\frac{1-\gamma}{2})\frac{\sigma}{\sqrt{N}}, \hat{\mu} + \Phi^{-1}(1-\frac{1-\gamma}{2})\frac{\sigma}{\sqrt{N}}]$$

Ex. $\gamma = 90\%$ or 0.9

$$\hat{\mu} + \underbrace{\Phi^{-1}(\frac{1-0.9}{2})}_{\Phi^{-1}(\frac{0.05}{2})}\frac{\sigma}{\sqrt{N}}, \hat{\mu} + \underbrace{\Phi^{-1}(1-\frac{1-0.9}{2})}_{\Phi^{-1}(1-0.05)}\frac{\sigma}{\sqrt{N}}]$$

CI : if I do some experiment many times, then 95% of those times the true μ

will be contained in the CI

CLT

coin flip : $x = \{0, 1\}$

$X \sim Ber(p)$ iid

$\hat{p} = \bar{x}$

$E(\hat{p}) = E(X) = p$

$var(X) = p(1-p)$

$\hat{p} \to N\left(p, \frac{p(1-p)}{N}\right)$

$$t = \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{N}} \times \frac{1/\sigma}{1/\sigma} = \left(\frac{\hat{\mu} - \mu}{\sigma/\sqrt{N}}\right) \Big/ \left(\frac{\hat{\sigma}}{\sigma}\right)$$

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

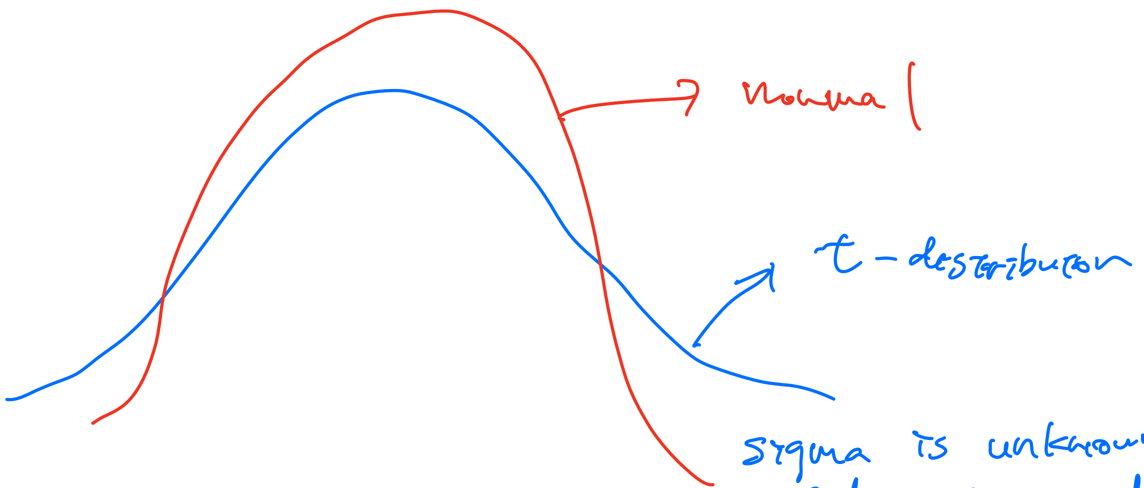
$$(N-1)\left(\frac{\hat{\sigma}}{\sigma}\right)^2 \sim \chi^2_{N-1}$$

$$Z \sim N(0, 1) \to \sum_{i=1}^{N} (Z_i - \bar{Z}) \sim \chi^2_{N-1}$$

T - Distribution

$$\frac{\text{standard normal}}{\sqrt{\text{chi-square / deg of freedom}}} \sim t \text{ deg of freedom}$$

when you standardize $\bar{X}$ or $\hat{\mu}$ with the sample standard diviation, it's gonna be t-distribution



sigma is unknown, confidence is spread out

Hypothesis Testing

1-sample test : input data is 1-D array

2-sample test : " 2 1-D array

scipy, statsmodel が使用できる

output : (test statistic , p-value)

the prob of observing a result as extreme or more extreme than what was observed assuming the null hypothesis is true

ex) $H_0: \mu = 0$

Observed $\bar{x} = 100$, var = 10

→ unlikely, p value will be small

CI では looking inside the middle 95%

hypothesis Testing では looking outside of the 5%

ex) significance threshold is 1%

→ p-value is 0.1

0.1 > 0.01 (not statistically significant)

if the cost of a false alarm is high,
set a strict significance threshold

Terminology

if p-value < significance threshold

→ reject null hypothesis

EL, p-value > significance threshold

→ not binary (H_0 or H_1)

only reject or fail to reject

Examples

- 正規分布かの検定
- データが異常か？

How do we calculate the p-value ?

$P_{right} = 1 - \Phi(|z|)$

$P_{left} : \Phi(-|z|)$

$p = P_{left} + P_{right}$

2 sample Z-test

convert it into a 1-sample test

Group 1 : $\{x_1, \cdots x_N\}$ with mean μ_1
Group 2 : $\{x_1 \cdots x_N\}$ with mean μ_2

$H_0 : \mu_1 = \mu_2 \iff H_0 : \mu_1 - \mu_2 = 0$

Let $y = \mu_1 - \mu_2$
$H_0 : y = 0$
$H_1 : y \neq 0$

$H_0 : \hat{y} \sim N(0, \sigma_{\hat{y}}^2)$ or $H_0 : \hat{y} \sim N(\mu_0, \sigma_{\hat{y}}^2)$

$$\hat{y} = \frac{x_1 + x_2 + \cdots + x_{N_1}}{N_1} - \frac{x_1' + x_2' + \cdots + x_{N_2}'}{N_2}$$

$$\begin{aligned}
\mathrm{var}(\hat{y}) &= \mathrm{var}\left(\frac{x_1}{N_1}\right) + \cdots + \left(\frac{x_{N_1}}{N_1}\right) + \mathrm{var}\left(\frac{x_1'}{N_2}\right) + \\
&= \frac{\sigma_1^2}{N_1^2} + \cdots + \frac{\sigma_1^2}{N_1^2} + \frac{\sigma_2^2}{N_2^2} + \cdots + \frac{\sigma_2^2}{N_2^2} \\
&= \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}
\end{aligned}$$

$$\sigma(\hat{y}) = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

Z-statistic

$$z = \frac{\hat{y} - \mu_0}{\sigma_{\hat{y}}} = \frac{\hat{y} - \mu_0}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

検定統計量：帰無仮説が正しいと仮定したときに、観測した事象よりも稀なことが起こる確率を計算するための値。

「

有意水準

← 以下 ; 以上 →

「頻出にないこと」; 「頻出にないこと」ではない

= 仮説が間違い ; = 仮説が間違っているとは言えない

」

p値 : 帰無仮説が正しいと仮定した時に、観測した事象よりも極端なことが起こる確率

T-Test

- stop making assumption that we know std.

$$t = \frac{\hat{\mu} - \mu}{\sigma / N}$$

Chi-square Test

- 0, 1 のデータについて正規分布を仮定している
  → この仮定をしたくない場合は chi-square test

Paired sample

- instead of having two groups, use only 1 group and test twice.

  ⟷ independent sample test

More than 2 groups.

- what if I test drug A, then B then drug C ....

$$\alpha_{\text{Bonferroni}} = \frac{\alpha}{\#\text{ tests}}$$

Frequentist Concerns.

- no peeking allowed (途中で検定できない)

Bayesian Approach

- we can answer, what is the probability that drug A > drug B?