

Recall

$X_1, \dots, X_n$ sample from a population mean normal (μ, σ^2)

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Assume σ^2 known

↑
variance

Using Central Limit Theorem

$$P\left(\mu \in \left(-\frac{\sigma}{\sqrt{n}}(1.96) + \bar{X}, \frac{\sigma}{\sqrt{n}}(1.96) + \bar{X}\right)\right) \approx 0.95$$

(I) $(-\frac{(1.96)\sigma}{\sqrt{n}} + \bar{X}, \frac{(1.96)\sigma}{\sqrt{n}} + \bar{X})$ - 95% confidence interval for μ

From the sample $X_1, \dots, X_n$ $\hat{\mu} = \bar{X}$

Can we say something more about μ ?

Suppose we believe $\mu = c$ for some $c \in \mathbb{R}$

Q: Can we use $X_1, \dots, X_n$ (data) to decide on the belief.

Testing: - we have an observed quantity $X_1, \dots, X_n$ & we wish

to compare this with an expected result
based on (our belief) the assumptions
we make about the distribution.

Null hypothesis - Specific conjecture made
on the distribution.

Alternate hypothesis :- Specifies a particular
manner in which the null hypothesis may
fail (if assumption on distribution are inaccurate)

Given:

$X_1, \dots, X_n$
data

Normal?

variance $= \sigma^2$

Null: $\mu = c$

Alternate: $\mu > c$

Q. If Null were true
then how likely will
we observe $X_1, X_2, \dots, X_n$?

Answer: (i) let $Y_1, \dots, Y_n$ be iid $\text{Normal}(c, \sigma^2)$

look at $P(\bar{Y} \geq \bar{X}) = P\left(\frac{\sqrt{n}(\bar{Y} - c)}{\sigma} \geq \frac{\sqrt{n}(\bar{X} - c)}{\sigma}\right)$

$= P\left(Z \geq \frac{\sqrt{n}(\bar{X} - c)}{\sigma}\right)$, where $Z \sim N(0,1)$.

(ii) [Before data is collected] decide on what is called Level of significance $= \alpha = 0.01$

ie - if $P(\bar{Y} \geq \bar{X}) \leq \alpha$

Then we conclude $\bar{X}$ is far from the assumed mean, so the assumption $\mu = c$ is incorrect.

[- we reject the null hypothesis)

If $P(\bar{Y} \geq \bar{X}) \geq \alpha$ then $\bar{X}$ may be consistent
with the assumptions made

[null hypothesis cannot be rejected]

Example: Take $n=16$, $\sigma=3$, $\bar{X}=10.2$ Null: $\mu=9$
 $\alpha=0.05$

$$P(\bar{Y} \geq \bar{X}) = P\left(2 > \frac{4}{3}(10.2 - 9.5)\right)$$

Alternate: $\mu >$

$\alpha=0.05$ and $p\text{-value} \approx 0.175$ — $p\text{-value}$
 $0.175 > 0.05 \Rightarrow$ we will not reject null hypothesis

(b) same as (a)

Null

$$\mu = 8.5$$

Alternate

$$\mu > 8.5$$

$$P(\bar{Y} > \bar{X}) = P\left(Z > \frac{4}{3}(10.2 - 8.5)\right)$$

As $\boxed{\text{P-value}} \approx 0.012$

$0.012 < 0.05$

we will reject the Null hypothesis.

ie If $\mu = 9.5$ assumption is made
then the sampling procedure ($Y_1, \dots, Y_{10}$)
will produce a result as
at least as large as $\bar{X} = 10.2$,
17.5% of the time. This larger
than 5% we will not reject the
null hypothesis.

Z-test σ -known. $n, \bar{X} \Rightarrow \begin{cases} x_1, \dots, x_n \text{ are Normal} \\ n \rightarrow \infty \text{ so CLT holds} \end{cases}$

Null Hypothesis	$\mu = c$	$\mu = c$	$\mu = c$
Alternate Hypothesis	$\mu > c$	$\mu < c$	$\mu \neq c$
Computation of P-value	$P(Z > \frac{\sqrt{n}(\bar{X}-c)}{\sigma})$	$P(Z < \frac{\sqrt{n}(\bar{X}-c)}{\sigma})$	$P(Z > \frac{\sqrt{n}(\bar{X}-c)}{\sigma})$

Fix α In each case if p-value $< \alpha$ we reject null hypothesis
 otherwise we do not reject null hypothesis

Facts from Probability

$X_1, \dots, X_n$ are iid Normal random variables (μ, σ^2)

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \stackrel{d}{=} N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad E[S^2] = \sigma^2$$

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

(chi-square with $n-1$ degrees of freedom)

induction

$\bar{X}$ and S^2 are independent random variables.

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \stackrel{d}{=} T_{n-1} \text{ where } T = \text{students } t_{n-1} \text{ distributed}$$

— Extensions :-

let $x_1, \dots, x_n$ be as above. μ & σ^2 unknown

From R or t-tables $P(|T| < t_{n-1}^{0.95}) = 0.95$

As before we can conclude

$$\left(-\frac{t_{n-1}^{0.95}}{\sqrt{n}} + \bar{X}, \frac{t_{n-1}^{0.95}}{\sqrt{n}} + \bar{X} \right) \text{ 95\% CI for } \mu$$

$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \stackrel{d}{=} T_{n-1}$ where $T =$ student's t_{n-1} distributed

— Extensions :-

let $X_1, \dots, X_n$ be as above. μ & σ^2 unknown

From

R & t-tables

$$P(|T| < t_{n-1}^{0.95}) = 0.95$$

As before we can conclude

$$\left(-\frac{t_{n-1}^{0.95}}{\sqrt{n}} + \bar{X}, \frac{t_{n-1}^{0.95}}{\sqrt{n}} + \bar{X} \right) \text{ 95\% CI for } \mu$$

Example: $\bar{X} = 10.2$, $\sigma = 3$, $n = 16$, $X_1, \dots, X_n$ are Normal
(Known)

(Recall)

$$\begin{aligned} 95\% \text{ C.I.} &= \left(-\frac{3}{4} (1.96) + 10.2, (1.96) \frac{3}{4} + 10.2 \right) \\ &= (8.73, 11.67) \end{aligned}$$

σ unknown

Take $S = 3$

95% C.I. for μ

$$n = 16 \quad P(|T| < 2.13) = 0.95$$

$$\begin{aligned} &\left(-\frac{3}{4} (2.13) + 10.2, (2.13) \frac{3}{4} + 10.2 \right) \\ &= (8.6, 11.8) \end{aligned}$$