Recall:

. $X_1, \ldots X_n$ iid X $\quad E[X] = \mu$, $Var[X] = \sigma^2$

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} \quad \& \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$\Rightarrow \quad E[\overline{X}] = \mu \qquad E[S^2] = \sigma^2$$

Example: - Suppose we have a coin with probability hea
occurring in toss was p. To gather information about p, w
the coin 100 times. $X_i = \begin{cases} 1 & \text{if } i\text{th toss was head} \\ 0 & \text{otherwise} \end{cases}$

Let $X_1, X_2, \ldots, X_{100}$ be the results. & we find that $\sum_{i=1}^{100} X_i = 6$

Q How do we use this information to infer something about p?

(A1) Find a good estimator $g(X_1, \ldots, X_{100})$ that provides an estimate for p. Typical statement: $\hat{p} = g(X_1, \ldots, X_n)$

(A2) Don't provide a single value but give a "confidence interval" for p. Typical statement: "With 90% confidence the actual probability of heads lies in the interval $(0.52, 0.68)$"

(A3) [Hypothesis test]. We make a conjecture about p & devise a computation to test the validity of the conjecture. Typical statement: "If the coin has 50% chance of getting a head then observing 60 heads or more in 100 tosses should occur only 5% of time. This suggest "$p = 0.5$" is not accurate"

Confidence Interval: let $X_1, X_2, \dots, X_n$ be iid X. $E[X] = \mu$ (unknown)

$Var[X] = \sigma^2$ (known)

$\bar{X} = \frac{X_1 + \dots + X_n}{n}$ let $\beta \in (0,1)$ be our "confidence level".

We want to find an interval of width "a" on each side such that

$$\mathbb{P}(|\bar{X} - \mu| < a) = \beta$$

= if X_i were normally distributed

$\approx$ otherwise it is an approximation

Use CLT: $Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim Normal(0,1)$ for large n

$$\beta = \mathbb{P}(\underbrace{|\bar{X} - \mu| < a}_{\bar{X} - a < \mu < \bar{X} + a}) = \mathbb{P}\left(\left|\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}\right| \leq \frac{\sqrt{n}}{\sigma} a\right) \approx \mathbb{P}\left(|Z| < \frac{\sqrt{n}}{\sigma} a\right)$$

Given: n, β, σ we can use the normal tables to solve/find a such that $P(|Z| < \frac{\sqrt{n}}{\sigma} a) = \beta$

Conclusion:- $(\bar{X} - a, \bar{X} + a)$ is β-confidence interval for μ.

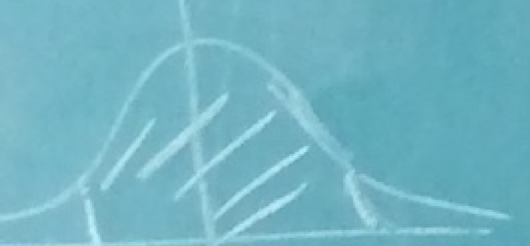
Rephrase:- Given n, $\bar{X}$ will produce an interval for μ of width a with probability β

Example 1

$X \sim \text{Normal}(\mu, 9)$ $\beta = 0.95$, $n = 16$, $\bar{X} = 10.2$

Find a: $\mathbb{P}(|\bar{X} - \mu| < a) = 0.95$

ie $$P\left(\left|\frac{\sqrt{16}\,(\bar{X} - \mu)}{3}\right| < \frac{\sqrt{16}\,a}{3}\right) = 0.95$$



ie $$P\left(|Z| < \frac{4a}{3}\right) = 0.95, \quad Z \sim \text{Normal}(0,1)$$

From Normal tables, we find $\frac{4a}{3} = 1.96 \Rightarrow a \sim 1.47$

$\therefore (\bar{X} - a, \bar{X} + a) = (8.73, 11.67)$ is a 95% confidence interval for μ from a sample of size 16.

Confidence interval for μ when σ^2 is not known

$X \sim \text{Normal}(\mu, \sigma^2)$

$X_1, \ldots, X_n$ be iid sample from X, $E[X] = \mu$, $\text{Var}[X] = \sigma^2$

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} \qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$\beta \in (0,1)$ Find a: $\mathbb{P}(|\bar{X} - \mu| < a) = \beta$

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \qquad T \sim t_{n-1} \text{ distribution}$$

$\boxed{\text{Given } n, \beta \text{ use } t\text{-Tables to find } a \text{ from } X_1, X_2, \ldots, X_n}$

Find a. $\beta = P(|\bar{X} - \mu| < a) = P\left(\left|\frac{\sqrt{n}(\bar{X} - \mu)}{S}\right| < \frac{\sqrt{n}\,a}{S}\right) = P\left(T < \frac{\sqrt{n}\,a}{S}\right)$

$\beta = 0.95$, $n = 16$, $\bar{X} = 10.2$

$S = 3$

Example 1

(continued)

$X \sim \text{Normal}(\mu, \sigma^2)$

Find a: $P(|\bar{X} - \mu| < a) = 0.9$

ie $P\left(4\frac{|\bar{x} - \mu|}{S} < \frac{4a}{S}\right) = 0.9$

ie $P\left(T < \frac{4a}{3}\right) = 0.9$

Find t-tables, $n=16$

$\frac{4a}{3} = 2.13 \Rightarrow a \sim 1.6$. $(8.6, 11.8)$ is

a 95% confidence interval for μ (σ - unknown) from a sample of size 16

Hypothesis testing

- This is a different approach to comparing an observed value to its expected value based on assumption on the distribution.

H_0: "Null hypothesis" – a specific conjecture about the nature of distribution of X.

(compared to)

H_1: "Alternative hypothesis" – a specific statement detailing a particular manner by which the null hypothesis is inaccurate

Goal:- To determine if there is enough statistical justification to reject the null - hypothesis. One arrive at such a conclusion based on a suitable statistic based on the sample

Test for sample mean when variance is known

Suppose $X \sim Normal(\mu, \sigma^2)$

$X_1, \ldots, X_n$ be (iid) sample from X. Let c be a value below $\overline{X}$

$H_0: \mu = c$ & $H_1: \mu > c$

If the null hypothesis was true, then how likely is it that we will see a sample mean as large as $\bar{X}$?

- View $\bar{X}$ as a deterministic number
- let $Y_1 \ldots Y_n$ be iid X [mimics the sampling]

Calculate (under null, is $\mu = c$)

$$\mathbb{P}(\bar{Y} > \bar{X})$$

$$= \mathbb{P}\left(\frac{\sqrt{n}(\bar{Y}-c)}{\sigma} > \frac{\sqrt{n}(\bar{X}-c)}{\sigma}\right) = \mathbb{P}\left(Z > \frac{\sqrt{n}(\bar{X}-c)}{\sigma}\right)$$

$$Z = \frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma} \sim \text{Normal}(0,1)$$

Fix $\alpha \in (0,1)$. If $\mathbb{P}\left(Z > \frac{\sqrt{n}(\bar{X}-c)}{\sigma}\right) < \alpha$

then we shall deem the hypothesis that $\mu = c$ is inaccurate ie we reject the null hypothesis

If $\mathbb{P}\left(Z > \frac{\sqrt{n}(\bar{X}-c)}{\sigma}\right) \geq \alpha$ then we shall accept that $\mu = c$ & not reject the null-hypothesis

Example 4

(a) $X \sim \text{Normal}(\mu, 9)$. $X_1, \ldots, X_{16}$ iid sample from X

$\bar{X} = 10.2$

$H_0: \mu = 9.5$

$H_1: \mu > 9.5$

$$\boxed{\begin{array}{l} n = 16, \ \sigma = 3 \quad c = 9.5 \\ \text{Fix } \alpha = 0.05 \end{array}}$$

Check $P\left(Z > \frac{\sqrt{n}(\bar{X} - c)}{\sigma}\right) < \alpha$ or not

when $Z \sim \text{Normal}(0,1)$

Find: $\mathbb{P}\left(Z > \frac{4(10.2 - 9.5)}{3} \right)$

$= \mathbb{P}\left(Z > \frac{2.8}{3} \right) \cong 0.175$

As $0.175 > 0.05 = \alpha$, the null hypothesis is not rejected.

---

4 (b) $H_0: \mu = 8.5$ $\bar{X} = 10.2, n = 16, \sigma = 3, c = 8.5$ $\alpha = 0.05$

$H_1: \mu > 8.5$ Compute $\mathbb{P}\left(Z > \frac{\sqrt{n}(\bar{X} - c)}{\sigma} \right)$

$$= P / 2 > \frac{4}{3}(10.2 - 8.5) \simeq 0.0117$$

As $0.0117 < 0.05$, we reject the null hypothesis & we reach the conclusion that $\mu > 8.5$

Another description :- If $\mu = 8.5$ assumption were true, the sampling procedure will produce a result as large a sample mean $\bar{X} = 10.2$ only 1.17 % of the time. This is so rare that we reject the null hypothesis that $\mu = 8.5$

4 (c) $X \sim \text{Normal}(\mu, 36)$ $\quad \bar{X} = 6.2 \quad n = 25$

$\alpha = 0.05$

$H_0 : \mu = 4$

$H_1 \quad \mu \neq 4$

Compute $\mathbb{P}\left(|\bar{Y} - 4| > |\bar{x} - 4| \right) = \mathbb{P}\left(|Z| > \frac{\sqrt{n}\,|\bar{x} - 4|}{\sigma} \right)$

$= P\left(|Z| > \frac{5}{6}(6.2 - 4) \right)$

$= P\left(|Z| > \cdot \right)$

$\cong 0.0668$

So as $0.0668 > 0.05$ we do not reject the null hypothes