Lecture 30

Wednesday, July 26, 2017 1:00 PM

Result 4 is true even if $X_1, \ldots, X_n$ are NOT indep.

e.g. $X \sim N(\mu, \sigma^2)$

$X_1 = X$, $X_2 = -X$

$X_1 \sim N(\mu, \sigma^2)$

$X_2 \sim N(-\mu, \sigma^2)$

$X_1 + X_2 = 0$ w.p. 1

Result 4 (with $a_1 = a_2 = 1$) $\Rightarrow X_1 + X_2 \sim N\Big(\underbrace{\mu_1+\mu_2}_{\mu-\mu=0},\ \underbrace{\sigma_1^2+\sigma_2^2+2Cov(X_1,X_2)}_{\sigma^2+\sigma^2+2\,cov(X,-X)\ =\ 2\sigma^2-2\sigma^2=0}\Big) = N(0,0)$ → This means $(X_1+X_2) = 0$ w.p. 1 (which we already saw)

$cov(X,-X) = E[(X-\mu)(-X+\mu)]$
$= -E(X-\mu)^2$
$= -\sigma^2$

- **Sec 5.3**: Using samples to estimate unknown popl^n parameters, e.g. μ, σ, $Cov(X,Y)$, $\rho_{X,Y}$ etc.

  We use sample quantities like $\bar{X}$, s_X, $Cov(x,y)$, $r_{x,y}$ etc.
  We call them 'statistic's.

  Inference: Using statistics to find info. on parameters.

  Dist^n of Sample Statistics: Sampling Dist^n (Depends on sample size n, sampling scheme and on the popl^n dist^n. as $n \to \infty$, dist^n becomes more accurate)

Inference
- Estimation
  - Point est. (Ch. 6)
  - Confidence Interval (Ch. 7)
- Hypothesis Testing (Ch 8, 9)

- **Sec 5.4**: $X_1, X_2, \ldots, X_n$ form a simple random sample (SRS) of size n, if (i) X_i's are indep. and (ii) Every X_i has the same dist^n } X_i's are iid.

Sample Mean: $\bar{X} = \dfrac{X_1 + \ldots + X_n}{n}$

$= \left(\frac{1}{n}\right)X_1 + \cdots + \left(\frac{1}{n}\right)X_n$: a linear combination of the X_i's. (with $a_1 = \ldots = a_n = \frac{1}{n}$)

Assumption: Popl^n dist^n (i.e. the dist^n of X_1, or any X_i) has mean μ and SD σ.

1. $E(\bar{X}) = \left(\frac{1}{n}\right).E(X_1) + \ldots + \left(\frac{1}{n}\right)E(X_n)$ (By result 1 of Sec 5.5)
$= \left(\frac{1}{n}\right).\mu + \ldots + \left(\frac{1}{n}\right).\mu = \mu$

i.e. $\mu_{\bar{X}} = \mu$ (Mean of the sampling dist^n of the sample mean is the popl^n mean)

2. $V(\overline{X}) = \left(\frac{1}{n}\right)^2 \cdot V(X_1) + \dots + \left(\frac{1}{n}\right) \cdot V(X_n)$ (By result 2 of Sec 5.5)

$$= \left(\frac{1}{n}\right)^2 \cdot \sigma^2 \cdot n = \frac{\sigma^2}{n}.$$

i.e. $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$. (As $\lim_{n \to \infty} \frac{\sigma}{\sqrt{n}} = 0$, the sampling dist^n of $\overline{X}$ becomes more concentrated around μ, i.e. it becomes "more accurate" as $n \to \infty$)

Now, if X_i's are iid $N(\mu, \sigma^2)$,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$ (By result 4 of Sec 5.5)

Similarly,

$$T_0 = X_1 + \dots + X_n \sim N(n\mu, n\sigma^2)$$ (It's another linear combination with $a_1 = a_2 = \dots = a_n = 1$)

Even if X_i's have non-Normal dist^n,

$$\overline{X} \overset{\text{approx.}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \text{ as } n \to \infty.$$ (This is called the Central Limit Theorem, or CLT)

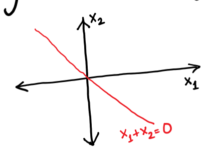
Lecture 31

Thursday, July 27, 2017 1:00 PM

Two conts. Variables may NOT share a joint density.

e.g. $X \sim N(\mu, \sigma^2)$

$X_1 = X$
$X_2 = -X$ } Both are conts.



Prob. of a line on the X_1-X_2 plane is 0 here which cannot be true if X_1 and X_2 have a joint density

So, $P(X_1 + X_2 = 0) = 1$

Eg. Weights of bags of a product $\overset{iid}{\sim} N(5, 1.5^2)$

Sample of 25 such bags : $\bar{X}_{25} \sim N\left(5, \frac{1.5^2}{25}\right)$

$$P(4.7 < \bar{X}_{25} < 5.3) = P\left(\frac{4.7-5}{\left(\frac{1.5}{5}\right)} < \frac{\bar{X}_{25}-5}{\left(\frac{1.5}{5}\right)} < \frac{5.3-5}{\left(\frac{1.5}{5}\right)}\right)$$

$$= P(-1 < Z < 1) = 0.6826 \approx 68\%.$$

CLT: $X_1, \ldots, X_n$ is a SRS from a dist^n with mean μ & SD σ. Then if n is sufficiently large, $\bar{X} \overset{\text{approx.}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$.

Thumb Rule: Apply CLT if $n > 30$.

- Ch 6 : (Point Estimation)

θ : parameter of interest (unknown)

$\hat{\theta}$: Sensible value for θ based on the sample — statistic / "Estimator" in this context

e.g.

Popl^n Parameters (θ)	Sample Estimators ($\hat{\theta}$)
Mean: μ	$\bar{X}$
Median: $\tilde{\mu}$	$\tilde{X}$
SD: σ	S
Proportion: p	$\hat{p}$

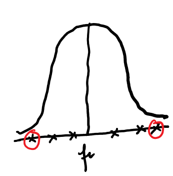
Eg. 1. $X \sim Bin(n, p)$

$\hat{p} = \frac{X}{n}$: Sample Proportion of S's.

2. $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\hat{\mu} = \begin{cases} \bar{X} \\ \tilde{X} \\ \frac{X_{min} + X_{max}}{2} \end{cases}$$

These are 3 different estimators



$\hat{\theta} = \theta + \text{error}$

Unbiased : $E(\text{error}) = 0 \Leftrightarrow E(\hat{\theta}) = \theta$ (No systematic error)

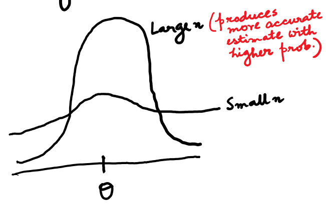
Eg 1. $\hat{p} = \frac{X}{n} \Rightarrow E(\hat{p}) = \frac{1}{n}\cdot E(X) = \frac{np}{n} = p$: Unbiased

2. $X_1, \dots, X_n \overset{iid}{\sim}$ Dist^n with mean μ

$E(\bar{X}) = \mu$: Unbiased

<u>Prop</u>: If a dist^n is conts. & symm. around μ, then $\tilde{X}$ and $\frac{X_{min} + X_{max}}{2}$ are unbiased for μ.

Among all unbiased estimators choose the one with the minimum Variance.



$Var(\hat{\theta})$

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

$= E(\text{error})^2$: Mean Squared error (MSE)

SD(error) $= \sqrt{MSE}$: Standard error of $\hat{\theta}$.

We pick the est. with smallest Var./MSE/SD

↳ Min. Var. Unb. Est. (MVUE)

<u>Prop</u>: 1. If $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\bar{X}$ is MVUE for μ.

2. If $X_1, \dots, X_n \overset{iid}{\sim} U(\mu - c, \mu + c)$, then $\frac{X_{min} + X_{max}}{2}$ is MVUE for μ.

<u>Construct Point Estimates from a sample:</u>

1. Maximum Likelihood Est. (MLE)

$$\hat{\theta}_{MLE} = \arg\max_{\theta} f(x_1, \dots, x_n | \theta)$$

(If $X_1, \dots, X_n$ are iid with the density $g(x;\theta)$ or $g(x|\theta)$, then $f(x_1, \dots, x_n | \theta) = g(x_1|\theta).g(x_2|\theta).\dots.g(x_n|\theta)$)

All the marginal densities are the same, i.e $g(\cdot|\theta)$ by identicality

Joint density is a product of marginal densities by indep.

2. Method of Moments (MOM)

$X_1, \dots, X_n \overset{iid}{\sim} f(x|\theta)$

<u>K^{th} Moment</u>: $\underbrace{E(X^K)}_{\text{Pop}^n\ K^{th}\text{ moment}} \approx \underbrace{\frac{X_1^K + \dots + X_n^K}{n}}_{\text{Sample } K^{th}\text{ moment}} = \overline{X^K}$ (for a representative sample)

MOM: $g(\theta) = \overline{X^K}$ (Set them to be equal to form an eqn. for θ)

$\Rightarrow \hat{\theta}_{MOM} = g^{-1}(\overline{X^K})$.

If you have p unknown parameters, then use the first p moments.

Lecture 32

Friday, July 28, 2017 1:00 PM

MOM: $E(X^k) = \frac{X_1^k + \dots + X_n^k}{n}$: Solve eqn.(s) for θ.

Now if you have p parameters to estimate, write p eqn^s for $k = 1, 2, \dots, p$.

Eg 1. $X_1, X_2, \dots, X_n \overset{iid}{\sim} Exp(\lambda)$

$\mu = E(X) = \frac{1}{\lambda}$.

MOM: $\theta = \lambda \to E(X) = \frac{X_1 + \dots + X_n}{n} = \bar{X}$

$\Leftrightarrow \frac{1}{\lambda} = \bar{X}$

$\Leftrightarrow \hat{\lambda} = \frac{1}{\bar{X}}$

Here, $E(\bar{X}) = \mu = \frac{1}{\lambda}$.

But, $E\left(\frac{1}{\bar{X}}\right)$ is not equal to $\underset{\text{Here}}{\frac{1}{\mu}}(=\lambda)$ in general.

i.e. $\hat{\lambda}$ is not unbiased for λ. [In fact, here $E\left(\frac{1}{\bar{X}}\right) = \frac{n\lambda}{n-1} \neq \lambda$

But, $E\left(\frac{1}{\bar{X}}\right) \to \lambda$ as $n \to \infty$ ($\because \lim_{n\to\infty} \frac{n}{n-1} = 1$)

i.e. $E(\hat{\lambda}) \to \lambda$ and equivalently

Bias$(\hat{\lambda}) \to 0$ as $n \to \infty$

This is called an "asymptotically unbiased estimator"]

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

2. $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$\theta = (\mu, \sigma^2)$

MOM: $E(X) = \frac{X_1 + \dots + X_n}{n} = \bar{X} \Leftrightarrow \mu = \bar{X}$

$E(X^2) = \frac{X_1^2 + \dots + X_n^2}{n} = \overline{X^2} \Leftrightarrow \sigma^2 + \mu^2 = \overline{X^2}$

$\Bigg\}$ $\hat{\mu} = \bar{X}$

$\hat{\sigma}^2 = \overline{X^2} - \hat{\mu}^2 = \overline{X^2} - \bar{X}^2 \underset{\uparrow \text{ Verify}}{=} S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$

- **Ch 7: (Confidence Intervals)**

Unknown parameter of interest: θ

Using data $X_1, \dots, X_n$, we want to find $l(x_1, \dots, x_n)$ and $u(x_1, \dots, x_n)$ s.t. we can say with some confidence that:

$$l(x_1, \dots, x_n) \le \theta \le u(x_1, \dots, x_n)$$

e.g. If $X \sim Bin(n, p)$, then with 100% confidence,

$0 \le p \le 1 \to 100\%$ CI for p : $[0, 1]$

Similarly, for any θ, 100% CI $= \mathbb{R}$

In most cases, CI for θ : $[l(x_1, \dots, x_n), u(x_1, \dots, x_n)] = [\underset{\text{Point est. of }\theta}{\hat{\theta}} - m, \hat{\theta} + \underset{\text{Margin of error}}{m}]$

(When the dist^n of $\hat{\theta}$ is symmetric about θ, this is the "best" CI)

Eg. $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ (σ^2 is known)

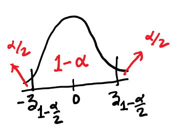
Take $\theta = \mu$

Our goal: Find m such that $P(\underset{\bar{X}}{\hat{\mu}} - m \le \mu \le \underset{\bar{X}}{\hat{\mu}} + m) = 1 - \alpha$ (α is some small value)

(We know that the dist^n of $\bar{X}$ is symmetric about μ)

We know: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

So, $P\Bigg(\underbrace{-z_{1-\frac{\alpha}{2}} \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le z_{1-\frac{\alpha}{2}}}_{\text{Rewrite the event as } l(\bar{x}) \le \mu \le u(\bar{x})}\Bigg) = 1 - \alpha$



Now, $-z_{1-\frac{\alpha}{2}} \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le z_{1-\frac{\alpha}{2}} \Leftrightarrow z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \ge \mu - \bar{X} \ge -z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ (Multiplying by $-\frac{\sigma}{\sqrt{n}}$)

$\Leftrightarrow \underbrace{\bar{X} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}}_{u(\bar{x})} \ge \mu \ge \underbrace{\bar{X} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}}_{l(\bar{x})}$ (Adding $\bar{X}$)

So, $\underbrace{100(1-\alpha)\%}_{\text{Confidence level / Confidence coeff.}}$ CI for μ : $\boxed{\left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right]}$

Some popular Values:

90% CI $\to \alpha = 0.1 \to z_{.95} = 1.645$

95% CI $\to \alpha = 0.05 \to z_{.975} = 1.96$

99% CI $\to \alpha = 0.01 \to z_{.995} = 2.576$

$95\% \ \ldots \ .975$

99% CI $\to \alpha = 0.01 \to z_{.995} = 2.576$

• Ch 8 : (Hypothesis Testing)

H_0 : Null hypothesis — Initial claim assumed to be true / the prior belief

H_1 / H_a : Alternative Hypothesis — Claim / belief that is contrary to H_0.

Test of hypothesis : To decide, based on sample info, which of H_0 or H_1 is correct.

H_0 is more favored to begin with (Protected hypo.)

Decisions : $\underbrace{\text{Reject } H_0}_{\text{If there is strong evidence in the data against } H_0}$ OR $\underbrace{\text{don't reject } H_0}_{\text{If there is not strong enough evidence in the data against } H_0}$.

Some known number

Eg. 1. $\underbrace{H_0 : \mu = \mu_0}_{\text{Simple null}}$ vs. $H_1 : \underbrace{\mu \neq \mu_0}_{\text{2-Sided Test}} \ | \ \mu > \mu_0 \ | \ \mu < \mu_0$

1-Sided Test

We'll work with this only → Simple null

2. $\underbrace{H_0 : \mu > \mu_0}_{\text{Composite null}}$ vs. $H_1 : \mu \leq \mu_0$

$X_1, \ldots, X_n$: Sample

Test Procedure :

① Choose a suitable Test Statistic (TS)

② A rejection region (RR) for the TS (If the value of the TS falls in the RR, we reject H_0. Otherwise, we don't reject H_0.)

Eg. $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ (σ^2 known)

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$

TS : $\bar{X}$

RR : $\{\bar{x} > c\}$ where $c > \mu_0$.

How do we find c ?

We use the distⁿ of $\bar{X}$ to find c just like finding n for CI's.

For that we look at the possible errors we can make in hypothesis testing.

Errors :

Truth \ Decision	Don't reject H_0	Reject H_0
H_0	OK	Type-I error
H_1	Type-II error	OK

More serious as H_0 is the protected hypothesis

We find c by making $P(\text{Type-I error}) = \alpha$ (Some small value)

So $P(\text{Type-I error})$ has a similar role in determining the test (through c) to the role the confidence level had in determining the CI (through n).

... the length of the CI,

So P(Type-I error) has a similar role in determining the test (through m). role the confidence level had in determining the CI

Moreover just like increasing the confidence level has a negative effect on the length of the CI, P(Type-I error) has a negative relation with P(Type-II error). Both cannot be made arbitrarily small. If you decrease one, the other will increase.

Since Type-I error is more serious, we like to make sure first that P(Type-I error) is small, say α.

Now in the test of $H_0: \mu = \mu_0$ vs. $H_1: \mu > \mu_0$,
Let us rewrite the RR: $\{\bar{X} > c\}$ for some $c > \mu_0$ as $\{\bar{X} - \mu_0 > c_1\}$ for some $c_1 > 0$.

Now, $P(\text{Type-I error}) = P(\text{Reject } H_0 \mid H_0 \text{ is true}) = P(\text{TS falls in RR} \mid H_0 \text{ is true})$

$\therefore$ In this example,

$$
\begin{aligned}
P(\text{Type-I error}) &= P(\bar{X} - \mu_0 > c_1 \mid \mu = \mu_0) \\
&= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c_1}{\sigma/\sqrt{n}} \,\middle|\, \mu = \mu_0\right) \\
&= P\left(Z > \frac{c_1}{\sigma/\sqrt{n}}\right) \quad \left[\because \text{ If } \mu = \mu_0, \ \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)\right] \\
&= 1 - \Phi\left(\frac{c_1}{\sigma/\sqrt{n}}\right)
\end{aligned}
$$

We set this to α: $1 - \Phi\left(\frac{c_1}{\sigma/\sqrt{n}}\right) = \alpha$

$$\Rightarrow \Phi\left(\frac{c_1}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

$$\Rightarrow \frac{c_1}{\sigma/\sqrt{n}} = z_{1-\alpha} \Rightarrow c_1 = \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$$

Similarly,
For $H_0: \mu = \mu_0$ vs. $H_1: \mu < \mu_0$, we take TS: $\bar{X}$ and RR: $\{\bar{X} < c\}$ for some $c < \mu_0$; OR equivalently $\{\bar{X} - \mu_0 < -c_2\}$ for some $c_2 > 0$.
For $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$, we take TS: $\bar{X}$ and RR: $\{\bar{X} > c\} \cup \{\bar{X} < c'\}$ for some $c > \mu_0$ and $c' < \mu_0$; OR equivalently $\{|\bar{X} - \mu_0| > c_3\}$ for some $c_3 > 0$.
(As dist^n of $\bar{X}$ is symmetric about μ)

Similar calculations as above lead to: $c_2 = \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$.
$c_3 = \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$.

(c_1, c_2 and c_3 are called 'Critical Values')

If we use the equivalent TS: $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$
We can rewrite the rejection regions for the three tests: (for some α)

1. $H_0: \mu = \mu_0$ vs. $H_1: \mu > \mu_0$
2. $H_0: \mu = \mu_0$ vs. $H_1: \mu < \mu_0$

and 3. $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$

as

RR: 1. $\{Z > z_{1-\alpha}\}$
2. $\{Z < -z_{1-\alpha}\}$
and 3. $\{|Z| > z_{1-\frac{\alpha}{2}}\} = \{Z > z_{1-\frac{\alpha}{2}}\} \cup \{Z < -z_{1-\frac{\alpha}{2}}\}$.

α is called the 'Significance Level' of the test.
In a similar manner, P(Type-II error) is denoted by β.

Unlike α, there is no single value for β as β depends on the 'true' value of θ under H_1. Unlike a simple null hypothesis, the alternative hypothesis is a collection of a range of θ values, each of which produces a different β value. So we write β as: $\underbrace{\beta(\theta)}_{\text{a function of alternative } \theta \text{ values.}}$.

$\therefore \beta = P(\text{Type-II error}) = P(\text{Don't reject } H_0 \mid H_0 \text{ false})$
So that $1 - \beta = P(\text{Reject } H_0 \mid H_0 \text{ false})$ is called the 'Power' of the test.

Clearly, we want our test to have as much power as possible.
But as we have already discussed, α and β cannot be made arbitrarily small at the same time.
Since our first priority is making α small, we fix α at a predetermined small level, eg. 0·1, 0·05 or 0·01 and try to make β as small as possible.

Lecture 33

Monday, July 31, 2017 1:00 PM

Eg. A manufacturer of sprinkler systems claims that the activation times are $N(\mu, (1.5)^2)$. They also claim that $\mu = 130$ (°F). A SRS of size $n=9$ yields $\bar{x}_9 = 131.08$.

Find out a 99% CI for the unknown population mean μ of the activation times.

99% CI for μ: $\left(\bar{x}_9 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}, \bar{x}_9 + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) = \left(131.08 - \frac{1.5}{3} \times 2.576, 131.08 + \frac{1.5}{3} \times 2.576\right)$

$= (129.792, 132.368)$

Perform a hypothesis test at 1% level of significance to conclude if the manufacturer's claim is correct.

Test: $H_0: \mu = 130$ vs. $H_a: \mu \neq 130$

TS: $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{131.08 - 130}{1.5/3} = 2.16$

RR: $\{|Z| > z_{1-\frac{\alpha}{2}}\} = \{|Z| > 2.576\}$ ($\alpha = 0.01 \rightarrow$ Test with 1% level of significance)

Clearly, $|Z| = 2.16 < 2.576$.

We don't reject H_0. (as TS is NOT in RR)

Conclusion: There is NOT enough evidence in the data to say that the manufacturer's claim is incorrect.

Steps: (For performing a test of hypotheses)

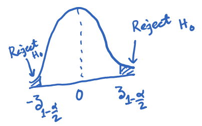
1. Formulate the problem and identify the parameter(s).
2. Determine H_0 and H_a.
3. Choose TS and find its dist^n under H_0.
4. Find the RR based on the level of significance.
5. Calculate the value of the TS from the sample.
6. Make a decision and state the conclusion.

- **One-sample Z-test and CI:**

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ (σ^2 Known)

$H_0: \mu = \mu_0$ vs. $H_a: \mu \neq \mu_0$

$Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$; RR: $\{|Z| > z_{1-\frac{\alpha}{2}}\}$ (Level-α Test)



CI for μ: $\bar{x} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$ ($100(1-\alpha)\%$ CI)

H_0 is NOT rejected $\Leftrightarrow |z| \leq z_{1-\frac{\alpha}{2}} \Leftrightarrow \left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| \leq z_{1-\frac{\alpha}{2}} \Leftrightarrow |\mu_0 - \bar{x}| \leq \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$

$\Leftrightarrow \mu_0 \in \underbrace{\left[\bar{x} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right]}_{100(1-\alpha)\%\text{ CI}}$

Level-α 2-Sided Test rejects H_0 iff $\mu_0 \notin 100(1-\alpha)\%$ CI

- **Length of a CI:**

For a $100(1-\alpha)\%$ CI, Width(w) $= 2m$ ← margin of error

$= 2 \underbrace{\frac{\sigma}{\sqrt{n}}}_{SE(\bar{X})} \cdot \underbrace{z_{1-\frac{\alpha}{2}}}_{\text{Critical Value}}$
(↑ if σ ↑ → more noisy data, less precision
↓ if n ↑ → bigger sample size, more precision
↓ if α ↑) → Less reliability, more precision

In general, (often) margin of error (m) of a 2-sided CI for θ $= SE(\hat{\theta}) \times$ Critical Value (where, $SE(\hat{\theta})$ is standard error of $\hat{\theta}$ and critical value comes from the dist^n of $\hat{\theta}$)

Width : Precision / accuracy of the CI
$100(1-\alpha)\%$: Confidence / Reliability of the CI

We need $100(1-\alpha)\%$ confidence and width $\leq w_0$. What's the minimum sample size required to achieve that?

$w = \frac{2\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} \leq w_0$.

$\Leftrightarrow \sqrt{n} \geq \dfrac{2\sigma z_{1-\frac{\alpha}{2}}}{w_0}$

$\Leftrightarrow n \geq \underbrace{\left[\dfrac{2\sigma z_{1-\frac{\alpha}{2}}}{w_0}\right]^2}_{\text{Go to the next integer, if fraction.}} = n_{min}$

- **Large Sample Z-test and CI:**

$X_1, \dots, X_n$: a SRS from some dist^n (Possibly non-Normal) but with finite μ and σ. (Unknown)

Large Sample : ($n > 30$) $\bar{X} \underset{\text{approx.}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ (By CLT)

(Now if popl^n dist^n is non-Normal but σ is known, One-sample Z-test and CI formula can still be used approximately — because of the above result by CLT)

Since σ is unknown, replace it by S, the sample SD.

$\left[S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$

Then the same formulae will work, as $S \to \sigma$ as $n \to \infty$. ($100(1-\alpha)\%$ CI: $\bar{x} \pm \frac{s}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$; TS: $Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$; RR same as before)

But we'll now need: $n > 40$ [10 more obs^{ns} needed so that $S \approx \sigma$]

- If σ is unknown and n is small (We need normality) → Otherwise knowledge about the popl^n dist^n will have to be used, e.g. if data $X \sim Bin(n,p)$ and you want CI or test for p)

$X_1, \dots, X_n$: SRS from $N(\mu, \underbrace{\sigma^2}_{\text{unknown}})$; $\underbrace{n \text{ is small}}_{\text{Can't use CLT.}}$.

We replace σ by S.

TS: $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$

(We can't use $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ as σ is unknown. T has more variability than Z because of the r.v. S in the denominator. Variability in Z comes only from $\bar{X}$ in the numerator but its denominator is only a constant)

It's more spread out!

In general, $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$ is no longer a $N(0,1)$ Variable. ... it's t-dist^n)

$\sigma/\sqrt{n}$ comes ...

In general, $T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ is no longer a $N(0,1)$ Variable. It's more spread out!

In fact, $T \sim T_{(n-1)}$: t-distⁿ with $(n-1)$ d.f. (Also called Student's t-distⁿ)

d.f.: degrees of freedom

Properties of the t-distⁿ:

- t-distⁿ is Bell-shaped, symmetric about 0 and unimodal (just like $N(0,1)$).
- It has higher variance, i.e. heavier tails and shorter peak compared to $N(0,1)$.
  As a result, the upper percentiles of it: t_α's are bigger than those of $N(0,1)$, i.e. $z_{1-\alpha}$'s. (The book calls these as z_α's.)
  $P(T_{(f)}>k) > P(Z>k)$ for any $k>0$ and for all f.
- As the d.f. increases, spread/variability of t-distⁿ decreases.
- As the d.f. $\uparrow \infty$, the t-curve approaches that of $N(0,1)$.
  [We know: $T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \to Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ as $n\to\infty$, because $\lim_{n\to\infty} S=\sigma$
  So the distⁿ of T approaches the distⁿ of Z as $n\to\infty$]

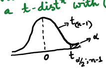
So for the CI and tests we replace σ by S and get the critical values t_α's from the t-table instead of the $z_{1-\alpha}$'s. These are called One-Sample t-test and CI.

$$100(1-\alpha)\%\ \text{CI}: \ \bar{X} \pm t_{\alpha/2;n-1}\cdot\frac{S}{\sqrt{n}}$$

Tests for $H_0: \mu=\mu_0$ vs. 1. $H_a: \mu>\mu_0$
2. $H_a: \mu<\mu_0$
3. $H_a: \mu\neq\mu_0$

use TS: $T=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}$ and RR: 1. $T>t_{\alpha;n-1}$
2. $T<-t_{\alpha;n-1}$
3. $|T|>t_{\alpha/2;n-1}$

Critical Value: $100(1-\alpha)^{th}$ percentile of a t-distⁿ with $(n-1)$ d.f.



Always round DOWN to the next integer when your df ($=n-1$) is not shown in the t-table!

Eg. A manufacturer of sprinkler systems claims that the activation times are $N(\mu,\sigma^2)$. σ is unknown. They also claim that $\mu=130$ (°F). A SRS of size $n=9$ yields $\bar{x}_9=131.08$ and $s=1.5$.

Find out a 99% CI for the unknown population mean μ of the activation times.

99% CI for μ: $\left(\bar{X}_n-\dfrac{S}{\sqrt{n}}\cdot t_{\alpha/2;n-1},\ \bar{X}_n+\dfrac{S}{\sqrt{n}}\cdot t_{\alpha/2;n-1}\right) = \left(131.08-\dfrac{1.5}{3}\times 3.355,\ 131.08+\dfrac{1.5}{3}\times 3.355\right)$

$= (129.403,\ 132.758)$

Perform a hypothesis test at 1% level of significance to conclude if the manufacturer's claim is correct.

Test: $H_0: \mu=130$ vs. $H_a: \mu\neq 130$

TS: $T=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}} = \dfrac{131.08-130}{1.5/3} = 2.16$

RR: $\{|T|>t_{\frac{\alpha}{2};n-1}\} = \{|T|>3.355\}$ ($\alpha=0.01 \to$ Test with 1% level of significance)

Clearly, $|T|=2.16<3.355$.

We don't reject H_0. (as TS is NOT in RR)

Conclusion: There is NOT enough evidence in the data to say that the manufacturer's claim is incorrect.

One-Sample t-test and CI share the same kind of relation that One-Sample Z-test and CI share.

e.g. In the above example $\mu_0=130 \in 99\%$ CI $\Leftrightarrow$ We don't reject H_0 in the 1% level test.

- Interpretation of confidence level of a CI:

In the previous example, saying
$P(\mu\in(129.403,\ 132.758))=0.99$ is wrong!

$P(\mu\in(\bar{X}-\frac{S}{\sqrt{n}}t_{.005;n-1},\ \bar{X}+\frac{S}{\sqrt{n}}t_{.005;n-1}))=0.99.$

In the previous example, saying

$P(\mu \in (129.403, 132.758)) = 0.99$ is wrong!

Even though $P\left(\mu \in \left(\bar{X} - \frac{S}{\sqrt{n}} t_{.005;n-1}, \bar{X} + \frac{S}{\sqrt{n}} t_{.005;n-1}\right)\right) = 0.99$.

That's because the randomness comes from $\bar{X}$, which is gone as soon as $\bar{X}$ is observed.

Based on the unknown value of μ, $P(\mu \in (129.403, 132.758))$ is 0 or 1.

If we repeat the sampling many times, the CI will keep on changing and roughly 99% of those CI's will contain the unknown μ.

That's why when we don't know μ, we can say that we have 99% confidence on this interval (129.403, 132.758). In reality, we have 99% confidence on the process that produced this CI, not any particular CI.

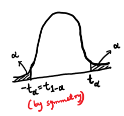
- One-sided intervals (Confidence bounds):

Leave entire α-probability on one tail: (instead of $\alpha/2$ on each tail that 2-sided CI's do)

$P(\mu < u(x_1, \dots, x_n)) = 1-\alpha$ (upper confidence bound /UCB)

OR, $P(\mu > l(x_1, \dots, x_n)) = 1-\alpha$ (lower confidence bound /LCB)

We know: $P\left(\frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\alpha;n-1}\right) = 1-\alpha$



$\Leftrightarrow P\left(\mu > \bar{X} - t_{\alpha;n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha$

So, $100(1-\alpha)\%$ LCB: $\left(\bar{x} - t_{\alpha;n-1}\cdot\frac{s}{\sqrt{n}}, \infty\right)$

Similarly, $100(1-\alpha)\%$ UCB: $\left(-\infty, \bar{x} + t_{\alpha;n-1}\frac{s}{\sqrt{n}}\right)$

These have similar relation to one-sided tests that CI's share with two-sided tests.

Similar CB's can be found with $z_{1-\alpha}$'s when σ is known.

- How to determine which CI or test you should use:

Look at three things: Sample size (n), Popl^n. SD (σ) and normality of the popl^n dist^n.

Sample size (n)
- Large
  - σ known
    - Normal dist^n: Exact Z
    - Non-Normal dist^n: Approx. Z ($n > 30$, by CLT)
  - σ unknown
    - Normal dist^n: Exact-T / Approx. Z (For large n, $T_{(n-1)} \approx Z$)
    - Non-Normal dist^n: Approx. Z ($n > 40$)
- Small
  - σ known
    - Normal dist^n: Exact Z
    - Non-Normal dist^n: Need Popl^n dist^n
  - σ unknown
    - Normal dist^n: Exact T
    - Non-Normal dist^n: Need popl^n dist^n

Lecture 35

Tuesday, August 1, 2017 3:00 PM

$$f(x|\theta) = \begin{cases} (\theta+1)x^{\theta}, & \text{if } 0<x<1 \\ 0, & \text{o.w.} \end{cases}$$

Verify it's a pdf ($\theta > -1$)

$$E(X) = \int_0^1 x.(\theta+1).x^{\theta}\,dx$$

$$= (\theta+1)\left.\frac{x^{\theta+2}}{\theta+2}\right|_0^1 = \frac{\theta+1}{\theta+2}$$

MOM: $E(X) = \bar{X}$

$$\Rightarrow \frac{\theta+1}{\theta+2} = \bar{X}$$

$$\Rightarrow \hat{\theta} = \frac{2\bar{X}-1}{1-\bar{X}}.$$ (The MOM estimator of θ)

If your SRS from this distⁿ is $\{0.2, 0.3, 0.4\}$, then $\bar{X} = 0.3$

MOM estimate of θ: $\frac{2\times 0.3 - 1}{1-0.3} = \frac{-0.4}{-0.7} = \frac{4}{7}$

One-Sample Z-test and CI have been derived from

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \Leftrightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$SE(\bar{X}) \rightarrow \sigma/\sqrt{n}$

Similarly, if you can show for some θ,

$$\frac{\hat{\theta}-\theta}{SE(\hat{\theta})} \sim N(0,1)$$ possibly approx.

Then, similar steps as the One-sample Z-test and CI would give:

Possibly approx. $\rightarrow$ $100(1-\alpha)\%$ CI (for θ): $\left[\hat{\theta} - SE(\hat{\theta}).z_{1-\frac{\alpha}{2}},\ \hat{\theta} + SE(\hat{\theta}).z_{1-\frac{\alpha}{2}}\right]$

Level α-test for $H_0: \theta=\theta_0$ vs. 1. $H_a: \theta>\theta_0$ 2. $H_a: \theta<\theta_0$ 3. $H_a: \theta\neq\theta_0$: $TS = \frac{\hat{\theta}-\theta_0}{SE(\hat{\theta})}$ and $RR: \begin{cases} 1.\ Z > z_{1-\alpha} \\ 2.\ Z < -z_{1-\alpha} \\ 3.\ |Z| > z_{1-\frac{\alpha}{2}} \end{cases}$

If $SE(\hat{\theta})$ involves θ, then plug-in the estimate of θ to get $\hat{SE}(\hat{\theta})$. If $SE(\hat{\theta})$ involves other unknown parameters, estimate $\hat{SE}(\hat{\theta})$ can still be found, possibly using S.

Distⁿ of $\frac{\hat{\theta}-\theta}{SE(\hat{\theta})}$ or $\frac{\hat{\theta}-\theta}{\hat{SE}(\hat{\theta})}$ can be something else.

e.g. for small n and unknown σ when popl distⁿ is Normal, $SE(\hat{\mu}) = SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ is unknown, so $\hat{SE}(\bar{X}) = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}}$ gives

$$T = \frac{\bar{X}-\mu}{\hat{SE}(\bar{X})} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

Following similar logic as above, we get One-Sample t-test and CI

- Power Calculation for One-Sample Z-test:

$$H_0: \mu=\mu_0 \text{ vs. } H_a: \mu>\mu_0$$

$RR: \{Z > z_{1-\alpha}\}$ where $TS: Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

$$\beta(\mu_a) = P(\text{Type-II error}) \quad (\text{at some } \mu_a > \mu_0)$$

$$= P(\text{Don't reject } H_0 \mid H_0 \text{ is false})$$

Clearly, $\beta(\mu_0) = 0$ as it's impossible to make a Type-II error if $\mu = \mu_0$.

$$= P\left(Z \le z_{1-\alpha} \,\middle|\, \mu = \mu_a\right)$$

$$= P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \le z_{1-\alpha} \,\middle|\, \mu = \mu_a\right)$$

$$= P\left(\frac{\bar{X}-\mu_a}{\sigma/\sqrt{n}} \le z_{1-\alpha} + \frac{\mu_0-\mu_a}{\sigma/\sqrt{n}} \,\middle|\, \mu = \mu_a\right)$$

$$= \Phi\left(z_{1-\alpha} + \frac{\mu_0-\mu_a}{\sigma/\sqrt{n}}\right)$$

Now, Power $= 1 - \beta(\mu_a)$.

Similarly with $H_a: \mu < \mu_0$, $\beta(\mu_a) = 1 - \Phi\left(-z_{1-\alpha} + \frac{\mu_0-\mu_a}{\sigma/\sqrt{n}}\right)$, at some $\mu_a < \mu_0$.

and with $H_a: \mu \ne \mu_0$, $\beta(\mu_a) = \Phi\left(z_{1-\alpha/2} + \frac{\mu_0-\mu_a}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{\mu_0-\mu_a}{\sigma/\sqrt{n}}\right)$, at some $\mu_a \ne \mu_0$.

These formulae are in Page 331 (ed. 9).

They also have formulae for the minimum sample size needed for a level α test to have $\beta(\mu_a) \le \beta$ for some small β and at some μ_a.

(Similar to n_{min} calculation for a $100(1-\alpha)\%$ CI to have width $\le \omega_0$)