

1/16/21

8. Statistical Inference

Estimation:

x_1, x_2, \dots, x_n statistic → Functions of sample observation

\bar{x}, s^2

Estimation is choose a good estimator to estimate unknown population parameter.

\bar{x} is estimator for μ .

$\bar{x} \rightarrow$ numerical

value is estimate

$$E(\bar{x}) = \mu$$

↓ estimator ↗ estimate

Point Estimation: single value as estimate

Interval Estimation: Range of values as estimate

$\mu \in (10, 20) \rightarrow$ Confidence interval

Methods for finding estimators:

Method of moments, Maximum likelihood estimation.

Properties of Estimators:

Unbiased Estimator: $T(x_1, x_2, \dots, x_n) = T(\bar{x})$ is unbiased estimator if $E(T(\bar{x})) = g(\theta)$

$T(\bar{x})$ is unbiased estimator of $g(\theta)$

$E(T(\bar{x})) = g(\theta) + b(\theta) \Rightarrow T(\bar{x})$ is biased estimator of $g(\theta)$

Ex. $X \sim \text{Bin}(n, p)$

$$E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{np}{n} = p = g(\theta)$$

$\frac{X}{n}$ is unbiased estimator of p .

$$E\left(\frac{X(X-1)}{n(n-1)}\right) = \frac{1}{n(n-1)} E(X(X-1))$$

$$= \frac{1}{n(n-1)} n(n-1)p^2$$

$$= p^2 = g(\theta)$$

$\frac{X(X-1)}{n(n-1)}$ is unbiased estimator of p^2 .

Ex. $x_1, x_2, \dots, x_n \sim \text{Poisson}(\lambda)$

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

$$E(\bar{x}) = \frac{1}{n} \cdot n\lambda = \lambda$$

\bar{x} is unbiased estimator of λ .

→ \bar{x} is unbiased estimator of μ ($E(\bar{x}) = \mu$)

s^2 is unbiased estimator of σ^2 ($E(s^2) = \sigma^2$)

Consistency: An estimator is said to be consistent estimator if

$$\lim_{n \rightarrow \infty} T(x_1, x_2, \dots, x_n) \rightarrow g(\theta)$$

$$\lim_{n \rightarrow \infty} P(|T(x_1, x_2, \dots, x_n) - g(\theta)| > \epsilon) \rightarrow 0$$

$$\underset{n \rightarrow \infty}{\text{LT}} P(|f(x) - g(x)| > \epsilon) \rightarrow 0$$

Ex: $x_1, x_2, \dots, x_n \quad E(x_i) = \theta, \quad \text{Var}(x_i) = \sigma^2$

\bar{x} is consistent for θ

$$P(|\bar{x} - \theta| > \epsilon) \leq \frac{\text{Var}(\bar{x})}{\epsilon^2}, \quad (\text{Chebyshev's Inequality})$$

$$\leq \frac{\sigma^2}{n\epsilon^2}$$

As $n \rightarrow \infty$, $\text{Var}(\bar{x}) \rightarrow 0$

\bar{x} is consistent estimator of θ

Ex: $T(x) = \frac{x_1 + x_2 + \dots + x_n}{n} = \bar{x}$ Normal distribution

$$\begin{aligned} P\{|T(x) - \mu| > \epsilon\} &= 2 P\{|\bar{x} - \mu| > \epsilon\} \\ &= 2 [1 - P\{|\bar{x} - \mu| \leq \epsilon\}] \\ &= 2 \left(1 - P\left\{\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \leq \frac{\epsilon\sqrt{n}}{\sigma}\right\}\right) \\ &= 2(1 - 1) \\ &= 0 \end{aligned}$$

Mean squared error:

$$\begin{aligned} \text{MSE} &= \frac{1}{n} \sum \left(\left(T(x_i) \overset{\text{actual value}}{\uparrow} - E(T(x_i)) \overset{\text{Expected}}{\uparrow} \right)^2 \right) \\ &= E(T(x_i) - E(T(x_i))^2 \end{aligned}$$

$\text{MSE} = \text{Var}(T(x))$
Smaller the value of variance of estimator better the estimate

Ex:

x_1, x_2, \dots, x_n - random sample

$$E(x_i) = \theta \quad \text{Var}(x_i) = \sigma^2$$

$$T_1(x) = \bar{x}, \quad T_2(x) = \tilde{x}$$

$$E(x_1) = \theta, \quad E(\bar{x}) = \theta \quad x_1, \bar{x} \text{ are unbiased}$$

$$\text{Var}(x_1) = \sigma^2, \quad \text{Var}(\bar{x}) = \frac{\sigma^2}{n}$$

$$\sigma^2 \geq \frac{\sigma^2}{n} \rightarrow \text{less variance}$$

\bar{x} is better estimator.

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

$$E\left(\frac{n-1}{n}s^2\right) = \frac{n-1}{n}\sigma^2$$

$$= \sigma^2 - \boxed{\frac{\sigma^2}{n}} \rightarrow \text{bias.}$$

$\frac{n-1}{n}s^2$ is biased estimator of σ^2 .

Method of Moments:

x_1, x_2, \dots, x_n from a population

$$\theta = (\theta_1, \theta_2, \dots, \theta_k)$$

parameter vector

Consider non central moments μ_j'

$$E(x_i^j) = \mu_j'$$

$$\mu_j' = \frac{1}{n} \sum x_i^j$$

↓
fixed probability

Each sample has
equal chance of choosing.

$$\mu_1' = g_1(\theta_1, \theta_2, \dots, \theta_k)$$

$$\mu_2' = g_2(\theta_1, \theta_2, \dots, \theta_k)$$

$$\vdots$$

$$\mu_k' = g_k(\theta_1, \theta_2, \dots, \theta_k)$$

Number of unknown parameters. Solve for $\theta_1, \theta_2, \dots, \theta_k$.

$$\theta_1 = h_1(\mu_1', \mu_2', \dots, \mu_k') \rightarrow \text{estimators}$$

$$\theta_2 = h_2(\mu_1', \mu_2', \dots, \mu_k')$$

$$\theta_k = h_k(\mu_1', \mu_2', \dots, \mu_k')$$

h_i is estimator for θ_i

Ex ① $X_1, X_2, \dots, X_n \sim \text{Poisson}(\lambda) \quad \lambda > 0$

unknown parameter

$$\mu_1' = \frac{1}{n} \sum x_i$$

$$= \frac{1}{n} \cdot n \lambda$$

$$\mu_1' = \lambda$$

estimator of λ is $\mu_1' = \frac{1}{n} \sum x_i = \bar{x}$

② $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$

unknown parameter

$$\mu_1' = \frac{1}{n} \sum x_i = \frac{1}{n} \cdot n\mu = \mu$$

$$\mu_2' = \sigma^2 + \mu^2$$

$$\text{Var}(x) = E(x^2) - (E(x))^2$$

$$\sigma^2 = E(x^2) - \mu^2$$

$$\mu_2' = \sigma^2 + \mu^2$$

$$\begin{aligned} \mu_1' &= \mu \\ \mu_2' &= \sigma^2 + \mu^2 \Rightarrow \mu_2' = \sigma^2 + (\mu_1')^2 \\ &\Rightarrow \sigma^2 = \mu_2' - (\mu_1')^2 \end{aligned}$$

So μ_1' is estimator for μ .

$$\begin{aligned} \mu_2' - (\mu_1')^2 &= \frac{1}{n} \sum x_i^2 - \left(\frac{1}{n} \sum x_i \right)^2 \\ &= \frac{1}{n} \sum x_i^2 - (\bar{x})^2 \text{ is estimator for } \sigma^2 \end{aligned}$$

→ Method of moments guarantees consistency but not unbiasedness.

Method of Likelihood Estimation: (discovered by R.A. Fisher)
 \downarrow
 F-dist)

→ Form of distribution.

x_1, x_2, \dots, x_n be the observed sample.

Joint probability pmf/pdf

$f(x_1, x_2, \dots, x_n, \theta)$ is called likelihood function

Independent
 $L(\theta, x) \quad x = (x_1, x_2, \dots, x_n)$
 \downarrow
 unknown parameter
 $f(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta); f(x_2, \theta) \dots f(x_n, \theta)$

$L(\theta, x) = \prod_{i=1}^n f(x_i, \theta)$ is likelihood function.

Most P.M.F's and P.Mf's are exponential in nature So we take log

$$l(\theta, x) = \log L(\theta, x) \rightarrow \text{linear function}$$

\downarrow
log likelihood function

→ A statistic $\Theta(x)$ or $T(x)$ is said to be more/mar likelyhood estimator of θ if

$$L(\Theta'(x), x) \geq L(\Theta, x) \text{ for } \forall \theta$$

Ex Let $X_1, X_2, \dots, X_n \sim \text{Ber}(1, P)$

$$L(P, x) = \prod_{i=1}^n P^{x_i} (1-P)^{n-x_i}$$

$$= P^{\sum x_i} (1-P)^{n-\sum x_i}$$

$$l(P, x) = \log L = \sum x_i \log P + (n - \sum x_i) \log (1-P)$$

$$\frac{\partial l}{\partial P} = \sum \frac{x_i}{P} + (n - \sum x_i) \frac{(-1)}{1-P}$$

$$= \frac{\sum x_i}{P} - \frac{(n - \sum x_i)}{1-P}$$

$$(1-P)\sum x_i - P(n - \sum x_i) = 0$$

$$(1-P)\sum x_i = P(n - \sum x_i)$$

$$\sum x_i - P\sum x_i = Pn - P\sum x_i$$

$$\sum x_i = nP$$

$$P = \frac{\sum x_i}{n} = \bar{x}$$

$P = \bar{x}$ is maximum likelihood estimator.

$$\text{Ex: } L(\lambda, x) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$= \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod_{i=1}^n (x_i!)}$$

$$\log L = -n\lambda + \sum x_i \log \lambda - \log(\prod_{i=1}^n (x_i!))$$

$$\frac{\partial L}{\partial \lambda} = -n + \frac{\sum x_i}{\lambda}$$

$$-n + \frac{\sum x_i}{\lambda} = 0$$

$$\lambda = \frac{\sum x_i}{n} = \bar{x}$$

$\lambda = \bar{x}$ is maximum likelihood estimator.

Confidence Interval: An interval is provided within the parameter is supposed to lie.

Ex: $X \sim N(\mu, \sigma^2)$, μ is unknown, σ is known.

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

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

$$P(-2.576 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 2.576) = 0.99$$

$$P\left(-2.576 \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq 2.576 \frac{\sigma}{\sqrt{n}}\right) = 0.99$$

$$P\left(-2.576 \frac{\sigma}{\sqrt{n}} - \bar{X} \leq -\mu \leq 2.576 \frac{\sigma}{\sqrt{n}} - \bar{X}\right) = 0.99$$

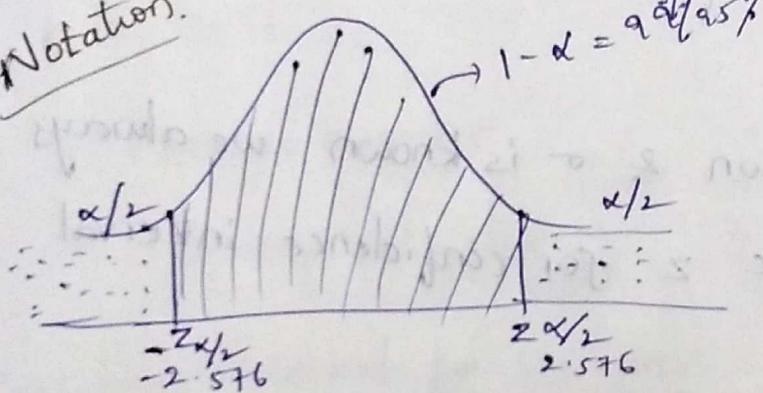
$$P\left(\bar{x} - 2.576 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 2.576 \frac{\sigma}{\sqrt{n}}\right) = 0.99$$

Population mean lies between $\bar{x} - 2.576 \frac{\sigma}{\sqrt{n}}$ and $\bar{x} + 2.576 \frac{\sigma}{\sqrt{n}}$ with confidence level 99%.

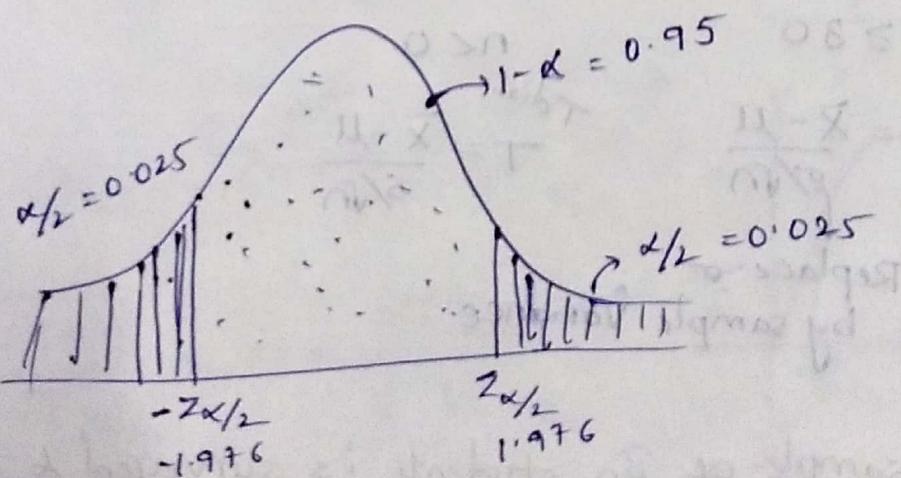
$(\bar{x} - z \frac{\sigma}{\sqrt{n}}, \bar{x} + z \frac{\sigma}{\sqrt{n}})$ is confidence interval in which μ is supposed to lie with 99% probability.

As you increase

23/6/21
Notation:



Confidence interval $(\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}})$



→ Bigger the confidence level wider the confidence interval.

→ Confidence level is $1-\alpha$.

Ex:

A random sample of 100 employees is taken & mean height 180cm & Variance of population is 49 cm^2 find 95% confidence interval for population.

Sol:

$$\mu \in \left(\bar{x} - z \frac{\sigma}{\sqrt{n}}, \bar{x} + z \frac{\sigma}{\sqrt{n}} \right)$$

$$\bar{x} = 180$$

$$\sigma^2 = 49$$

$$\mu \in \left(180 - 1.976 \cdot \left(\frac{7}{10} \right), 180 + 1.976 \cdot \left(\frac{7}{10} \right) \right)$$

→ μ is unknown & σ is known we always use statistic z for confidence interval for any n .

→ μ is unknown & σ is unknown

$$n \geq 30$$

$$\text{Normal } Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Replace σ by sample variance

$$n < 0$$

$$T \text{ dist} \quad T = \frac{\bar{x} - \mu}{S / \sqrt{n}}$$

Ex: A sample of 80 students is surveyed & Avg amount spent by them is 593.84 & S.D is 369.34. For 95% confidence level find confidence interval for mean.

$$\mu \in \left(593.84 \pm 1.976 \cdot \frac{369.34}{\sqrt{80}} \right)$$

A sample size 15 is taken from a population where $\bar{x} = 12$ & $s^2 = 25$ what is 95% confidence interval for μ ?

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

$$\text{Interval } \left(\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \right)$$

$$n - 1 = 15 - 1 = 14$$

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

$$t_{0.025, 14} = 2.145$$

$$\mu \in \left(12 \pm 2.145 \cdot \frac{5}{\sqrt{15}} \right)$$

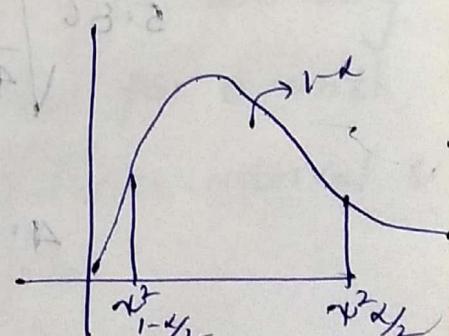
Confidence interval for Variance:

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

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

$$\frac{1}{\chi^2_{\alpha/2}} < \frac{\sigma^2}{(n-1)s^2} < \frac{1}{\chi^2_{1-\alpha/2}}$$

$$\sqrt{\frac{s^2(n-1)}{\chi^2_{\alpha/2}}} < \sigma < \sqrt{\frac{s^2(n-1)}{\chi^2_{1-\alpha/2}}}$$



Ex: A statistician chooses 27 dates when examining occupancy records of particular hotel finds s.D on 5.86 rooms rented. Find 95% confidence interval for population s.D.

sol

$$n = 27 \text{ s.d. not known} \quad \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$n-1 = 26$$

$$\sqrt{\frac{s^2(n-1)}{\chi^2_{\alpha/2}}} < \sigma < \sqrt{\frac{s^2(n-1)}{\chi^2_{1-\alpha/2}}} \quad 1-\alpha = 0.95 \\ \alpha = 0.05 \\ \frac{\alpha}{2} = 0.025$$

$$\sqrt{\frac{(5.86)^2(26)}{\chi^2_{0.025, 26}}} < \sigma < \sqrt{\frac{(5.86)^2(26)}{\chi^2_{0.975, 26}}}$$

$$\chi^2_{0.025, 26} = 41.923$$

$$\chi^2_{0.975, 26} = 13.844$$

$$5.86 \sqrt{\frac{26}{41.923}} < \sigma < \sqrt{\frac{(5.86)^2 26}{13.844}}$$

$$4.614 < \sigma < 8.0307$$

Test of Hypothesis: There will be two claims about values of parameter & determine which is correct.

→ Decision making problem

$H_0: \mu = 50$ ✓
 $H_1: \mu \neq 50$ ✗
mutually exclusive

Statistical Hypothesis:

Statement about the population.

Null Hypothesis: (H_0)

Prediction is not true.

Alternative Hypothesis: (H_A or H_1).

Prediction to be true.

25/6/21

Confidence Interval: Let x_1, x_2, \dots, x_n be the sample observation θ be the unknown parameter of population if there exist statistic L_n & U_n such

that $P(L_n < \theta < U_n) = 1 - \alpha$ then we say

(L_n, U_n) as confidence interval for θ with

$\alpha(100-\alpha)\%$, $100(1-\alpha)\%$ confidence interval &

$(1-\alpha)$ is called confidence level.

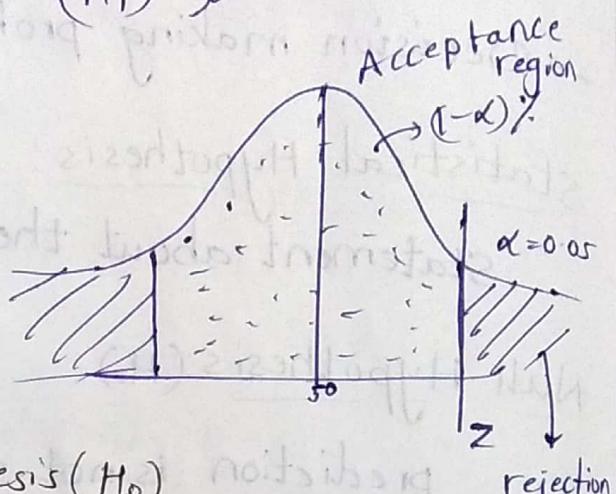
Testing of Hypothesis

Ex:

Vaccine 1 for covid 19 (H_0) $\mu = 50$

Vaccine 2

(H_1) $\mu > 50$



Condition of Null Hypothesis (H_0)

		True	False
Possible Action	Accepting	Correct action	Type II error $P(\text{II}) = \beta$
	Reject H_0	Type I error $P(\text{I}) = \alpha$	Correct action

Type I error \rightarrow rejecting Null hypothesis when it is true
 $\alpha = P$.

Type 2 error \rightarrow Failed to reject Null Hypothesis.

Testing hypothesis for the mean μ :

When the value of sample size (n)

Population is normal or not
 normal ($n \geq 30$)

Population is normal
 $(n < 30)$

σ is known

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

σ is not known

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

σ is known

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

σ is not known

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Test Procedures:

Hypotheses	$H_0: \mu = \mu_0$ $H_A: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_A: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_A: \mu < \mu_0$
Test statistic	Calculate the value of :		
Rejection Region & Acceptance Region of H_0	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$		
Critical value(s)	$Z_{\alpha/2}$ & $Z_{-\alpha/2}$	$Z_{1-\alpha} = -Z_\alpha$	Z_α
Decision	We reject H_0 (and accept H_A) at the significance level α if :		
	$Z < Z_{\alpha/2}$ or $Z > Z_{1-\alpha/2} = -Z_{\alpha/2}$ Two sided test	$Z > Z_{1-\alpha}$ $= -Z_\alpha$ One sided test	$Z < Z_\alpha$ One sided test

$$\alpha = P(\text{type I error})$$

$$\beta = P(\text{type II error}) = P(\text{failing to reject } H_0 \text{ when it is false})$$

Size of the test = α

Power of test = $1 - \beta = P(\text{reject } H_0 \text{ when it is false})$

Ex: Average mean age of population of random sample of 10 individuals from the population has mean age 27 & Variance of population is 20. Can we conclude mean is different from 30.

Sol

$$\bar{x} = 27, \sigma^2 = 20$$

$$H_0: \mu = 30$$

$$H_1: \mu \neq 30$$

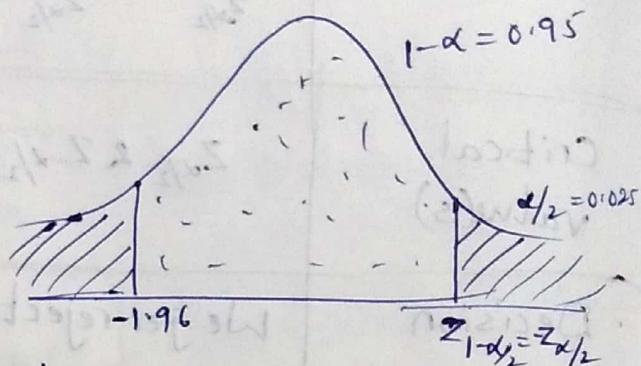
$$Z_c = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Two sided test

$$= \frac{27 - 30}{\sqrt{20}/\sqrt{10}}$$

Calculated statistic

$$Z_c = -2.121$$



Test: $Z_c > |z|$ Accept H_0 .

$$Z_{0.025} = 1.96$$

Since

$$-2.12 < -1.96 \rightarrow \text{Reject } H_0$$

Ex:

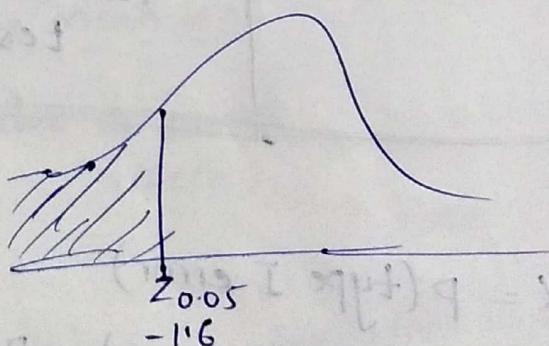
$$H_0: \mu = 30$$

$$H_1: \mu < 30$$

$$Z_c = -2.12$$

Reject H_0 .

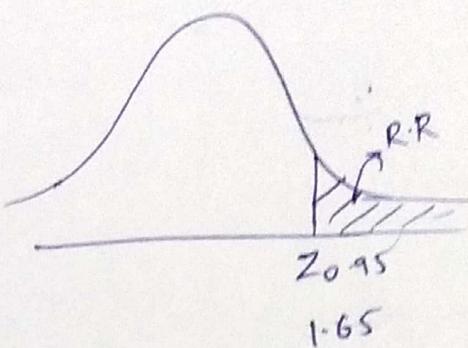
(if $Z_c < Z_{0.05}$ reject H_0 .)



Ex $n = 157$ Avg BP $\bar{X} = 146$
 $S = 27$
 $\alpha = 0.01$

$$H_0: \mu = 140$$

$$H_1: \mu > 140$$



$$Z_c = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

$$= \frac{146 - 140}{\frac{27}{\sqrt{157}}}$$

$$= \frac{6}{\frac{27}{\sqrt{157}}} = \frac{6}{2.7} \sqrt{157}$$

$$Z_c = 2.78$$

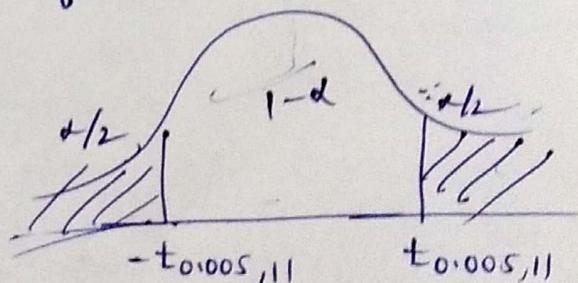
$Z_c > Z_{1-\alpha}$ Reject H_0 .

Ex T test $n = 12$, $\bar{X} = 672.6$
 $\alpha = 0.01$ $S = 43.72$

New process yield is different from 690.

$$H_0: \mu = 690$$

$$H_1: \mu \neq 690$$



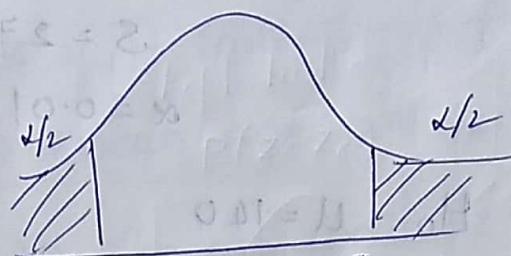
$$T_c = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{672.6 - 690}{43.72/\sqrt{11}} = -1.319$$

$$T_{0.005,11} = 3.106$$

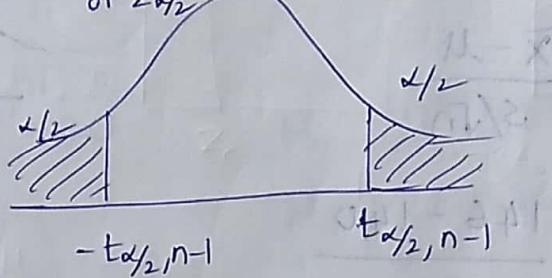
Accept H_0 .

28/6/21
Two tailed

Z test

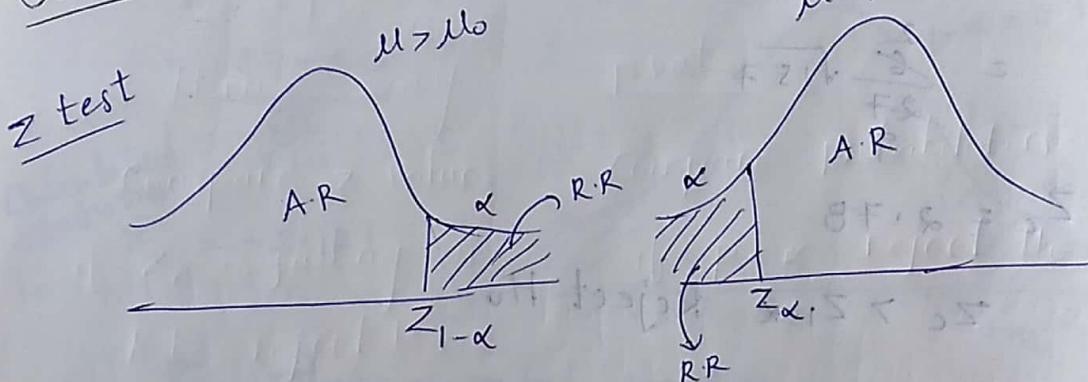


t test



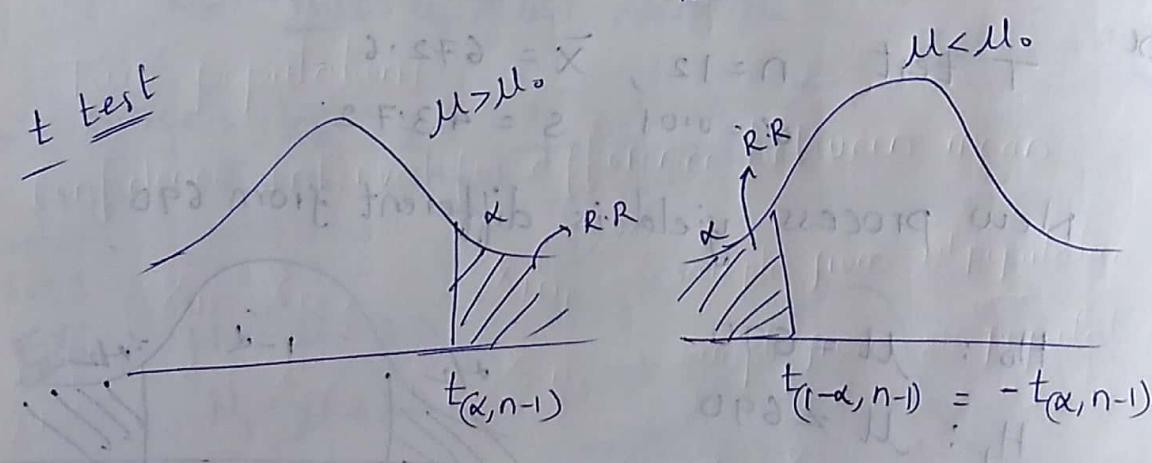
One tailed

Z test



$$\mu < \mu_0$$

t test



$$\mu < \mu_0$$

$$t_{(1-\alpha, n-1)} = -t_{(\alpha, n-1)}$$

Chi square test for Variance:

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

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 \neq \sigma_0^2$$

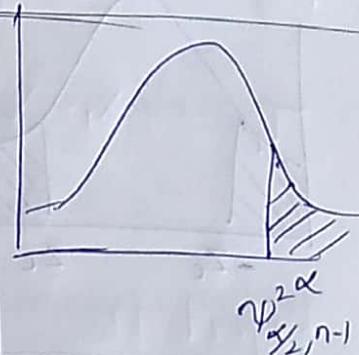
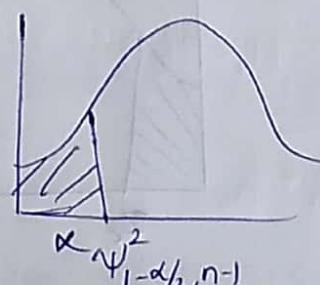
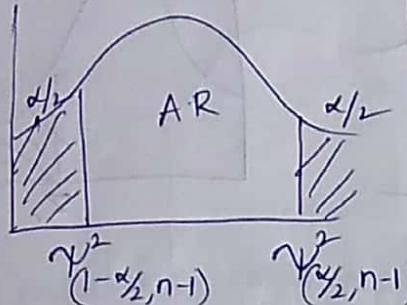
$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 < \sigma_0^2$$

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 > \sigma_0^2$$

statistic $\chi_c^2 = \frac{(n-1)s^2}{\sigma^2}$



Reject H_0
 $\chi_c^2 < \chi_{1-\alpha/2}^2, n-1$ or
 $\chi_c^2 > \chi_{\alpha/2}^2, n-1$

Reject H_0
 $\chi_c^2 < \chi_{1-\alpha/2}^2, n-1$

Reject H_0
 $\chi_c^2 > \chi_{\alpha/2}^2, n-1$

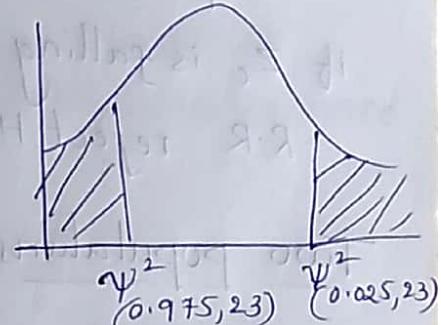
Ex: $H_0: \sigma^2 = 0.04$ $n = 24$

$$H_1: \sigma^2 \neq 0.04 \quad \alpha = 0.05$$

$$S = 0.11$$

$$\chi^2_{0.975, 23} = 11.689$$

$$\chi^2_{0.025, 23} = 38.076$$



$$\chi_c^2 < \chi_{1-\alpha/2}^2 \Rightarrow \text{Reject } H_0.$$

Procedure:

① Write H_0 & H_1

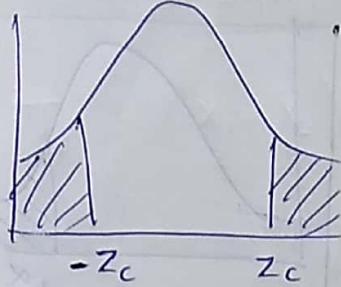
② Calculate appropriate statistic

③ Fix α & find the tabulated values / critical values

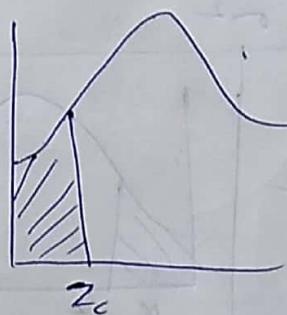
④ Compare both & take decision.

Probability
P-value:

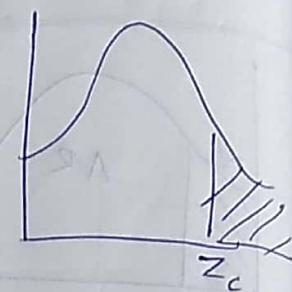
calculate P Two tailed $2P(z > z_c)$
 ~~$P(z > |z_c|)$~~



Left tailed $P(z < z_c)$



Right tailed $P(z > z_c)$



Compare with α

Reject H_0

$$P < \alpha$$

Normal way

Compare z values

If z_c is falling in
R.R reject H_0

p value

Compare probabilities

If $p < \alpha$ reject H_0

Two populations:

Confidence Interval for difference between mean.

independent & Normal $X_{11}, X_{12}, \dots, X_{1n_1} \rightarrow$ sample from 1st population
 $X_{21}, X_{22}, \dots, X_{2n_2} \rightarrow$ sample from 2nd population

Estimator for $\mu_1 - \mu_2$ is $\bar{X}_1 - \bar{X}_2$

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

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

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

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

$$\mu_1 - \mu_2 \in \left((\bar{X}_1 - \bar{X}_2) - Z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{X}_1 - \bar{X}_2) + Z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

when σ_1^2 & σ_2^2 are unknown we can replace by s_1^2 & s_2^2

Confidence interval for difference of means when n_1 or $n_2 \leq 30$

we use 't' statistic

Pooled variance \rightarrow Combined variance

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \rightarrow \text{estimator of population variance } \sigma^2$$

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad d = n_1 + n_2 - 2$$

$$\mu_1 - \mu_2 \in \left((\bar{X}_1 - \bar{X}_2) - t_{\alpha/2, d} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2, d} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

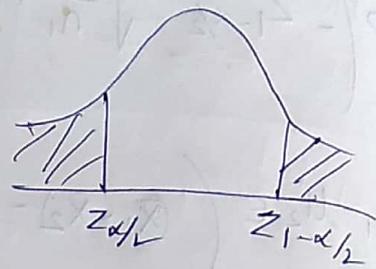
Testing of hypothesis for difference of means :

$$H_0: \mu_1 - \mu_2 = d$$

$$H_1: \mu_1 - \mu_2 \neq d, \quad \mu_1 - \mu_2 < d, \quad \mu_1 - \mu_2 > d$$

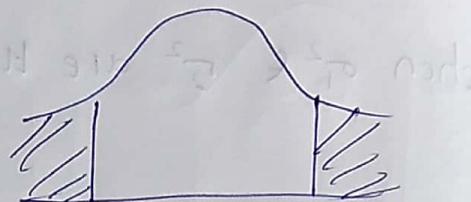
$$Z_C = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Reject H_0 if $z_0 > z_{1-\alpha/2}$ or $z < z_{\alpha/2}$



For small sample

$$T_c = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$



Reject H_0 :
if $T_c < -t_{\alpha/2, n_1+n_2-2}$ or
 $T_c > t_{\alpha/2, n_1+n_2-2}$

$$T_c > t_{\alpha/2, n_1+n_2-2}$$

$$\text{to minimize } \frac{s^2(1-\alpha) + s^2(1-\alpha)}{s^2(1-\alpha) + s^2(1-\alpha)} = s^2$$

$$\frac{(\mu_1 - \mu_2) - (\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = T$$

$$\left(\frac{1}{n_1} s_1^2 + \frac{1}{n_2} s_2^2 \right) + \left(\bar{x}_1 - \bar{x}_2 \right)^2 \cdot \frac{1}{n_1 + n_2} = b$$

sum to make not significant to null

$$b = 31 - 11 = 20$$

$$b > 31 - 11 = 20$$

$$\frac{(\mu_1 - \mu_2) - (\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 5$$