Lecture 3

Recall that $X_1, X_2, \dots X_n$ are i.i.d R.v's form any R.S.

$$\bar{X} = \frac{\sum X_i}{n}$$

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

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

, $\bar{X}, S^2$ are R.v.s

* If $X_1, X_2, \dots, X_n$ is a R.S. from normal Population

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

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

* If $X_1, X_2, \dots, X_n$ is a R.S. from any distribution, $n \geq 30 \Rightarrow \bar{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$

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

* $E(aX_1 + bX_2) = a\,E(X_1) + b\,E(X_2)$

$V(aX_1 + bX_2) = a^2 V(X_1) + b^2 V(X_2)$

$V(aX_1 - bX_2) = a^2 V(X_1) + b^2 V(X_2)$

Case of two Populations

If two indep random samples of size n_1, n_2 are drawn from any two Populations with mean μ_1, μ_2 and Variances σ_1^2, σ_2^2

What is the distribution of $\bar{X}_1 - \bar{X}_2$??

Population #1		Population #2
μ_1	P. mean	μ_2
σ_1^2	P. variance	σ_2^2
n_1	sample size	n_2
$\bar{X}_1$	sample mean	$\bar{X}_2$
S_1^2	Sample Variance	S_2^2

$\bar{X}_1 - \bar{X}_2$ is a R.V.

$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$

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

If $\bar{X}_1, \bar{X}_2$ from two normal population

$$\therefore \bar{X}_1 - \bar{X}_2 \sim Normal\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right.$$

$$\bar{X}_2 - \bar{X}_1 \sim Normal\left(\mu_2 - \mu_1, \frac{\sigma_1^2}{n_1^2} + \frac{\sigma_2^2}{n_2^2}\right)$$

$$\therefore Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

⇒ Similarly, if $n_1 \geq 30$, $n_2 \geq 30$ (large enough)
by central limit theorem

$\bar{X}_1 - \bar{X}_2 \sim$ Normal

regardless of the shape of two population

⇒ If σ_1^2, σ_2^2 are unknown, $n_1 \geq 30, n_2 \geq 30$

C.L.T still valid with using sample variances S_1^2, S_2^2 instead of σ_1^2, σ_2^2

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$$

or

① Sample taken from normal Population
σ^2 is known
⇓
$\bar{X} \sim$ Normal$(\mu, \frac{\sigma^3}{n})$

② • $n \geq 30$ Sample taken from any distr.
• $n \geq 30$, σ^2 is unknown
$\sigma^2 \to S^2$
⇓
$\bar{X} \sim$ Normal$(\mu, \frac{\sigma^3}{n})$

③ When $n < 30$ and σ^2 is unknown

??

Ex ① The electric light bulb of manufacturer A have a mean lifetime of 1400 hrs with a standard deviation of 200 hrs, while those of manufa B have mean life time of 1200 hrs with st. dev = 100 hrs. IF random sample of 125 bulbs of each brand are tested what is the Pr. that brand A bulbs will have a mean lifetime which is at least 160 hrs more than brand B bulbs?

A: $\boxed{\mu_A = 1400 \quad \sigma_A = 200 \quad n_A = 125}$

B: $\boxed{\mu_B = 1200 \quad \sigma_B = 100 \quad n_B = 125}$

$P(\bar{X}_A - \bar{X}_B \geq 160) = ?$

$\bar{X}_A$ mean lifetime of sample of brand A

$\bar{X}_B$ 〃 〃 〃 〃 〃 B

Since $n_A \geq 30$ && $n_B \geq 30$

$\therefore$ C.L.T is valid

$$\bar{X}_A - \bar{X}_B \sim N\left(\mu_{A-B} = \mu_A - \mu_B = 200,\ \sigma^2_{A-B} = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \frac{200^2 + 100^2}{125}\right)$$

$\sigma^2_{A-B} = 400$

$\sigma_{A-B} = 20$

$$P\left(\frac{(\bar{X}_A - \bar{X}_B) - \mu_{A-B}}{\sigma_{A-B}} \geq \frac{160 - 200}{20}\right)$$

$$P(Z \geq -2) = 1 - P(Z \leq -2)$$

$= 1 - \Phi(-2)$

$= 1 - (1 - \Phi(2))$

$= \Phi(2)$

$= 0.977$ by Minitab or R.

To get $\Phi(2)$ using Minitab

Menubar/Calc/Probability distributions/Normal [Mean 0, st. dev. 1, Input const 2]

Distributions derived from normal distr.

1. Chi-squared distr.
2. student-t distr.
3. F- distr.

The chi-squared Distribution χ^2_K

1. Let $Z_1, Z_2, \dots, Z_K$ are indep standard normal R.U.s ($i.i.d\ N(0,1)$)

let $U = Z_1^2 + Z_2^2 + Z_3^2 \dots + Z_K^2$

$U \sim \chi^2_K$ with K degrees of freedom

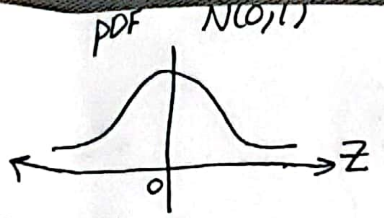
degrees of freedom No. of indep Z's

χ^2_K with K degrees of freedom is the distribut of a sum of the squares of K indep. standard normal variables

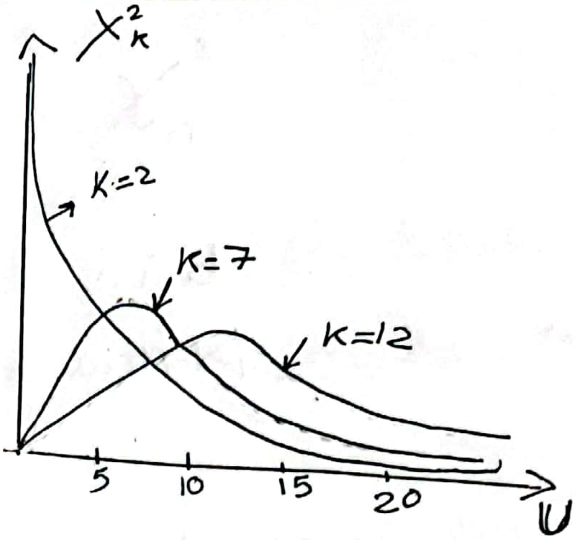
$$\boxed{\text{if } U \sim \chi^2_K \Rightarrow E(U) = K \quad , \quad V(U) = 2K}$$

$U_1 = Z_1^2 \Rightarrow U_1 \sim \chi^2_1 \quad (M = 1, Var = 2)$

$U_2 = Z_1^2 + Z_2^2 \Rightarrow U_2 \sim \chi^2_2 \quad (M = 2, Var = 4)$



$\Rightarrow$

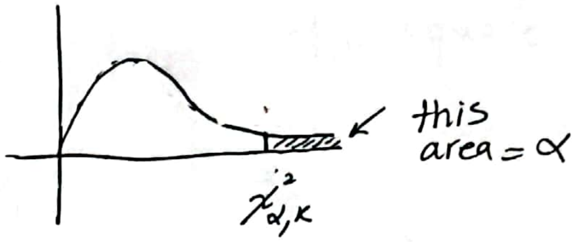


chi-squared curve

$\Rightarrow$ Not symmetric

$\Rightarrow$ No -ve value for χ^2

To get $\chi^2_{\alpha,K}$



$$F(\chi^2_{\alpha,K}) = 1-\alpha$$

CDF

$$\therefore \chi^2_{\alpha,K} = F^{-1}(1-\alpha)$$

Using Minitab

To get $\chi^2_{\underset{\alpha}{0.05},\overset{K}{12}}$

$$\therefore F(\chi^2_{0.05,12}) = 1-0.05 = 0.95$$

$$\therefore \chi^2_{0.05,12} = F^{-1}(0.95)$$

Menu bar / Calc / prob. distr. / Chi_square

- inverse Cumulative

degrees of freedom = 12

Input constant = 0.95

$$\chi^2_{0.05,12} = 21.02$$

$\boxed{2}$ let $X_1, X_2, \ldots, X_n$ are i.i.d $N(\mu, \sigma^2)$

then $\frac{X_1-\mu}{\sigma}, \frac{X_2-\mu}{\sigma}, \ldots, \frac{X_n-\mu}{\sigma}$ are i.i.d $N(0,1)$

hence, $U = \left(\frac{X_1-\mu}{\sigma}\right)^2 + \left(\frac{X_2-\mu}{\sigma}\right)^2 \cdots + \left(\frac{X_n-\mu}{\sigma}\right)^2 \sim \chi^2_n$

$$\therefore \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi_n^2$$

$\boxed{3}$ If $U_1, U_2, U_3, \dots, U_n$ are i.i.d R.V.'s having Chi squared distribution with $k_1, k_2, k_3, \dots, k_n$ d.o.f then

$$V = \sum_{i=1}^{n} U_i \quad \text{has Chi-square distr.}$$

with $k_1 + k_2 - \dots + k_n$ dof

e.g. $U_1 \sim \chi_3^2 \qquad \therefore U_1 = Z_1^2 + Z_2^2 + Z_3^2$

Z_1, Z_2, Z_3 are indp.

$U_2 \sim \chi_5^2 \qquad \therefore U_2 = Z_4^2 + Z_5^2 + Z_6^2 + Z_7^2 + Z_8^2$

$Z_4, Z_5, Z_3, \dots, Z_8$ are indep

$$\therefore U_1 + U_2 \sim \chi_8^2$$

Ex what is the distribution of $n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2$

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

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

$$U = Z^2 = \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 = n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2$$

$$\therefore n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2 \sim \chi_1^2 \Rightarrow \boxed{1}$$

The distribution of s^2

① $\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n \rightarrow ①$

② $n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2 \sim \chi^2_1 \rightarrow ②$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \Rightarrow (n-1)s^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 \rightarrow ③$$

Proof

$$\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} \left[(x_i - \bar{X}) + (\bar{X} - \mu)\right]^2$$

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

$$= \sum_{i=1}^{n} (X_i - \bar{X})^2 + 2(\bar{X} - \mu)\sum (X_i - \bar{X}) + \sum (\bar{X} - \mu)^2$$

$n\bar{X} = \sum X_i$

$$\boxed{\textbf{Note} \quad \sum_{i=1}^{n} K = nK \qquad \sum_{i=1}^{n} (X_i + K) = \sum X_i + nK}$$

$$\sum (X_i - \mu)^2 = \sum (X_i - \bar{X})^2 + \underbrace{2(\bar{X} - \mu)(\sum X_i - n\bar{X})}_{0} + n(\bar{X} - \mu)^2$$

deviding both sides by σ^2

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} \quad \text{from ③}$$

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

$$\therefore \frac{(n-1)S^2}{\sigma^2} = \frac{\sum (x_i-\mu)^2}{\sigma^2} - n\frac{(\bar{x}-\mu)^2}{\sigma^2}$$

$\downarrow$ from ① χ^2_n $\qquad\qquad$ $\downarrow$ from ② χ^2_1

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

if sample from normal distr. or $n \geq 30$

$E(S^2)$, $V(S^2)$??

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

$\boxed{U \sim \chi^2_n \quad E(U)=n \quad V(U)=2n}$

$$\frac{(n-1)}{\sigma^2}E(S^2) = n-1$$

$$\boxed{E(S^2) = \sigma^2}$$

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

$$\frac{(n-1)^2}{\sigma^4}V(S^2) = 2(n-1) \Longrightarrow \boxed{V(S^2) = \frac{2\sigma^4}{n-1}}$$

$\therefore$ Sample variance S^2 estimates the population variance σ^2 regardless of the distribution from which we are sampling