Introduction to Probability & Statistics
Prof Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 11
Lecture - 41
Student's t-distribution

ab hum teesra case dekhenge confidence interval ke liye jahan pe hamara population button hai normal par ab hum assume karenge ki mu aur sigma square dono unknown hain; pehle wale case mein sigma known tha, doosre case mein large-sample approximation tha, par yeh teesra case us situation ke liye hai jahan n large assume nahi karna; yahan mu ka estimator ab bhi x-bar hi hai aur sigma-square ka estimator s-square hi hai; humne pehle dekha tha ki x-bar − mu / (s / sqrt(n)) approx normal hota hai jab sample size large ho, par agar original population bhi normal hai aur sigma unknown hai to yeh standardized quantity ka exact distribution t-distribution hota hai with n-1 degrees of freedom; matlab

$$T = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

yeh t distribution ek continuous distribution family hai jiska ek parameter hota hai jise bolte hain degree of freedom ya degrees of freedom, jisko generally Greek μ (nu) se denote karte hain aur μ koi positive integer hota hai; properties of t-distribution: (1) t-distribution bell-shaped hota hai standard normal ki tarah aur symmetric around 0; (2) har t-curve standard normal se zyada spread out hota hai, matlab tails heavy hoti hain; (3) jaise-jaise degrees of freedom μ badhta hai, curve ka spread kam hota hai; (4) $\mu \to \infty$ hone par t-distribution standard normal ko approximately approach karta hai isi liye standard normal ko kabhi-kabhi t-∞ distribution bhi kaha jata hai; ab jab confidence interval nikalna ho mu ke liye aur sigma unknown ho, tab hum Z-values nahi use karte hum use karte hain T-critical values, jise denote karte hain $t_{(\alpha,\mu)}$; yeh woh value hoti hai jahan random variable T (with μ degrees of freedom) ke $t_{(\alpha,\mu)}$ se right-side mein rehne ki probability hoti hai α, yaani $P(T > t_{(\alpha,\mu)}) = \alpha$; isi logic ke saath hum confidence interval banayenge using t-distribution instead of z. ye to t-alpha-nu woh vastavik sankhya hai jiske right mein t-nu curve ke neeche wali area hoti hai alpha, bilkul waise hi jaise Z curve ke liye hota tha; lekin normal (Z) ke case mein hum har normal ko standard normal mein convert kar sakte the, par t-distribution ke saath hum waise nahi kar sakte phir bhi t-values tabulate ki ja sakti hain; ise kaafi log Student's t-distribution kehte hain (Student ek scientist ka naam tha jinke naam par distribution rakha gaya), aur ye jo table aap dekh rahe ho usme degrees of freedom pehle column mein diya hota hai (1,2,3,…50,60,70,100,∞), aur jo column headings hoti hain wo upper-tail probability α hoti hai (0.2, 0.1, 0.05, etc). ab ek zaroori baat isse "degrees of freedom" kyun kehte hain: kyunki S^2 ki definition mein jo deviations from mean aate hain $x1-\bar{x}$, $x2-\bar{x},\ldots,xn-\bar{x}$ unme se n values freely choose nahi ki ja sakti; inka sum hamesha 0 hota hai, isliye hum $n-1$ deviations freely choose kar sakte hain; isliye T statistic ke liye degrees of freedom hoti hain $n-1$; ab ek proposition likhte hain: jab $\bar{X}$ aur S sample mean aur sample standard deviation ho from a normal μ,σ^2 population, tab $100(1-\alpha)\%$ confidence interval for μ hota hai:

$\bar{x} \pm t(\alpha/2, n-1) \cdot S/\sqrt{n}$;

yahi normal-case formula tha, bas Z ki jagah T aagaya; ek udaharan: sample of joint specimens mein sample mean stress aaya 8.48 MPa, sample SD 0.79 MPa, assume stress ~ Normal(μ,σ^2), n = 14; humein 95% CI chahiye; to interval =

$8.48 \pm t(0.025,13) \cdot 0.79/\sqrt{14}$;

ab table use karenge: df=13 row, alpha=0.025 column, value=2.16; to answer ban gaya:

$8.48 \pm 2.16 \times (0.79/\sqrt{14})$,

ye hoga 95% confidence interval for the true average proportional limit stress.