Introduction to Probability & Statistics

Prof Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi

Week - 12
Lecture - 44
Tests for Population Means

to hum test of hypothesis dekh rahe hain population mean ke liye jaha normal population hai with known sigma, aur H0: mu = mu0, H1: mu > mu0 ke case me humne Z statistic use kiya tha:

$$Z = (\frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}})$$

jiska distribution H0 ke true hone par standard normal 0,1 hota hai; level-alpha test me rejection region hota hai: reject H0 if Z ≥ Z_alpha, jaha Z hum sample se compute karte hain aur Z_alpha standard normal table se read karte hain; isi tarah agar H1: mu < mu0 ho to rejection region banega: Z ≤ −Z_alpha (lower-tailed test), aur agar H1: mu ≠ mu0 ho to rejection region: Z ≤ −Z_{alpha/2} OR Z ≥ Z{alpha/2} (two-tailed test), kyunki total tail area alpha hota hai left tail me alpha/2 aur right tail me alpha/2; ab ek example: paint ka drying time mu aur sigma² ke saath assume normal, sigma = 9 minutes, existing paint ka mu = 75 minutes; ab ek additive propose kiya gaya hai jo drying time reduce karne ka claim karta hai; new paint ke sample of n = 25 test specimens se drying times collect kiye gaye aur unka sample mean mila: x̄ = 71; question: kya ye data sufficient evidence deta hai ke claim sahi hai; step-by-step: (1) parameter identify: mu = true average drying time for new paint; (2) null hypothesis: H0: mu = 75 (prior info); (3) alternate hypothesis: H1: mu < 75 (since drying time reduce hone ka claim hai); (4) test statistic choose karein: since sigma known hai, x̄ is normally distributed, to hum use karenge: Z = (x̄ − mu0)/(sigma/√n); ye test statistic H0 ke true hone par standard normal distribution follow karega; aage ke steps me hum critical value aur compute Z se comparison karke decision lenge.to x̄ normal hai isliye hum test statistic use karenge

$$Z = \left(\frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}\right)$$

lekin yaha sigma unknown hai to hum sigma ki jagah sample standard deviation s lagaenge: Z = (x̄ − mu0)/(s/√n); jab H0 true hota hai to Z approx standard normal hota hai (large sample case), phir step-5 me hum rejection region define karte hain is example me H1: mu > 2.5 hai jo ek upper-tailed test hai, to rejection region hoga: reject H0 if Z ≥ Z_alpha; ab step-6 me test statistic compute karte hain: x̄ = 2.6, mu0 = 2.5, s = 0.42, n = 64, so small-z = (2.6 − 2.5) / (0.42/√64) = 0.1 / (0.42/8) = 0.1 / 0.0525 ≈ 1.90; ab step-7 me hum conclusion nikalte hain:

alpha = 0.05 ke liye Z_alpha = 1.645, aur hamari computed value 1.90 > 1.645, matlab small-z rejection region me aata hai, isliye H0 ko reject karte hain; interpretation: data support karta hai claim ke against yani voltage drop 2.5 volts se zyada ho sakta hai, to claim "at most 2.5 volts" supported nahi hota; isi tarah har test me ye 7 steps systematically follow karenge 1) parameter, 2) hypotheses, 3) alternate form, 4) test statistic, 5) rejection region, 6) compute statistic, 7) conclusion. to Z greater than equal to Z alpha yani $Z \geq Z0.05$ aur $Z \geq 1.645$, agar hamara data aisa hai jiske liye Z ki value 1.645 se zyada ho to hum H0 reject karenge; step-6 me hum Z compute karte hain: $Z = (2.6 - 2.5) / (0.42/\sqrt{64}) = 1.90$; final step me conclusion: kyunki 1.90 > 1.645, matlab hamari computed Z value rejection region me aati hai, isliye hum H0 ko reject karte hain in favor of H1, yani jo average voltage drop hai wo specified limits ke andar nahi milta yaani 2.5 volt se zyada ho sakta hai, lekin dhyan rahe ye conclusion level alpha = 0.05 ke liye hai; is tarah humne case-2 pura kiya; ab case-3 me population normal maana jata hai par sigma unknown hota hai aur sample size chhota ho sakta hai; is case me statistic

$$t = \left(\frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}\right)$$

hota hai jiska distribution t-distribution hota hai (standard normal nahi); H0: mu = mu0 hi rahega aur H1 ke 3 type ke test upper tail, lower tail, or two-tail isme rejection region t-table se milega; finally chemical factory wala example: glycerol concentration ke 5 sample values (2.67, 4.62, 4.14, 3.81, 3.83), parameter mu define karna, H0: mu = 4, H1: mu ≠ 4, test statistic $t = (\bar{x} - 4)/(s/\sqrt{5})$, jisme $\bar{x} = 3.814$ and $s = 0.718$, $t = -0.58$ aata hai; alpha = 0.05 ke liye critical $t = \pm 2.776$ hota hai (df=4), aur kyunki −0.58 rejection region me nahi aata (yani $-2.776 < -0.58 < 2.776$), hum H0 ko reject nahi karte, yani data suggest nahi karta ki true average glycerol concentration desired level 4 se different hai; isi tarah hum sampling, hypothesizing, test statistic, rejection region, test value compute, aur conclusion ye 7 steps systematically follow karke hypothesis testing complete karte hain.