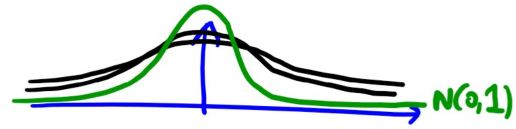
3Y03 – 3J04 PROBABILITY & STATISTICS FOR (CIVIL) ENGINEERING

C01 - Lecture 24

Last time (STUDENT) t – DISTRIBUTION

Used when: $X_1, \dots, X_n \sim N(\mu, \sigma^2)$, but μ, σ^2 unknown and n small

Then $T = \dfrac{\bar{X} - \mu}{(S/\sqrt{n})} \sim T_{n-1}$



t – distribution with "$n-1$ degrees of freedom"

$100(1-\alpha)\%$ C.I. for μ:

$$\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$$

Example 12 students write Test #2 at an alternate time; for these students $\bar{x} = 77\%$ and $s = 13.61\%$. Find a 95% C.I. for the class average for Test #2.

Solution $1 - \alpha = 0.95$ $\qquad$ $n - 1 = 11$

so $\frac{\alpha}{2} = 0.025$

Look up t – table : go to $\frac{\alpha}{2}$ i.e. 0.025 column & $n-1$ i.e. 11 row

[Table V in Textbook p. 745]

Read off $t_{0.025, 11} = 2.201$

Then $\bar{x} \pm t_{\alpha/2, n-1} \cdot \left(\frac{s}{\sqrt{n}}\right) = 77 \pm 2.201 \left(\frac{13.61}{\sqrt{12}}\right)$

$= 77 \pm \underbrace{8.647}_{= \text{Margin of Error}}$

i.e. $(68.353, 85.647)$.

8.4 C.I for Population Proportion (Large Sample)

Recall: If X is Binomial $\sim Bin(n, p)$

(# successes in n independent Bernoulli trials, where prob. success in 1 trial $= p$.)

If n is large, then $Z = \dfrac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$

i.e. $np, n(1-p) > 5$

$\hat{P} = \dfrac{X}{n}$ is an estimator of p ~~(proportion)~~

probability above

$E(\hat{P}) = \frac{1}{n} E(X) = \frac{np}{n} = p$ so in fact an unbiased estimator of p

If we have a large population & we are interested in proportion of population in a given category

we could say success = in category
failure = not in category

So $\hat{p}$ is an unbiased estimator for p, the proportion of the population in category of interest.

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}\,\cancel{n}\hat{p} - \overset{\sqrt{n}}{\cancel{n}}p}{\sqrt{\cancel{n}p(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

$Z \sim N(0,1)$

$$P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

We can find a $100(1-\alpha)\%$ C.I. for p in an analagous way to arguments above:

$$\left(\hat{p} - z_{\frac{\alpha}{2}}\underbrace{\sqrt{\frac{p(1-p)}{n}}}, \ \hat{p} + z_{\frac{\alpha}{2}}\underbrace{\sqrt{\frac{p(1-p)}{n}}}\right)$$

Notice this is the standard error of $\hat{p}$ & <u>depends</u> on p — so estimate!

Replace p with $\hat{p}$ in above:

$$\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \ \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

where $\hat{p}$ is calculated using data. ↖ <u>THIS</u> is then the $100(1-\alpha)\%$ C.I. for p.

<u>Example</u> In a random sample of 50 engineering students, 18 have blond hair.

Find a 99% C.I. for proportion of all engineering students having blond hair.

<u>Solution</u> We need to find $z_{\frac{\alpha}{2}}$ and $\hat{p}$.

$$\frac{\alpha}{2} = 0.005 \quad \text{so} \quad z_{\frac{\alpha}{2}} = 2.58.$$

So $\hat{p} = \dfrac{18}{50}$. So our 99% C.I. for p

has endpoints
$$\frac{18}{50} \pm \underbrace{2.58\sqrt{\frac{\frac{18}{50}\left(\frac{32}{50}\right)}{50}}}_{\text{Marginal Error}}$$

$$= 0.36 \pm 0.175.$$

Now suppose we want a $100(1-\alpha)\%$ C.I. within a specified margin of error (ME). How big must our sample size n be?

Want
$$z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq ? = \text{specified ME}$$

i.e.
$$n \geq \left(\frac{z_{\alpha/2}}{?}\right)^2 p(1-p)$$

But we didn't take sample yet! So how to estimate p ?!!

2 methods: (1) Use an earlier sample (e.g. small pilot study) & compute $\hat{p}$ from that & use it to estimate p

(2) The biggest $p(1-p)$ can possibly be is $\frac{1}{2} \times \frac{1}{2} = 0.25$

optimization problem! $p(1-p) = p - p^2$
Differentiate & set to 0: $1 - 2p = 0$ i.e. $p = 1/2$
1st derivative test: at $p = 1/2$, derivative is decreasing so local max.

So substitute in 0.25 for $p(1-p)$.

<u>Example</u> We want a 95% C.I. for the proportion of engineering students with blond hair with ME at most 0.07. How many students should be in our sample?

<u>Solution</u>

$$ME \leq z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

$$n \geq \left(\frac{z_{\alpha/2}}{0.07}\right)^2 p(1-p)$$

$z_{\alpha/2} = 1.96$

2 methods: (1) Use earlier Example where we ~~found~~ had, for that sample, that $\hat{p} = 18/50$.

Then $n \geq (1.96/0.07)^2 (18/50)(32/50) = 180.6$ $\to n = 181$

(2) Use worst case $p = 0.5$. Then

$$n \geqslant \left(\frac{1.96}{0.07}\right)^2 \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 196.$$