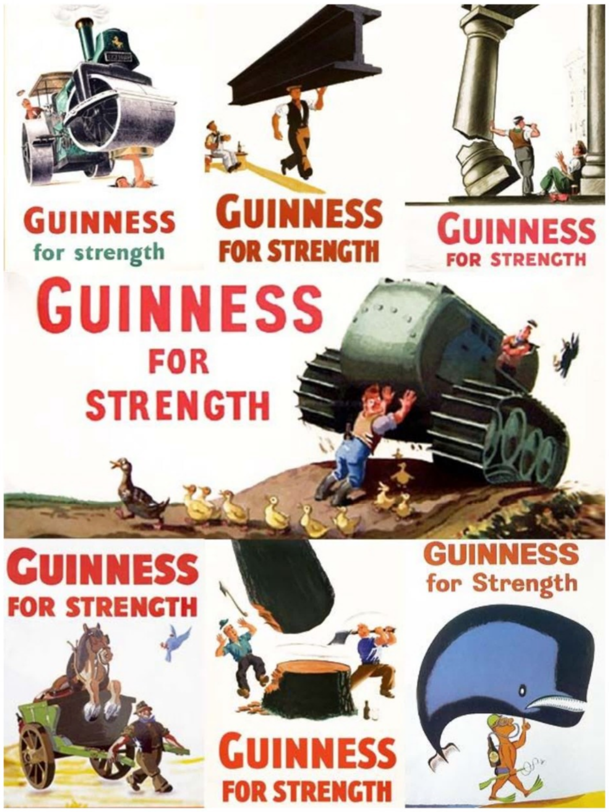
3Y03 - PROBABILITY AND STATISTICS FOR ENGINEERING

WS19 Lecture 23



Last time Confidence Intervals for the mean

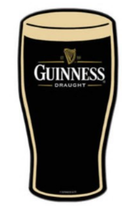
e.g.

If $X_1, \dots, X_n \sim N(\mu, \sigma^2)$, but σ^2 UNKNOWN, n NOT large:

then use that $T := \dfrac{\bar{X} - \mu}{(S/\sqrt{n})}$ has a

(student) t distribution

$n-1$ degrees of freedom.



$100(1-\alpha)\%$ C.I.: $\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \dfrac{s}{\sqrt{n}}$

See website for more on William S. Gosset AKA "student".

Example 12 $=n$ students write Test #2 at an alt. time; for these students $\bar{x} = 77\%$ & $s = 13.61\%$

Find a 95% C.I. for the class average for Test 2.

Solution

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

$1 - \alpha = 0.95$

$\alpha = 0.05$

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

$$77 \pm \underbrace{t_{0.025, 11}} \frac{13.61}{\sqrt{12}}$$

don't get confused by the fact that the columns are labelled with "α" values. Look up the # that you need!

Look up t-table, 0.025 column, 11 row

p.745 of textbook

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

So C.I. $\quad 77 \pm (2.201)\left(\dfrac{13.61}{\sqrt{12}}\right) = 77 \pm \underbrace{8.647}_{\text{M.E.}}$

Explicitly: $(68.353, 85.647)$.

C.I.s for variance — not in this course

8.4 C.I. for Population Proportion

("Lage" Sample)

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

X = # successes in n trials (each of which has prob. p of success, & all trials independent)

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

when n lage

$np, n(1-p) > 5$

Define $\hat{p} = \dfrac{X}{n}$; is an estimator for p < prob. / proportion

" !!! ∴ it's $\dfrac{X}{n}$ i.e. $\dfrac{\#\text{successes}}{\text{total observations}}$

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

So $\hat{p}$ unbiased estimator for p = prob.

= prop. in given category

$$\text{So } Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{X/n - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

So a $100(1-\alpha)\%$ C.I. for p analagous to above for mean :

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

Problem: M.E. depends on p !!!

$\longrightarrow$ so we estimate with $\hat{p}$:

$100(1-\alpha)\%$ C.I. for p :

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

Now everything can be found using data.

Example Random sample of 50 engineering students. 18 have blond hair.

Find a 99% C.I. for the proportion of all eng. students that have blond hair.

Solution We need $z_{\frac{\alpha}{2}}$, $\hat{p} \leftarrow \frac{18}{50} = 0.36$

$1-\alpha = 0.99$

$\alpha = 0.01$

$\frac{\alpha}{2} = 0.005$

$1 - \alpha/2 = 0.995$ $\longrightarrow$ $z_{\alpha/2} = 2.58$

Get used to commonly appearing $z_{\frac{\alpha}{2}}$ values e.g.

$100(1-\alpha)\% = 99\%$: $z_{\alpha/2} = 2.58$.

So C.I. : $\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$$= 0.36 \pm 2.58 \sqrt{\frac{0.36(1-0.36)}{50}}$$

$$= 0.36 \pm 0.175.$$

Now suppose we want a $100(1-\alpha)\%$ C.I. for proportion with a certain Margin of Error (ME) (or interval length = $2 \times$ ME).

i.e. Want $z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq ME$

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

Problem: at this stage in process, maybe don't have an estimate $\hat{p}$ for p!

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

② The biggest that $p(1-p)$ could be is 0.25 (i.e. $p = 0.5$) so substitute $\hat{p} = 0.5$ i.e. $n \geq \left(\frac{z_{\alpha/2}}{ME}\right)^2 (0.25)$.

Example Want 95% C.I. for proportion of eng. students with blond hair with ME at most 0.07. How many students should we check.

Solution $$n \geqslant \underbrace{\left(\frac{z_{\alpha/2}}{ME}\right)^2}\hat{p}(1-\hat{p})$$

$z_{\alpha/2}$: $1-\alpha = 0.95$ $\quad 1-\frac{\alpha}{2} = 0.975$

$\alpha = 0.05$ $\quad \rightarrow z_{\alpha/2} = 1.96$

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

↑ another common $z_{\frac{\alpha}{2}}$ value, for $100(1-\alpha)\% = 95\%$.

So want $$n \geqslant \left(\frac{1.96}{0.07}\right)^2(\hat{p}(1-\hat{p})).$$

2 methods: ① Use earlier example if we had that info : $\hat{p} = \frac{18}{50} = 0.36$

Then $$n \geqslant \left(\frac{1.96}{0.07}\right)^2(0.36(1-0.36))$$

$= 180.63$ So $n \geqslant 181$.

② Worst-case scenario: $\hat{p}(1-\hat{p}) = 0.25$

Get $$n \geqslant \left(\frac{1.96}{0.07}\right)^2(0.25) = 196.$$