C.I. on Population Proportion

- let's say x/n observations belong to a "class of interest."

  ex: # of green m&ms out of n total m&ms

- we can compute a sample proportion as a point estimator for population proportion p:

$$\hat{p} = \frac{x}{n}$$

- We would like a confidence interval on the unknown population parameter p, population proportion.

- whether an observation belongs or doesn't belong to a class of interest: binomially distributed.

  - either an m&m is green, or isn't!

.. a $(1-\alpha)\times 100\%$ C.I. on p is :

$$p: \hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

... where $\hat{p} = \frac{x}{n}$

ex: $n = 85$ crankshaft bearings

.. we just found that $x = 10$ have surface finish rougher than allowed

.. write 95% confidence interval on the population proportion of bad bearings

.. Sample pop. proportion :

$$\hat{p} = \frac{x}{n} = \frac{10}{85} = 0.12$$

[or 12%]

.. need $Z_{\alpha/2}$;

95% C.I. $\rightarrow \alpha = 0.05 \rightarrow Z_{.05/2} = Z_{.025} = 1.960$

$$0.12 \pm 1.960\sqrt{\frac{0.12(1-0.12)}{85}}$$ ← Binomial !

$$0.05 < p < 0.19$$

or $5\% < p < 19\%$

∴ not particularly useful !

∼ how do we narrow the interval w/ same level of confidence?

∴ increase n, just like C.I. on μ, known σ^2

∴ if you want a one-sided width E:

$$n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 p(1-p)$$

∴ Wait ! We don't know p, and never will !
(population proportion!)

∴ somebody smart noticed a local minimum

$$p(1-p) = 0.25$$

∴ no matter what p is, $p(1-p)$ can't exceed 0.25

∴ we may use

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \cdot 0.25$$

... and it will work!

∴ don't forget to round up to next integer value of n

∴ what if we need a width of ±2%?

∴ then $E = 0.02$

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \cdot 0.25 = \left(\frac{1.960}{0.02}\right)^2 \cdot 0.25$$

$n = 2401$ ∴ that's a lot of crankshaft bearings!

- one-sided conf. bounds : use appropriate upper or lower bound of C.I., and substitute z_α for $z_{\alpha/2}$.

Prediction Interval

- different from a C.I. on μ; this is an interval on the next value of X.

$$X_{n+1} : \quad \bar{x} \pm t_{\alpha/2, n-1}\, S \sqrt{1 + \frac{1}{n}}$$

reminds you that it's the next value; hence, prediction!

widens interval compared to a C.I. on μ due to uncertainty in prediction!

ex: tensile adhesion test problem;

$n = 22, \bar{x} = 13.71, s = 3.55$

95% C.I. on μ:

$$12.14 < \mu < 15.28$$

∴ now write a 95% prediction interval on the tensile strength of the 23rd sample

$\alpha = 0.05$

$$t_{\alpha/2, n-1} = t_{.025, 21} = 2.080$$

n is still 22!

$$X_{23}: \quad 13.71 \pm 2.080 \cdot 3.55 \sqrt{1 + \frac{1}{22}}$$

$$6.16 < X_{23} < 21.26$$

∴ **much** wider than C.I. on μ due to high sample std. dev. of $s = 3.55$