Approximate Confidence Intervals for p.

Consider a population in which a specific trait occurs w/ an unknown probability p.
Let $(Y_1, Y_2, \dots, Y_n)$ be a random sample from the Bernoulli dist'n w/ the unknown p.

Goal: Designing a confidence interval for p.

Idea: Look @ the natural point estimator for p.

$$\rightarrow \quad \bar{Y} = \boxed{\frac{1}{n}}\Big(\underbrace{Y_1 + Y_2 + \dots + Y_n}_{\sim B(n,p)}\Big)$$

Note: $\mathbb{E}[\bar{Y}] = p$ Of course, the dist'n of $\bar{Y}$ depends on p.

We need a pivotal quantity.

By de Moivre-Laplace Thm:

Set $S_n = Y_1 + \dots + Y_n$

Then,
$$\boxed{\frac{S_n - np}{\sqrt{np(1-p)}}} \xrightarrow{\mathcal{D}} N(0,1)$$

We note $\bar{Y} = \frac{1}{n} S_n$

$$U = \frac{\bar{Y} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathcal{D}} N(0,1)$$ is still not a pivotal quantity

But we can still create an approximate confidence interval based on it!

Say, C is a confidence level.

Let $$z^* = \Phi^{-1}\left(\frac{1+C}{2}\right) = \text{qnorm}((1+C)/2)$$

$$\mathbb{P}\left[-z^* \leq \mathcal{U} \leq z^*\right] \approx C$$

dM·L

$$\mathbb{P}\left[-z^* \leq \frac{\bar{Y}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z^*\right] \approx C$$

$$\mathbb{P}\left[\bar{Y}-z^*\sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{Y}+z^*\cdot\sqrt{\frac{p(1-p)}{n}}\right] \approx C$$

Usually, we write $\boxed{\bar{Y} = \hat{p}}$

We construct the confidence interval for p as

$$\boxed{p = \hat{p} \pm z^*\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Problem 16.2. Gallup's inaugural measure of global loneliness shows over one in five people worldwide (23%) said they felt loneliness "a lot of the day yesterday."[1] *However, there were considerable variations between countries. For instance, out of 1000 individuals polled in Taiwan, 11% reported having felt loneliness "a lot of the day" before. What 90%-confidence would you report for the population proportion of Taiwanese who had felt lonely the day before?*

$\longrightarrow$: $\hat{p} = 0.11$

$z^* = \Phi^{-1}(0.95) = 1.645$

$$p = \hat{p} \pm \boxed{z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = 0.11 \pm 1.645 \cdot \sqrt{\frac{0.11 \cdot 0.89}{1000}}$$

$\square$

Example. How do we figure out the necessary sample size w/ a required margin of error m?

$\longrightarrow$: $n = ?$ $\boxed{z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le m}$ $/^2$

w/ z^* the critical value for the confidence level C

$$(z^*)^2 \cdot \frac{\hat{p}(1-\hat{p})}{n} \le m^2$$

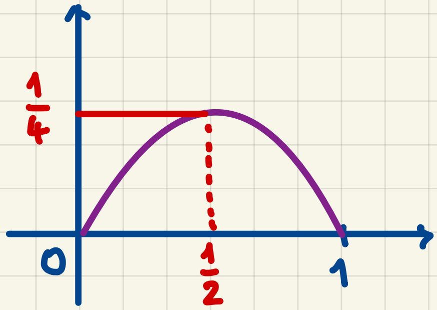
$$(z^*)^2 \cdot \frac{\boxed{\hat{p}(1-\hat{p})}}{m^2} \le n$$

We cannot know $\hat{p}$ before knowing n.

---

[1] https://news.gallup.com/poll/646718/people-worldwide-feel-lonely-lot.aspx

Consider $\hat{p}(1-\hat{p})$ as a function of $\hat{p}$.



The conservative choice for $\hat{p}$ is $\frac{1}{2}$.

The conservative choice for the sample size

$$\boxed{n \geq \left(\frac{z^*}{2m}\right)^2}$$

$\square$