M378K: November 6^{th}, 2024.

Approximate Confidence Intervals.

Consider a population in which a specific trait appears w/ an unknown probability p.

Let $(Y_1, Y_2, \dots, Y_n)$ be a random sample from a Bernoulli dist'n w/ an unknown parameter p

Goal: To design a confidence interval for p.

Idea: A good point estimator is

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

Say, C = confidence level = $1-\alpha$

We need a pivotal quantity.

$$U = \frac{\bar{Y} - p}{\sqrt{\frac{p(1-p)}{n}}}$$ is still not a pivotal quantity.

But using the normal approximation to the binomial, we can construct an approximate confidence interval since, for "large enough" n

$$\boxed{U \approx N(0,1)}$$

z^* ... critical value of $N(0,1)$ at the conf. level C

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

Problem 14.2. 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?*

→:

$\hat{p} = 0.11$

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

$$p = 0.11 \pm 1.645 \cdot \sqrt{\frac{0.11 \cdot 0.89}{1000}}$$

$$\boxed{p = 0.11 \pm 0.01627636}$$

□

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

<u>Example.</u> How do we figure out the sample size w/ a given required margin of error m ?

$\boxed{n=?}$

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

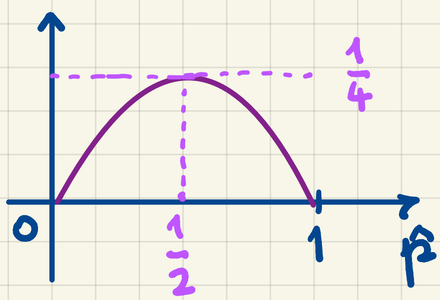
w/ z^* the critical value of $N(0,1)$ @ the confidence level C

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

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

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

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



The conservative choice for the sample size

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

Confidence Intervals for the Sample Variance.

Consider a normal model $N(\mu, \sigma)$ w/ both parameters unknown.

A good point estimator for σ^2:

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Theorem. Consider a random sample $Y_1, \dots, Y_n$ from $N(\mu, \sigma)$.
Let

$$\bar{Y} = \frac{1}{n}(Y_1 + \dots + Y_n)$$

and

$$Q^2 = \boxed{\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- $\bar{Y} \sim N(\text{mean} = \mu, \text{sd} = \frac{\sigma}{\sqrt{n}})$
- Q^2 is a pivotal quantity for σ^2; it has the $\chi^2(df = n-1)$
- $\boxed{\bar{Y} \text{ and } Q^2 \text{ are independent}}$

$\square$

Problem 14.3. *What is the unbiased estimator for σ^2.*

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

Problem 14.4. *Assume a random sample $Y_1, Y_2, \ldots, Y_n$ from a normal distribution with mean μ and standard deviation σ - both unknown. What's the distribution of*

$$\frac{\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2}{\sigma^2} = \boxed{\frac{(n-1)S^2}{\sigma^2}}?$$

$\sim \chi^2(df = n-1)$

$\boxed{\text{PIVOTAL QUANTITY}} = U$

Problem 14.5. *Assume that you are assigned a* confidence level $\boxed{1-\alpha.}$ *What does it mean to find a* confidence interval *for S^2?*

$$\mathbb{P}[a \leq U \leq b] = 1-\alpha$$

$$\mathbb{P}\left[\chi^2_L \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_R\right] = 1-\alpha$$

Problem 14.6. *Are $\hat{\chi}^2_L$ and $\hat{\chi}^2_U$ as above uniquely defined?*