ADM2304B

September 2015

Applied Statistical Methods in Business

Lecture: on Mondays 19:00-22:00 in CBY-B205

DGD: Please Check the Course Outline For Details

Text Book: Business Statistics, A-W Pearson, 2014

Authors: Sharpe et al(DVW) Canadian Ed# 2.

Prof.: Dr. Suren Phansalker

Office: DMS 5142

Office Hours: Thursdays 13:00 – 14:00 hrs

(By Tuesday morning, write me an e-mail and let me know you would like to see me.)

Introduction: Proportion
Dr. Suren Phansalker

- **1. Sample Proportion:** $\hat{p}$

  Here $\hat{p} = \dfrac{X}{n}$ and $\hat{q} = (1 - \hat{p})$, X~ b(n, p, x) where 'p' is the population proportion. Note that $\hat{p}$ is found with X as the # of successes and n as the Sample Size. Please Note: Many times $\hat{p}$ & $\bar{p}$ are written inter-changeably.

- **2. CLT for** $\hat{p}$

  If $np > 10$, $nq > 10$ then

  $$\hat{p} \sim N\left(p, \sigma^2(\hat{p}) = \frac{pq}{n} \right)$$

  Or, $$\hat{p} \sim N\left(p, \sigma(\hat{p}) = \sqrt{\frac{pq}{n}} \right)$$

- The CLT for the Sample Proportion, states that when the product of the sample size and p and sample size and q and both products are more than 10, the $\hat{p}$ has a Normal Distribution with mean given by 'p' the population proportion and standard deviation given by $\sqrt{\frac{pq}{n}}$ This theory is very useful if 'p' and 'q' are known. Please Refer to other files in "One Population Proportion" folder.
- **3. Example 1 for CLT:**

  Recently, in an election in a foreign country, it was found that when 100 people were checked at random on the voters' list, 90 seemed to have voted. It was well known from past experience that 80% people generally voted. What can you say about what was found?

Here p = 80% = 0.80 and q = (1 – p) = (1 – 0.80)

Thus, q = 0.20

Hence (n p) = 100 (0.80) = 80 > 10,

and (n q) = 100 (0.20) = 20 > 10.

Hence CLT for Sample Proportion applies.

$$\hat{p} \sim N\left(p, \sigma(\hat{p}) = \sqrt{\frac{pq}{n}} \right)$$

Now $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.80(0.20)}{100}} = \sqrt{0.0016} = 0.04$

Here $\hat{p} = \frac{X}{n} = \frac{90}{100} = 0.90$

- Thus:

$$P[\hat{p} > 0.90] = P\left[Z > \left(\frac{\hat{p} - p}{\sigma(\hat{p})} = \frac{0.90 - 0.80}{0.04} = 2.50\right)\right]$$

P[Z > 2.50] = 1 – P[Z < 2.50]

P[Z > 2.50] = 1 – 0.993790

= 0.00621 (6.21 in 1000)

This is a rather small probability; all though it is possible, it is not very probable! It is likely that there was a massive fraud in voting or past voting patterns were an unreliable guide.

Please notice that the ‘Z’ value was 2.50, a rather large value. The larger is the ‘Z’ value, the less probable is the event!

- **4. Confidence Intervals (CI):**

From Example 1, the question that needs to be answered is, if the Sample Proportion, $\hat{p} = 0.90$, and n = 100, then what must be the Population Proportion, ‘p’?

When we answer this question, we would have found the Confidence Interval for ‘p’ the Population Proportion.

Confidence Intervals are generally Symmetrical. Occasionally, One-Sided Confidence Intervals are found; but in this course we will stay with the Symmetrical Confidence Intervals (CI)

Confidence Interval is defined by the following relation:

$\hat{p} \pm z^{*} SE(\hat{p})$ This mean that:

$$\hat{p} - z^{*} SE(\hat{p}) < p < \hat{p} + z^{*} SE(\hat{p})$$

- Please notice that the mathematical expression defines an interval, the Confidence Interval; the Sample Proportion $\hat{p}$ minus 'something' to $\hat{p}$ plus 'something'. This 'something' is called the Margin of Error of $\hat{p}$ or ME($\hat{p}$) . Here the

  ME($\hat{p}$) = $z^* SE(\hat{p})$ where

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$ the Standard Error, SE.

  Since the true Standard Deviation of $\hat{p}$, is given by:

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}}$$

, it is impossible to find, since 'p' is unknown. Instead we use $\hat{p}$ to find the Standard Error, SE.

Here z^*, the critical value of z, is sometimes called $z_{\alpha/2}$ where α is known as Level of Significance (LS) and $(1 - \alpha)$ is known as the Confidence Coefficient (CC).

The following table will make some of these 'z' values obvious.

$CC = (1 - \alpha)$	$LS = \alpha$	$z^* = z_{\alpha/2}$
90%(0.90)	10%(0.10)	1.645
95%(0.95)	5%(0.05)	1.960
98%(0.98)	2%(0.02)	2.326
99%(0.99)	1%(0.01)	2.576

It is easy to check that based on symmetry:

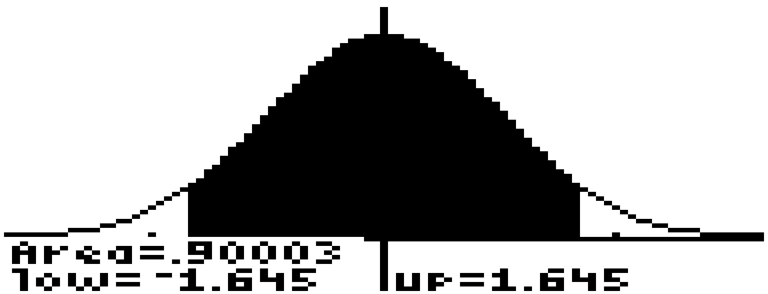
$P[-1.645 < Z < 1.645] = 0.90$

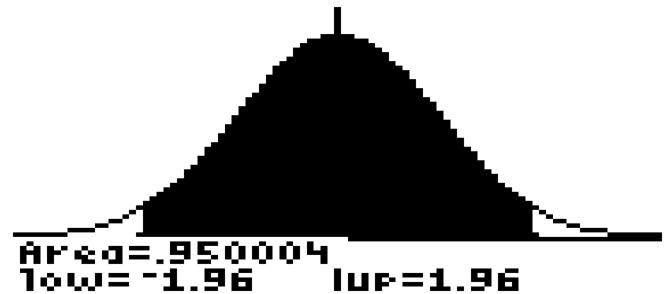
$P[-1.960 < Z < 1.960] = 0.95$

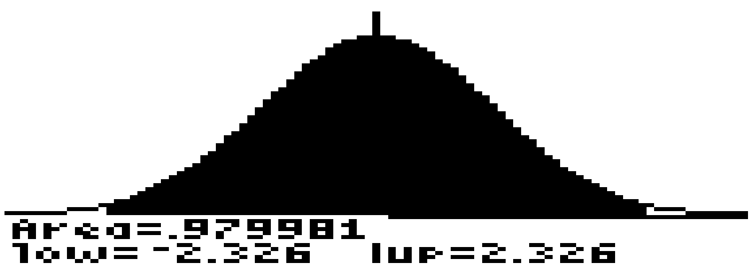
$P[-2.326 < Z < 2.326] = 0.98$, and

$P[-2.576 < Z < 2.576] = 0.99$.

It is obvious that the higher is the value of CC, higher is the value of $z^* = z_{\alpha/2}$. Some typical Symmetrical Areas of the Normal Curve are shown below.







- **5. Example 2 for CI:**

  Given that Sample Size was 100 and Sample Proportion was 0.90, find the 90% Confidence Interval for the Population Proportion.

  Solution:

  Given: n = 100, $\hat{p} = 0.90$, and CC = 90% or $z^* = z_{Crit} = 1.645$,

  Find:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{0.90(0.10)}{100}} = \sqrt{0.0009} = 0.03$$

  CI: $\hat{p} - z^* SE(\hat{p}) < p < \hat{p} + z^* SE(\hat{p})$

  Thus, CI: $0.90 - 1.645(0.03) < p < 0.90 + 1.645(0.03)$

  Or, $0.85065 < p < 0.94935$

  What does this mean?

- **The Correct Interpretation of CI:**

  If 100 samples were taken, and 100 Confidence Intervals were found by using the specified mathematical technique, then 90 of these Confidence Intervals (90% CC), would in fact contain the true population proportion, ‘p’.

  N.B.: Often students misinterpret the meaning of Confidence Interval: it is NOT the probability that the population proportion will lie anywhere within a given Confidence Interval. Remember, population proportion is a ‘Unique Parameter’ with a unique value and it most certainly is NOT a statistic with some probability value attached to it!

- **6. Confidence Intervals with Different Values of CC:**

For the same data in Example 2 for CIs,

CC	$z^* = z_{Crit}$	CI
90%	1.645	0.85065 to 0.94935
95%	1.960	0.84120 to 0.95800
98%	2.326	0.83022 to 0.96978
99%	2.576	0.82272 to 0.97728

Please note that the width of the Confidence Interval becomes bigger and bigger as the CC becomes higher and higher.

Question: Why is it that the 100% Confidence Interval has NO meaning?

- **7. Sample Size:**

What should be the sample size, if the Margin of Error, ME is 2% and the Confidence Coefficient is 95%.

Solution: $ME(\hat{p}) = ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} ==> n = \frac{(z^*)^2 \hat{p}\hat{q}}{(ME)^2}$

Since no sample has been taken, it is not possible to know the values of $\hat{p}$ or $\hat{q}$. In the event that Case I and Case II (see below) arguments cannot be justified, this equation is used as a necessity. This gives rise to two distinct possibilities.

Case 1: Historical Value of population proportion is unknown. Then in the worst case scenario, the maximum value of (this can be proved with Differential Calculus)

$$\hat{p}\hat{q} = 0.5(0.5) = 0.25$$

- Thus,

$$n = \frac{(z^*)^2 0.25}{(ME)^2} = \frac{(z^*)^2}{4(ME)^2} = \left(\frac{z^*}{2(ME)}\right)^2 = \left(\frac{1.96}{2(0.02)}\right)^2$$

$$n = 49^2 = 2401$$

Case 2: Historical Value of population proportion is known: In our case p = 0.80. We assume that this will be the value that can be used.

$$n = \frac{(z^*)^2 \hat{p}\hat{q}}{(ME)^2} = \frac{(z^*)^2 pq}{(ME)^2} = \frac{(1.96)^2 0.8(0.2)}{(0.02)^2} = 1536.64 \approx 1537$$

Please Note that in Statistics, 'n' is always rounded up!

N.B.: The same situation would be true if the Hypothesis Test for "H_0: p = 0.8" could not be rejected. Many times assuming $\hat{p} = \hat{q} = 0.5$ may not be reasonable.