

ADM2304X

Summer 2011

Statistics for Management II

Location: DMS-1120

Day and Time:

Section X: Thursdays 19:00 – 22:00

**Text Book: Business Statistics, A-W Pearson, 2010
(Custom Edition for UOttawa)**

Authors: Sharpe et al

Prof.: Dr. Suren Phansalker

Office: DMS 5144

Office Hours: Tuesdays -13:30-14:45

(Write me an e-mail and let me know before you come.)

Lecture#1A: Recpitulation

Sample Mean Distributions: 'Z' & 't' Distributions

Dr. Suren Phansalker

- **Central Limit Theorem (CLT):**

As seen before, P-S. Laplace proved the main assertion of the CLT. However, in its modern form, it does have different forms. The following three major cases bring out the variations.

- **Case I of CLT:** If a large sample with size, $n \geq 30$, is drawn from any much larger population 'X', of unknown distribution, then:

$$\bar{X} \sim N\left(E(\bar{X}), \sigma^2(\bar{X})\right) \quad \text{But, } E(\bar{X}) = \mu \text{ and } \sigma^2(\bar{X}) = \frac{\sigma^2}{n}$$

where μ , σ , and σ^2 are the population parameters.

Then simply written:
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

It sometimes is written as:
$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- **Case II of CLT:**

If the Population RV, 'X' is itself Normally Distributed, **then if 'σ' is known**, then for any sample size, (even 'n' < 30):

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{or} \quad \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- **Case III of CLT:**

If the Population RV, 'X' is itself Normally Distributed, **then if 'σ' is unknown and must be estimated by 's'**, the sample standard deviation, **then:**

$$\bar{X} \sim t_{n-1}\left(\mu, \frac{s^2}{n}\right) \quad \text{or} \quad \bar{X} \sim t_{n-1}\left(\mu, \frac{s}{\sqrt{n}}\right)$$

- **Special Condition for Case III of CLT:**

If 'n', the sample size is fairly large (n ≥ 30 or n ≥ 120), then:

$$\bar{X} \sim t_{n-1}\left(\mu, \frac{s^2}{n}\right) \approx N\left(\mu, \frac{s^2}{n}\right) \quad \text{or} \quad \bar{X} \sim t_{n-1}\left(\mu, \frac{s}{\sqrt{n}}\right) \approx N\left(\mu, \frac{s}{\sqrt{n}}\right)$$

In other words, the 't' Distribution and 'Z' Distribution become almost equal.

- **Some Examples of the CLT:**

- **Example 1 for Case I:**

A large sample of size 100 is taken from a population of marks on a standardized test with unknown distribution. The population mean is 550 marks and the standard deviation is 80 marks. Find the probability of the following events:

a. $P[\text{Sample Mean} < 525 \text{ marks}]$

b. $P[\text{Sample Mean} > 574]$

c. $P[534 < \text{Sample Mean} < 566]$

d. If $P[\text{Value1} > \text{Sample Mean} > \text{Value2}] = 0.0026$, what are the two symmetrical values?

Solution 1: $E[\bar{X}] = \mu = 550$ and $\sigma[\bar{X}] = \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{100}} = 8$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N(550, 8)$$

a. $P[\bar{X} < 525] = P\left[Z < \frac{525 - 550}{8} = -3.125\right] = 0.5 - 0.49915 = 8.50 \times 10^{-4}$

$$\text{b. } P[\bar{X} > 574] = P\left[Z > \frac{574-550}{8} = 3.000\right] = 1 - 0.9987 = 0.0013$$

$$\text{c. } P[534 < \bar{X} < 566] = P\left[\frac{534-550}{8} < Z < \frac{566-550}{8}\right] = P[-2.0 < Z < 2.0]$$

Draw the diagram and with, A(z), the Table Values

$$\begin{aligned} P[-2.0 < Z < 2.0] &= A[Z = -2.0] + A[Z = 2.0] = 2 * A(Z = 2) \\ &= 2(0.4772) = 0.9544 \text{ or } 95.44\% \end{aligned}$$

$$\text{d. } P[\text{Value}_1 < \text{Sample Mean} < \text{Value}_2] = 0.9974 \text{ (99.74\%)}$$

Since, $P[0 < Z < 3] = A(z = 3) = 0.4987$, $P[-3.0 < Z < 3.0] = 0.9974$,

$$3 = \frac{\text{Value}_2 - \mu}{\sigma(\bar{X})} = \frac{\text{Value}_2 - 550}{8} \implies \text{Value}_2 = 550 + 3(8) = 574$$

Similarly, $\text{Value}_1 = 550 - 3(8) = 526$

Hence $P[526 > \text{Sample Mean} > 574] = 0.0026$

- **Example 2, for Case II:**

Assume that the standardized test marks are normally distributed, (which generally they are) and have the same $\mu = 550$ marks and $\sigma = 80$ marks as before. But now, the sample size is only 25. Find:

a. $P[\text{Sample Mean} < 525]$

b. $P[\text{Sample Mean} > 574]$

c. $P[534 < \text{Sample Mean} < 566]$

d. $P[\text{Value1} > \text{Sample Mean} > \text{Value2}] = 0.0026$

Solution 2: $E[\bar{X}] = \mu = 550$ and $\sigma[\bar{X}] = \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{25}} = 16$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N(550, 16)$$

a. $P[\bar{X} < 525] = P\left[Z < \frac{525 - 550}{16} = -1.5625\right] = 0.5 - A(z = 1.5625) = 0.0591$

Please note that a smaller sample size of 25 leads to a higher standard deviation of sample mean (16), with all the expected consequences. The Intervals become broader, the probabilities become larger for similar situations.

b. $P[\text{Sample Mean} > 574] \rightarrow$

$$P[\bar{X} > 574] = P\left[Z > \frac{574 - 550}{16} = 1.500\right] = 0.5 - A(z = 1.5) = 0.0668 (6.68\%)$$

c. $P[534 < \text{Sample Mean} < 566] \rightarrow$

$$P[534 < \bar{X} < 566] = P\left[\frac{534 - 550}{16} < Z < \frac{566 - 550}{16}\right] = P[-1.0 < Z < 1.0]$$

$$P[-1.0 < Z < 1.0] = A[Z = -1.0] + A[Z = 1.0]$$

$$= 2 * A(z = 1) = 2 * 0.3413 = 0.6826 \text{ or } 68.26\%$$

d. $P[\text{Value}_1 > \text{Sample Mean} > \text{Value}_2] = 0.0026 \rightarrow$

$$P[\text{Value}_1 < \text{Sample Mean} < \text{Value}_2] = 1 - 0.0026 = 0.9974$$

Since $P[0 < Z < 3] = A(z = 3) = 0.4987$, $P[-3.0 < Z < 3.0] = 0.9974$

$$3 = \frac{\text{Value}_2 - \mu}{\sigma(\bar{X})} = \frac{\text{Value}_2 - 550}{16} \implies \text{Value}_2 = 550 + 3(16) = 598$$

Similarly, $\text{Value}_1 = 550 - 3(16) = 502$, and, therefore,

$$P[502 > \text{Sample Mean} > 598] = 0.0026$$

- **Example 3 for Case III:**

Assume that the standardized test marks are normally distributed (which generally they are) and have the same $\mu = 550$ marks. But now assume that the standard deviation, ‘ σ ’, is unknown. A sample of size 25 is drawn and the sample standard deviation, ‘ s ’, is found to be 100 marks. Find: (MiniTab ‘CDF’ command has been used.)

a. $P[\text{Sample Mean} < 525]$

b. $P[\text{Sample Mean} > 574]$

c. $P[534 < \text{Sample Mean} < 566]$

d. $P[\text{Value1} > \text{Sample Mean} > \text{Value2}] = 0.0026$

Solution 3: $E[\bar{X}] = \mu = 550$ and $s[\bar{X}] = \frac{s}{\sqrt{n}} = \frac{100}{\sqrt{25}} = 20$

$$\bar{X} \sim t_{n-1} \left(\mu, \frac{s}{\sqrt{n}} \right) = t_{24} (550, 20)$$

a. $P[\bar{X} < 525] = P \left[t_{24} < \frac{525 - 550}{20} = -1.250 \right] = 0.1117$

Please note that this probability is higher than 0.0591 obtained in Example II. ‘ t ’ distribution is broader than the ‘ Z ’ distribution!

b. $P[\text{Sample Mean} > 574] \rightarrow$

$$P[\bar{X} > 574] = P\left[t_{24} > \frac{574-550}{20} = 1.200\right] = 1 - 0.8883 = 0.1117(11.17\% !)$$

c. $P[534 < \text{Sample Mean} < 566] \rightarrow$

$$P[534 < \bar{X} < 566] = P\left[\frac{534-550}{20} < t_{24} < \frac{566-550}{20}\right] = P[-0.80 < t < .80]$$

$$\begin{aligned} P[-0.80 < t_{24} < 0.80] &= P[t_{24} < 0.80] - P[t_{24} < -0.80] \\ &= 0.7842 - 0.2158 = 0.5684 \end{aligned}$$

d. $P[\text{Value}_1 > \text{Sample Mean} > \text{Value}_2] = 0.0026 \rightarrow$

$$P[\text{Value}_1 < \text{Sample Mean} < \text{Value}_2] = 0.9974$$

Use: MTB> invcdf 0.0013;

Subc> t 24. This will give you -3.3601 , or

Use MTB> invcdf 0.9987;

Subc> t 24. This will give a symmetrical value of 3.3601 .

Since $P[-3.3601 < t_{24} < 3.3601] = 0.9974$

Thus Value2 can be found by:

$$3.3601 = \frac{\underline{Value}_2 - \mu}{s(X)} = \frac{\underline{Value}_2 - 550}{20} \implies \underline{Value}_2 = 550 + 3.3601(20) = 617.2020$$

Similarly, $\underline{Value}_1 = 550 - 3.3601(20) = 482.7980$

Hence $P[482.798 > \text{Sample Mean} > 617.202] = 0.0026$

Please note that by using ‘s’ and the consequent ‘t’ distribution, all the relevant probability values and the intervals have become even broader.

- Please see the ‘Week1AMTB2304.pdf’ file to know how you can use the ‘CDF’ command in MiniTab. Also see pages 11 & 13 of this file.

- **Example 4 for Special Condition in Case III:**

All the data in this example is exactly the same except as in Example 3 with one exception. The sample size, 'n', is now 100.

All the Probabilities to be found are exactly the same. DF = 99.

Solution 4: $E[\bar{X}] = \mu = 550$ and $s[\bar{X}] = \frac{s}{\sqrt{n}} = \frac{100}{\sqrt{100}} = 10$

$$\bar{X} \sim t_{n-1} \left(\mu, \frac{s}{\sqrt{n}} \right) = t_{99} (550, 10)$$

a. $P[\bar{X} < 525] = P \left[t_{99} < \frac{525 - 550}{10} = -2.50 \right] = 0.0070$

Please note: MTB> cdf -.25;

Subc> t 99. This gives you 0.0070

Also, MTB> cdf -2.5;

Subc> normal 0 1. This gives you 0.0062. These

values are pretty close.

Also, MTB> cdf -2.5;

Subc> t 120. This gives you 0.0069, almost the same as 'Z'.

b. $P[\text{Sample Mean} > 574] \rightarrow$

$$P[\bar{X} > 574] = P\left[t_{99} > \frac{574-550}{10} = 2.400\right] = 1 - 0.9909 = 0.0091 (0.9132\% !)$$

Please note:

$$P[\bar{X} > 574] = P\left[Z > \frac{574-550}{10} = 2.400\right] = 0.5 - 0.4918 = 0.0082 (0.8198\% !)$$

The values obtained by using 'Z' distribution and 't' distribution are pretty close. This is so because 'n' is reasonably large and the degrees of freedom, df, are getting to be large too. The degrees of freedom, here, are given by: $df = n - 1 = 99$

When $df \geq 120$, t-values \approx z-values. This is the special condition!

c. $P[534 < \text{Sample Mean} < 566] \rightarrow$

$$P[534 < \bar{X} < 566] = P\left[\frac{534-550}{10} < t_{99} < \frac{566-550}{10}\right] = P[-1.60 < t < 1.60]$$

$$\begin{aligned} P[-1.60 < t_{99} < 1.60] &= P[t_{99} < 1.6] - P[t_{99} < -1.6] \\ &= 0.9436 - 0.0564 = 0.8872. \end{aligned}$$

N.B.: $\{P[-1.6 < Z < 1.6] = 0.8904\} \approx \{P[-1.6 < t_{99} < 1.6] = 0.8872\}!$

d. $P[\text{Value}_1 > \text{Sample Mean} > \text{Value}_2] = 0.0026 \rightarrow$

$$P[\text{Value}_1 < \text{Sample Mean} < \text{Value}_2] = 0.9974$$

Use: MTB> invcdf 0.9987;

Subc> t 99. This will give you 3.0898

$$\text{Thus } P[-3.0898 < t_{99} < 3.0898] = 0.9974$$

As you remember, $P[-3.0 < Z < 3.0] = 0.9974$

Obviously $\{t_{99} = 3.0898\} \approx \{Z = 3.0\}$. Thus,

$$3.0898 = \frac{\text{Value}_2 - \mu}{s(\bar{X})} = \frac{\text{Value}_2 - 550}{10} \implies \text{Value}_2 = 550 + 3.0898(10) = 580.898$$

Similarly, $\text{Value}_1 = 550 - 3.0898(10) = 519.102$

$$\text{Thus } P[519.102 > \text{Sample Mean} > 580.898] = 0.0026$$

Had you used the ‘Z’ distribution, you would have got:

$P[520 > \text{Sample Mean} > 580] = 0.0026$, almost the same interval as you obtained by using the ‘t’ distribution, “the limiting behavior!”