GE2262 Business Statistics

Topic 5 Confidence Interval Estimation for Population Mean

Lecturer: Dr. Iris Yeung
Room : LAU-7239
Tel No.: 34428566
E-mail: msiris@cityu.edu.hk

Outline

- Confidence Interval Estimation for the Population Mean
- Determining Required Sample Size for Estimating Mean

Reference

Levine, D.M., Krehbiel, T.C. and Berenson, M.L., *Business Statistics: A First Course*, Pearson Education Ltd, Chapter 8

Part One

- Confidence Interval Estimation for the Population Mean
- Determining Required Sample Size for Estimating Mean

Population Parameter, Sample Statistic and Inferential Statistics

- A population contains **all** the items or individuals about which we want to study
- A sample contains only a **portion** of the population of items or individuals
- A variable is a characteristic of an item or individual
- A population parameter summarizes the value of a specific variable for a population
- A sample statistic summarizes the value of a specific variable for sample data
- Two types of Statistics
  - Descriptive Statistics (Topic 1)
    - to describe, summarize and present data via tables, graphs, and summary measures
  - Inferential Statistics (Topics 5 – 7)
    - to infer, conclude, and make decisions about a large group (population) from a small group (sample).

Two Types of Inferential Statistics

- Estimation
  - Estimate the unknown population parameter
  - Examples
    - we want to estimate the mean waiting time of bank service, …
    - We want to estimate the proportion of customers being satisfied with bank service
- Hypothesis Testing
  - Test whether a hypothesis (claim or statement) about the population parameter holds or not
  - Example: suppose a bank manager claims that (1) the mean waiting time for their service is no more than 10 mins and (2) the proportion of customers being satisfied with their service is at least 0.9. We want to estimate whether the manager's claims hold or not

Measure	Population parameter	Sample statistic	Lecture
Mean	μ	$\bar{X}$	Topic 5 (estimation) Topic 6 (hypothesis testing)
Proportion	p	$\hat{p}$	Topic 7 (estimation and hypothesis testing)

Sampling Distribution of the Sample Mean

Sampling with replacement from Population (24, 26, 28, 30)

Sample	Sample members	Sample data values	Sample mean	Prob
1	A,A	24,24	24	0.0625
2	A,B	24,26	25	0.0625
3	A,C	24,28	26	0.0625
4	A,D	24,30	27	0.0625
5	B,A	26,24	25	0.0625
6	B,B	26,26	26	0.0625
7	B,C	26,28	27	0.0625
8	B,D	26,30	28	0.0625
9	C,A	28,24	26	0.0625
10	C,B	28,26	27	0.0625
11	C,C	28,28	28	0.0625
12	C,D	28,30	29	0.0625
13	D,A	30,24	27	0.0625
14	D,B	30,26	28	0.0625
15	D,C	30,28	29	0.0625
16	D,D	30,30	30	0.0625

Sample Mean $(\bar{X})$	Probability $P(\bar{X})$	$\bar{X}P(\bar{X})$	$(\bar{X}-\mu_{\bar{X}})^2P(\bar{X})$
24	0.0625	1.5	0.5625
25	0.125	3.125	0.5
26	0.1875	4.875	0.1875
27	0.25	6.75	0
28	0.1875	5.25	0.1875
29	0.125	3.625	0.5
30	0.0625	1.875	0.5625
Total	1.000	27	2.5

- Mean value (expected value) of the sampling distribution of sample mean:

$$\mu_{\bar{X}} = \sum \bar{X}P(\bar{X})$$
$$= 24\left(\frac{1}{16}\right) + \cdots + 30\left(\frac{1}{16}\right) = 27$$

- Standard deviation of the sampling distribution of of the sample mean:

$$\sigma_{\bar{X}} = \sqrt{\sum(\bar{X}-\mu_{\bar{X}})^2P(\bar{X})}$$
$$= \sqrt{(24-27)^2\left(\frac{1}{16}\right) + \cdots + (30-27)^2\left(\frac{1}{16}\right)} = \sqrt{2.5} = 1.5811$$

- For sampling with replacement, or sampling from large population without replacement
  - Mean of the sampling distribution of $\bar{X}$ = Mean of the population
    - $\mu_{\bar{X}} = \mu$
  - Standard deviation of the sampling distribution of $\bar{X}$ (also called standard error of the mean)
    - $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
    - As n increases, $\sigma_{\bar{X}}$ decreases

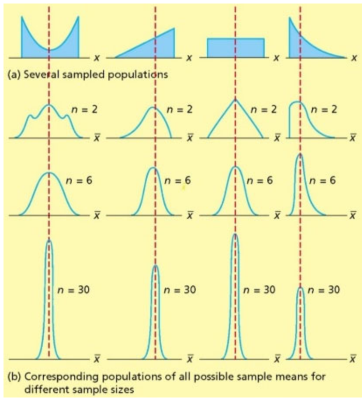
Sampling Distribution of the Sample Mean

- If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normally distributed with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
- $X \sim N(\mu, \sigma^2)$

  $\rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

- If the population is not normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ will be approximately normal as long as the sample size is large enough based on Central Limit Theorem
- X not normal but $n \geq 30$

  $\rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

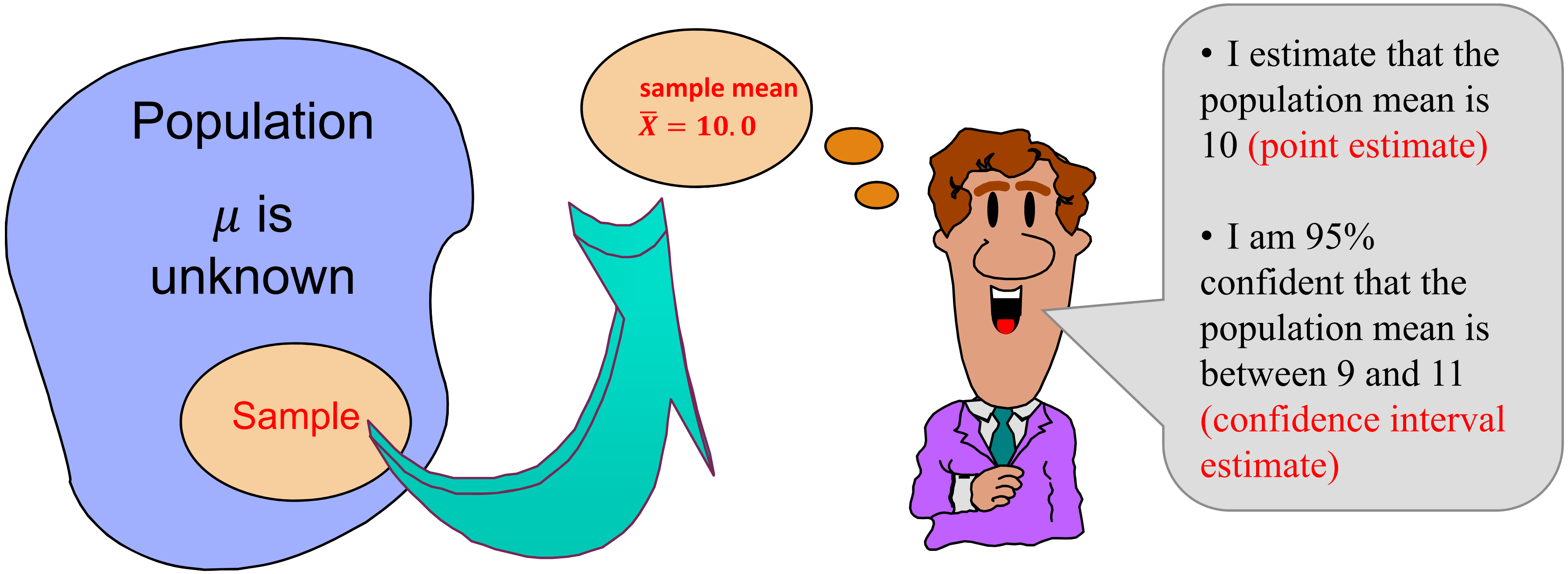


Two Types of Estimation

1. Define the Population

2. Select a Random Sample and Obtain the Sample Statistics

3. Estimate population parameters based on the sample statistics calculated from sample data

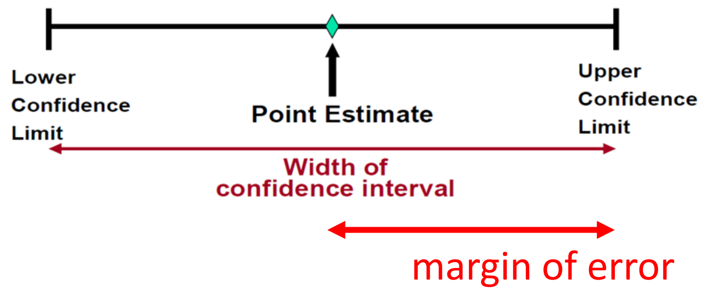


Point and Interval Estimation for Population Parameter

- Point estimation - use the value of a sample statistic (a single number) to estimate unknown population parameter
  - Example: use sample mean (10) to estimate population mean

	Population Parameter	Sample Statistic
Mean	μ	$\bar{X}$
Proportion	p	$\hat{p}$

- Confidence interval estimation - use a range (or an interval) of numbers to estimate unknown population parameter
  - Example: use an interval (9, 11) to estimate population mean and state the level of confidence
  - Confidence interval = point estimate ± margin of error
  - The level of confidence is $100(1-\alpha)\%$. Most common confidence levels are: 90% (α=0.10) , 95% (α=0.05) , and 99% (α=0.01). Note that it can never be 100% confident



$100(1-\alpha)\%$Confidence Interval Estimation for Population Mean μ

Population distribution	Sample size n	σ known	σ unknown
Normal	Large (n≥30)	$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ use z value if no t value for large n in the table
	Small (n<30)	$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$
Not normal	Large (n≥30) Due to central limit theorem	$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ use z value if no t value for large n in the table

Case 1: σ known, (Population normal or n large), use Z

Case 2: σ unknown, (Population normal or n large), use *t* (use *Z* if you cannot find *t* value for large *n* in the *t* table)

Confidence Interval for Population Mean μ (Case 1)

Assumptions

- Population standard deviation is known
- Population is normal
- If population is not normal, use large *n*

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$P\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1-\alpha$$

$$\rightarrow P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\rightarrow P\left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \ge \mu-\bar{X} \ge -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\rightarrow P\left(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$100(1-\alpha)\%$ Confidence interval estimate for population mean is:

$$\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$2 < 5 \rightarrow -2 > -5$

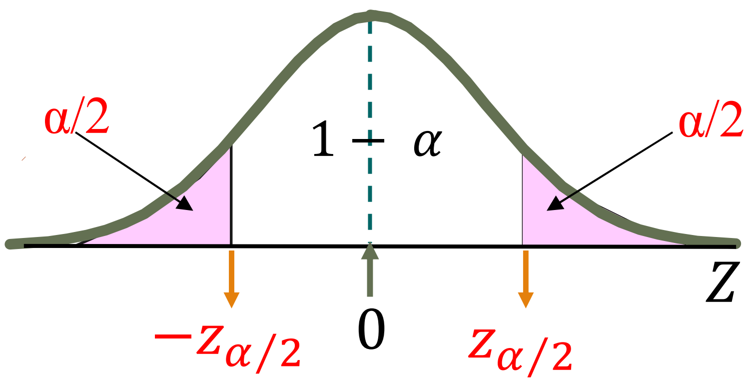
$$(\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

The inequality can be written as two inequalities

$$\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{X} \text{ and } \bar{X} \le \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \mu \le \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ and } \mu \ge \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$



Confidence Interval for Population Mean μ (Case 1)

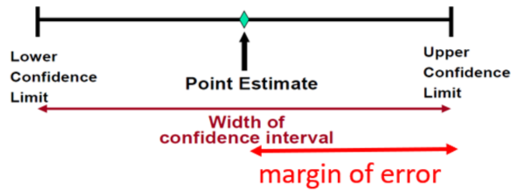
- $100(1-\alpha)\%$ Confidence interval estimate

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

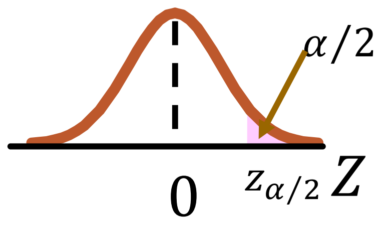
Standard Error, $\sigma_{\bar{X}}$

Sampling Error, E
(Also called margin of error)

Critical Value



- Confidence interval = point estimate ± margin of error
- Z-value (Critical Value)
  - $z_{\alpha/2}$ is the value corresponding to an upper-tail probability of $\alpha/2$ from the standardized normal distribution
  - $z_{\alpha/2}$ is based on the confidence level $100(1-\alpha)\%$
- Standard Error $\sigma_{\bar{X}}$ is the standard deviation of the sample statistic (sample mean $\bar{X}$)



- Sampling Error (Margin of Error)
  - Half of the width of the confidence interval
  - Sampling error measures how far off the estimation results are likely to be from the result that they would have gotten if the entire population are surveyed instead of merely a sample
  - $E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Example 1: Meaning of Confidence Interval for μ

Consider a **normal population** with μ=368, **σ=15**.

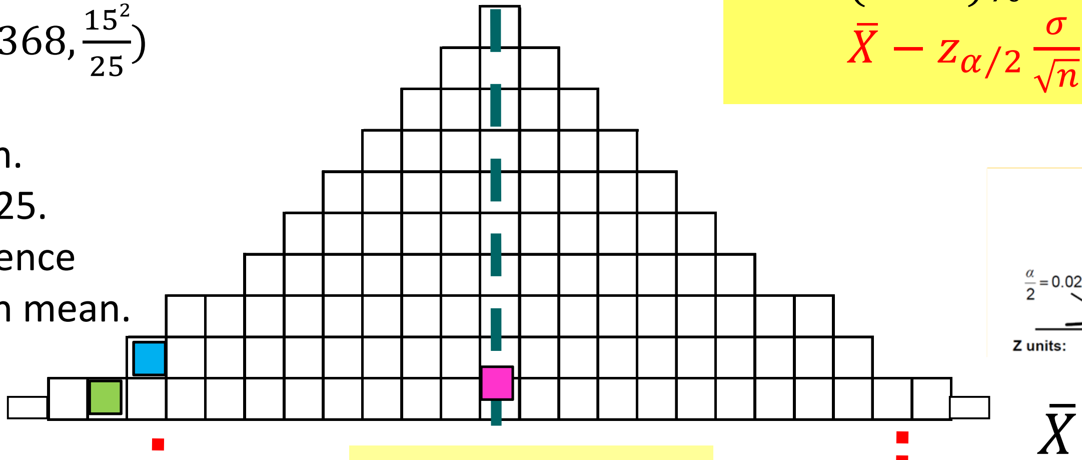
$X \sim N(\mu, \sigma^2) \rightarrow \bar{X} \sim N(368, \frac{15^2}{25})$

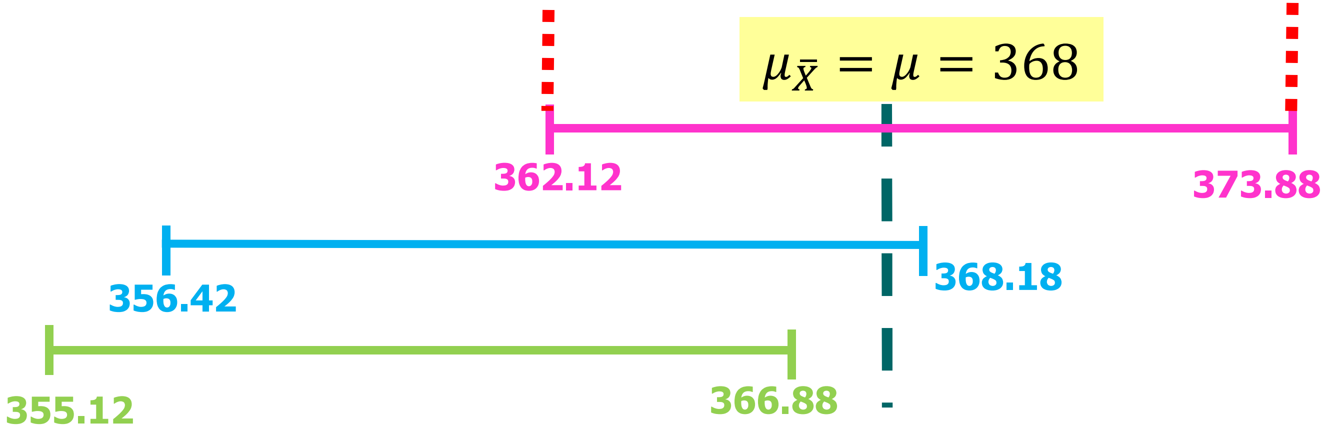
Suppose μ is unknown.
Take a sample of size 25.
Construct 95% confidence interval for population mean.

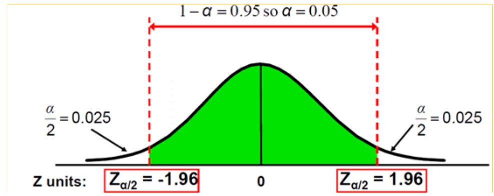
Case 1

$100(1-\alpha)\%$ confidence interval:

$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ to $\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$







Sample 1: $\bar{X} = 368$
95% CI: $368 \pm 1.96 \times \frac{15}{\sqrt{25}}$
$= [362.12, 373.88]$

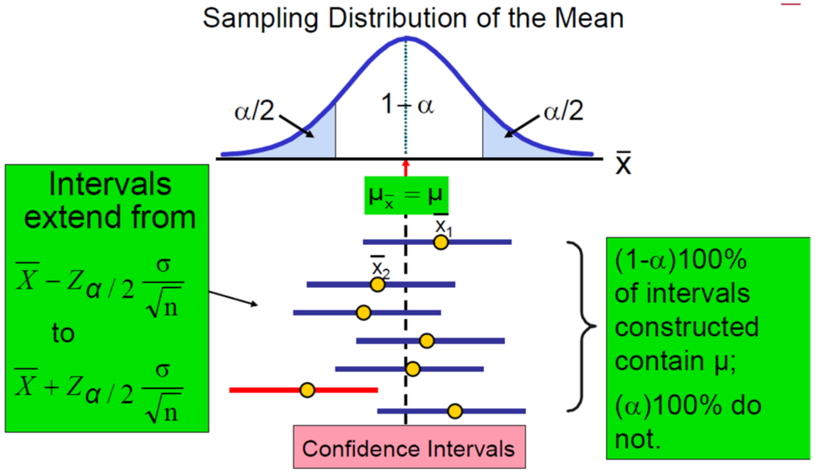
Sample 2: $\bar{X} = 362.3$
95% CI: $362.3 \pm 1.96 \times \frac{15}{\sqrt{25}}$
$= [356.42, 368.18]$

Sample 3: $\bar{X} = 361$
95% CI: $361 \pm 1.96 \times \frac{15}{\sqrt{25}}$
$= [355.12, 366.88]$

- If you repeat the sampling by 100 times, you will find that 95% of intervals so constructed cover μ; 5% do not
- Based on the one sample you actually selected, you can be 95% confident your interval will contain μ (this is a 95% confidence interval).

Meaning of Confidence Interval for Population Mean μ

100(1-α)% Confidence Interval
$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ to $\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$



- A relative frequency interpretation
  - In the long run, $100(1-\alpha)\%$ of all the confidence intervals that can be constructed will cover the unknown population parameter
- A conventional interpretation
  - We are $100(1-\alpha)\%$ confident that the unknown population parameter lies between $\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and $\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
    - Because the interval is obtained by a method that gives correct results $100(1-\alpha)\%$ of the time
- A specific interval will either cover or not cover the population parameter

Factors Affecting Interval Width (Precision)

100(1-α)% Confidence Interval
$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ to $\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

- Data variation
  - Measured by σ
  - As σ ↑→ $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ ↑→width of interval ↑
- Sample size
  - Measured by $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
  - n ↑→ $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ ↓→width of interval ↓
- Level of confidence
  - Measured by $100(1-\alpha)\%$
  - $(1-\alpha)$ ↑→ $|Z\text{-value}|$ ↑ →width of interval ↑
- $\bar{X}$ affects the location of the interval, but not the width

The Cumulative Standardized Normal Distribution (Continued)
Entry represents area under the cumulative standardized normal distribution from -∞ to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952

Confidence level	Confidence coefficient 1-α	$Z_{\alpha/2}$ value
90%	0.90	1.645
95%	0.95	1.96
99%	0.99	2.575

Example 2: σ Known, Population Normal (Case 1)

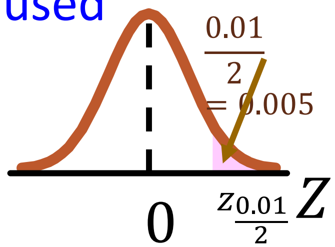
- A random sample of 15 stocks traded on the Hang Seng Index showed an average shares traded to be 215,000
- From the past experience, it is believed that the population standard deviation of shares traded is 195,000 and the shares traded are very close to a Normal distribution
- Construct a 99% confidence interval for the average shares traded on the Hang Seng Index. Interpret the result.

Example 2: σ Known, Population Normal (Case 1)

Since the population number of shares traded (X) follows Normal distribution, the distribution of sample means also follows Normal distribution, i.e.

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

With known population standard deviation (σ), Z distribution is used



99% confidence interval (C.I.) for μ

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 215000 \pm z_{0.01/2}\frac{195000}{\sqrt{15}}$$

$$= 215000 \pm 2.575\frac{195000}{\sqrt{15}} = [85351.88, 344648.12]$$

Confidence level	Confidence coefficient 1-α	$Z_{\alpha/2}$ value
90%	0.90	1.645
95%	0.95	1.96
99%	0.99	2.575

Interpretation

- If all possible samples of size 15 are taken and the corresponding 99% confidence intervals are constructed, 99% of these intervals will cover the unknown population mean
- We are 99% confident that the population average number of shares traded on the Hang Seng Index is between 85351.88 and 344648.12
- There is 99% chance that the unknown population mean will be in between 85351.88 and 344648.12

Confidence Interval for Population Mean μ (Case 2)

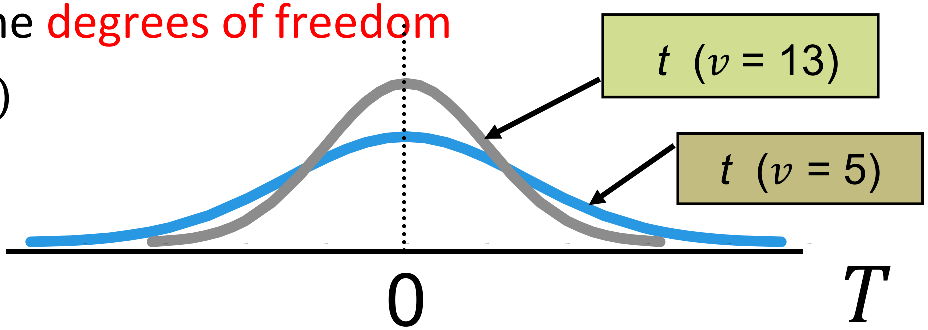
- Assumptions
  - Population standard deviation is unknown
  - Population is normal
  - If population is not normal, use large *n*
- The variable $\mathrm{T} = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ follows a distribution called Student's t-distribution (or simply called t-distribution)
  - Developed by Gosset who used "Student" as the pen name in pu the paper

- The probability density function of t-distribution is

$$f(t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})}(1+\frac{t^2}{v})^{-\frac{v+1}{2}} \quad , v > 0$$

where Γ is the gamma function and v is the parameter of the function which is often called the degrees of freedom

- Often denoted as $T \sim t(v)$



Degrees of Freedom (Df) in t-distribution

- Df is the number of observations that are free to vary in the final calculation of a statistic
- Df equals the total number of observations used in the analysis minus the number of parameters estimated as intermediate steps in the estimation of the parameter itself
- Example: Sample variance has $n - 1$ degrees of freedom
  - Suppose there are three numbers. Two variation values can be any numbers, but the third is not free to vary.
  - Since it is computed from n observations minus one parameter estimated (sample mean) as intermediate step. If we know the mean of three numbers is equal to 2, then the third number must be equal to 3

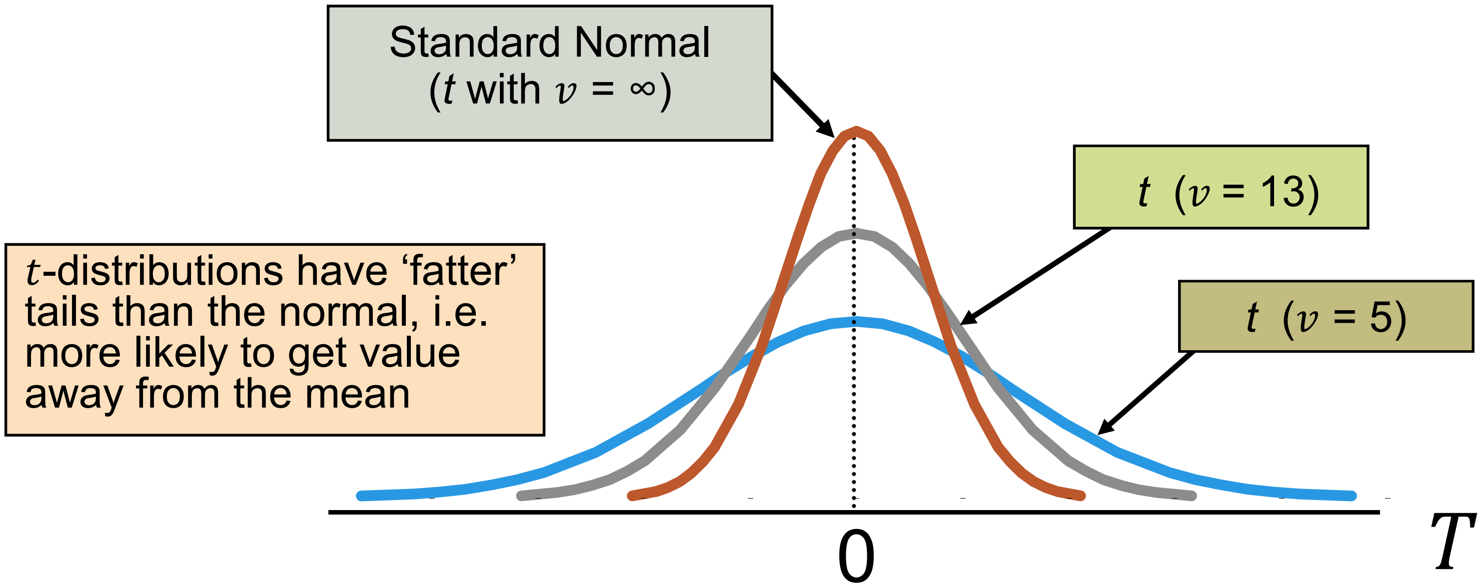
X	X-$\bar{X}$
1	-1
2	0
3	?
Total	**0**

$$\sum_i (X_i - \bar{X}) = 0$$

X
1
2
?=2*3-1-2
Mean = 2

Properties of Student's t-distribution with DF= ν

- Mean & Standard Deviation
  - Mean = 0 for v >1, otherwise it is undefined
  - Standard deviation = $\nu/(\nu - 2)$ for v > 2, = ∞ for $1 < v \leq 2$, otherwise undefined (Standard deviation >1)
- The shape of the density function
  - The theoretical range of T is infinite, i.e. $-\infty$ to $+\infty$
  - Bell shaped and symmetric about $zero$, but have fatter tails than normal
  - Median = mode = 0
  - As v increases, the density curve approaches the *N*(0, 1) curve



Confidence Interval for Population Mean μ (Case 2)

Assumptions

- Population standard deviation is unknown
- Population is normal
- If population is not normal, use large *n*

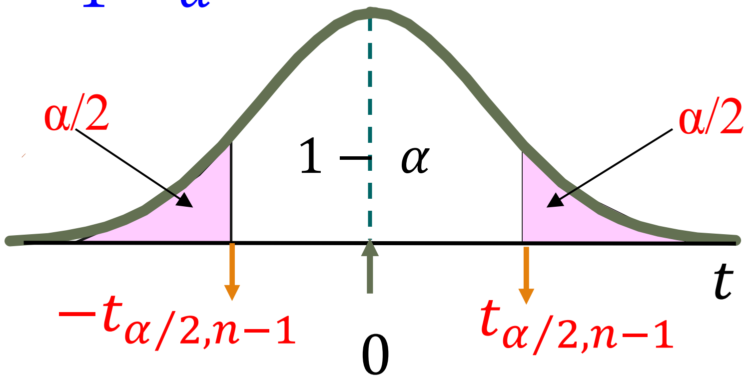
$$T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

$$P\left(-t_{\alpha/2,n-1} \le \frac{\bar{X}-\mu}{s/\sqrt{n}} \le t_{\alpha/2,n-1}\right) = 1-\alpha$$

$$\rightarrow P\left(-t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \bar{X}-\mu \le t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

$$\rightarrow P\left(t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \ge \mu-\bar{X} \ge -t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

$$\rightarrow P\left(\bar{X}-t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{X}+t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$



$100(1-\alpha)\%$ Confidence interval estimate for population mean is:

$$\bar{X}-t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{X}+t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

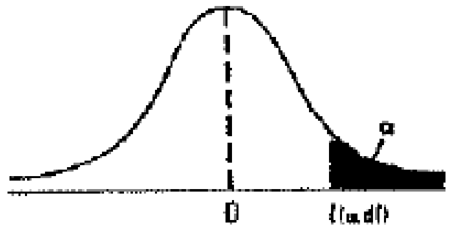
$$or \quad \bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

Student's t-distribution

The column gives the upper tail area

Critical Values of t

For a particular number of degrees of freedom, entry represents the critical value of t corresponding to a specified upper-tail area (α)



t Distribution: Critical Values of t

Degrees of freedom		*Significance level*					
	Two-tailed test:	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408

The row shows the degrees of freedom

The value within the table gives the t-value corresponding to a particular degrees of freedom and upper-tail area

At 7 degrees of freedom, $P(t > 2.365) = 0.025$

Comparing Student's t-distribution and Standard Normal Distribution

Confidence level	Confidence coefficient 1-α	$Z_{\alpha/2}$ value	$t_{\alpha/2}$ value (n=30)
90%	0.90	1.645	1.6991
95%	0.95	1.96	2.0452
99%	0.99	2.575	2.7564

Critical Values of t
For a particular number of degrees of freedom, entry represents the critical value of t corresponding to a specified upper-tail area (α)

Degrees of Freedom	Upper-Tail Areas 0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
110	0.6767	1.2893	1.6588	1.9818	2.3607	2.6213
120	0.6765	1.2886	1.6577	1.9799	2.3578	2.6174
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

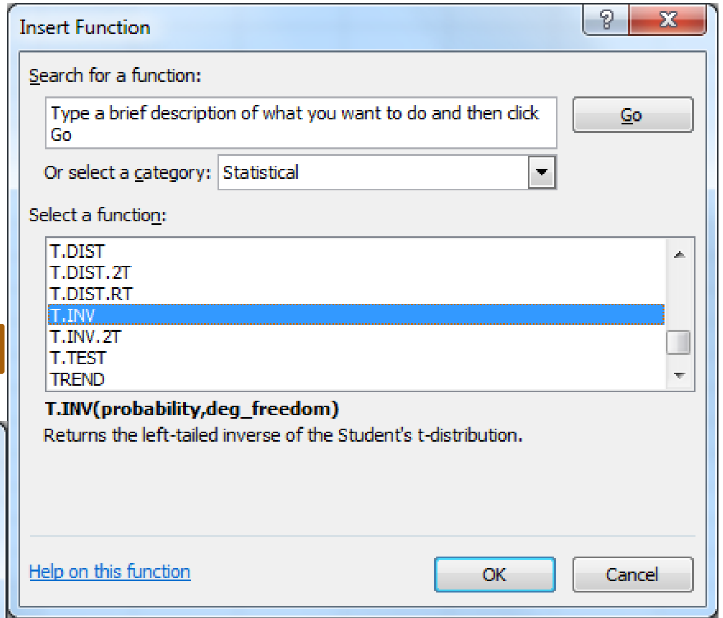
Finding t-Value in Excel

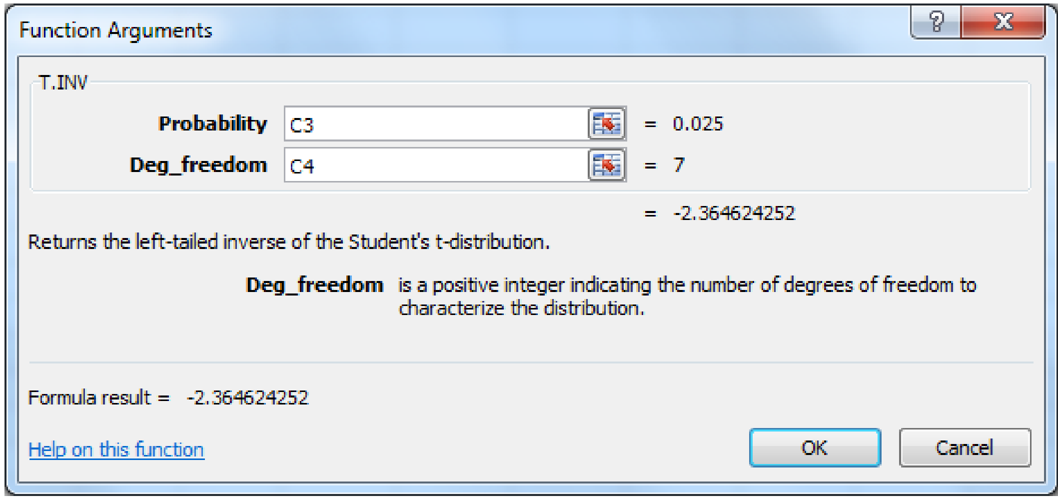
Step 1: Type the given information $(\frac{\alpha}{2}, df)$

Step 2: Insert the “T.INV” function

	A	B	C
1	t Distribution		
2			
3	Lower Tail Probability	α/2 =	0.025
4	Degrees of Freedom	df =	7
5			
6	t-Value =	-2.3646	

=T.INV(lower tail probability, df)





Example 3: σ Unknown, Population Normal, *n* Small (Case 2)

The monthly salary of brokers is found to be Normally distributed. A random sample of 25 brokers has a mean monthly salary HK$80K and a standard deviation of HK$16K. Set up a 95% confidence interval estimation for the population mean

95% confidence interval (C.I.) for μ

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} = 80 \pm t_{0.05/2,25-1}\frac{16}{\sqrt{25}}$$

$$= 80 \pm 2.064\frac{16}{\sqrt{25}} = [73.396, 86.604]$$

We are 95% confident that the population mean monthly salary of brokers is between HK$73.396K and HK$86.604K

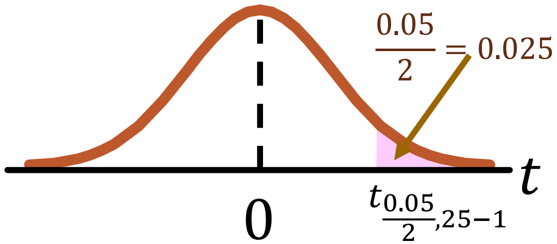


Table A.2

t* Distribution: Critical Values of *t

Degrees of freedom		Significance level			
	Two-tailed test:	10%	5%	2%	1%
	One-tailed test:	5%	2.5%	1%	0.5%
1		6.314	12.706	31.821	63.657
2		2.920	4.303	6.965	9.925
3		2.353	3.182	4.541	5.841
4		2.132	2.776	3.747	4.604
5		2.015	2.571	3.365	4.032
6		1.943	2.447	3.143	3.707
7		1.894	2.365	2.998	3.499
8		1.860	2.306	2.896	3.355
9		1.833	2.262	2.821	3.250
10		1.812	2.228	2.764	3.169
11		1.796	2.201	2.718	3.106
12		1.782	2.179	2.681	3.055
13		1.771	2.160	2.650	3.012
14		1.761	2.145	2.624	2.977
15		1.753	2.131	2.602	2.947
16		1.746	2.120	2.583	2.921
17		1.740	2.110	2.567	2.898
18		1.734	2.101	2.552	2.878
19		1.729	2.093	2.539	2.861
20		1.725	2.086	2.528	2.845
21		1.721	2.080	2.518	2.831
22		1.717	2.074	2.508	2.819
23		1.714	2.069	2.500	2.807
24		1.711	2.064	2.492	2.797
25		1.708	2.060	2.485	2.787

Example 4: σ Unknown, Population Distribution Unknown, *n* Large (Case 2)

- The branch manager of an outlet of a nationwide chain of pet supply stores want to study characteristics of her customers. In particular, she would like to estimate the population mean amount spent in the pet supply store. A random sample of 200 customers is selected. The sample mean of amount of money spent is \$21.34, and the sample standard deviation is \$9.22. Construct a 95% confidence interval estimate for the population mean amount spent in the pet supply store.

No knowledge about population distribution, *n*=200 large

95% confidence interval (C.I.) for μ

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} = 21.34 \pm 1.96 \frac{9.22}{\sqrt{200}} = [20.0622, 22.6179]$$

We are 95% confident that the population mean amount spent is between \$20.0622 and \$22.6179

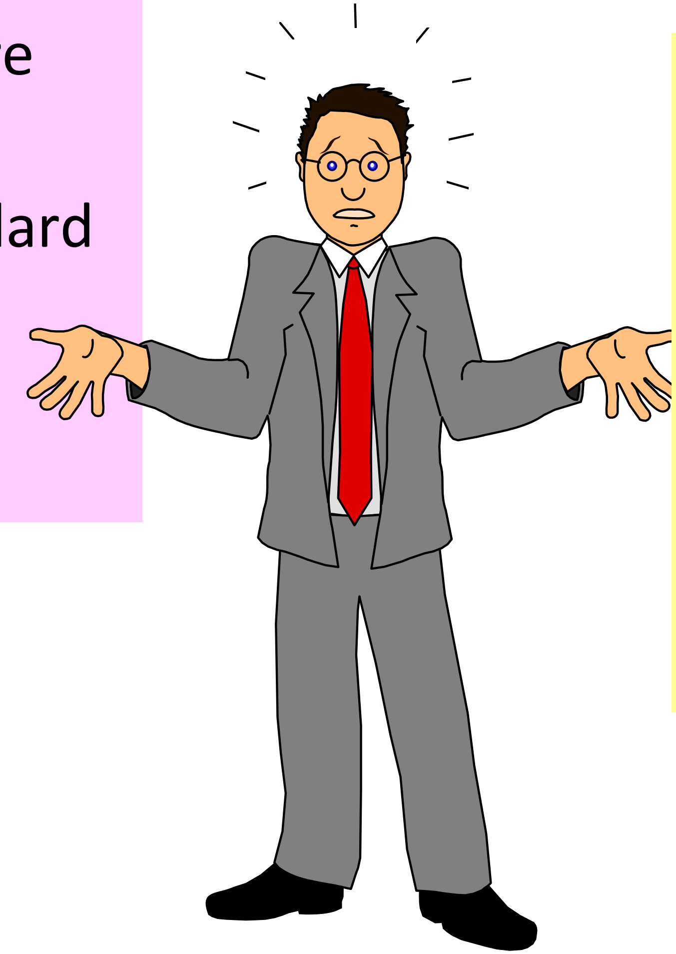
Degrees of Freedom	Upper-Tail Areas					
	0.25	0.10	0.05	0.025	0.01	0.005
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
110	0.6767	1.2893	1.6588	1.9818	2.3607	2.6213
120	0.6765	1.2886	1.6577	1.9799	2.3578	2.6174
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

Part Two

- Confidence Interval Estimation for the Population Mean
- Determining Required Sample Size for Estimating Mean

Determining Sample Size

- Large sample
  - Requires more resources
  - Smaller standard error (more precise estimation)



- Small sample
  - Requires less resources
  - Larger standard error (Less precise estimation)

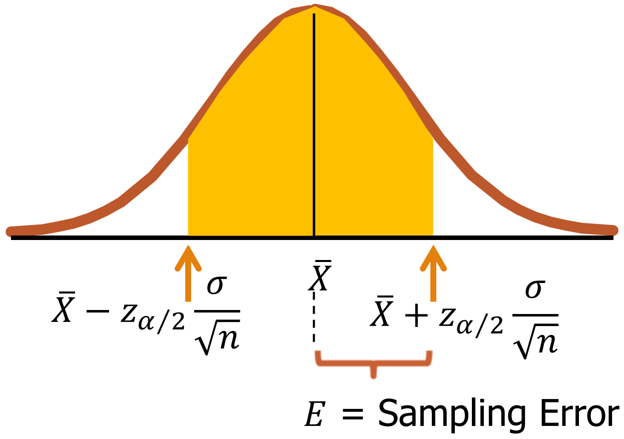
Determining Sample Size

- What sample size is needed to be $100(1-\alpha)\%$ confident of being correct to within $\pm E$?
- Assume σ is known

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

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



- If the population standard deviation is unknown,
  - Use prior information such as the sample standard deviation in earlier similar studies to guess σ value
  - If no prior information is available, estimate the range of the data and then estimate the standard deviation as range/4 (or range/6)
    - If population is normal or near-normal, 95.4% of the observations are within 2σ of the mean (99.7% of the observations are within 3σ of the mean)
  - Conduct a small-scale study and estimate the standard deviation from the resulting data

Example 5: Determining Sample Size (σ Known)

- A poll was conducted to study the degree of support of residents toward a policy using a scale of 0 to 100. A random sample of 1038 respondents gave a mean score of 20.23 and standard deviation of 28.84

1. In order to be 90% confident of being correctly reflecting the population opinion to within ± 2.5 points, what sample size is needed?

Use s to replace σ when σ is unknown

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{1.645 \times 28.84}{2.5}\right)^2 = 360.12 \cong 361$$

Round Up

Confidence level	Confidence coefficient 1-α	$Z_{\alpha/2}$ value
90%	0.90	1.645
95%	0.95	1.96
99%	0.99	2.575

Example 5: Determining Sample Size (σ Known)

2. If we want to increase our confidence to 95%, how many individuals should we interview? Keep all other factors remain unchanged.

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{1.96 \times 28.84}{2.5}\right)^2 = 511.24 \cong 512$$

Confidence level	Confidence coefficient 1-α	$Z_{\alpha/2}$ value
90%	0.90	1.645
95%	0.95	1.96
99%	0.99	2.575

Example 6: Determining Sample Size (σ Unknown)

- Suppose you want to estimate the mean GPA (μ) of all the students at your university at a margin of error of 0.3 and 95% confidence. How many students should be sampled?
  - For 95% confidence level, $\alpha = 0.05$, then $z_{\alpha/2} = 1.96$
  - Suppose the range of GPA is 4.3, we may estimate the standard deviation as 4.3/4
  - The required sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{1.96 \times \left(\frac{4.3}{4}\right)}{0.3}\right)^2 = 49.33 \cong 50$$

Confidence level	Confidence coefficient 1-α	$Z_{\alpha/2}$ value
90%	0.90	1.645
95%	0.95	1.96
99%	0.99	2.575

Determining Sample Size

- Warning:
  - The value of n does not depend on the size of the population. This is true as long as the population is much larger than the sample
  - The derived sample size should only be taken as a rough indicator for the desired margin of error
  - The true required sample size, which is unknown to us in practice, might be larger or smaller than the computed value