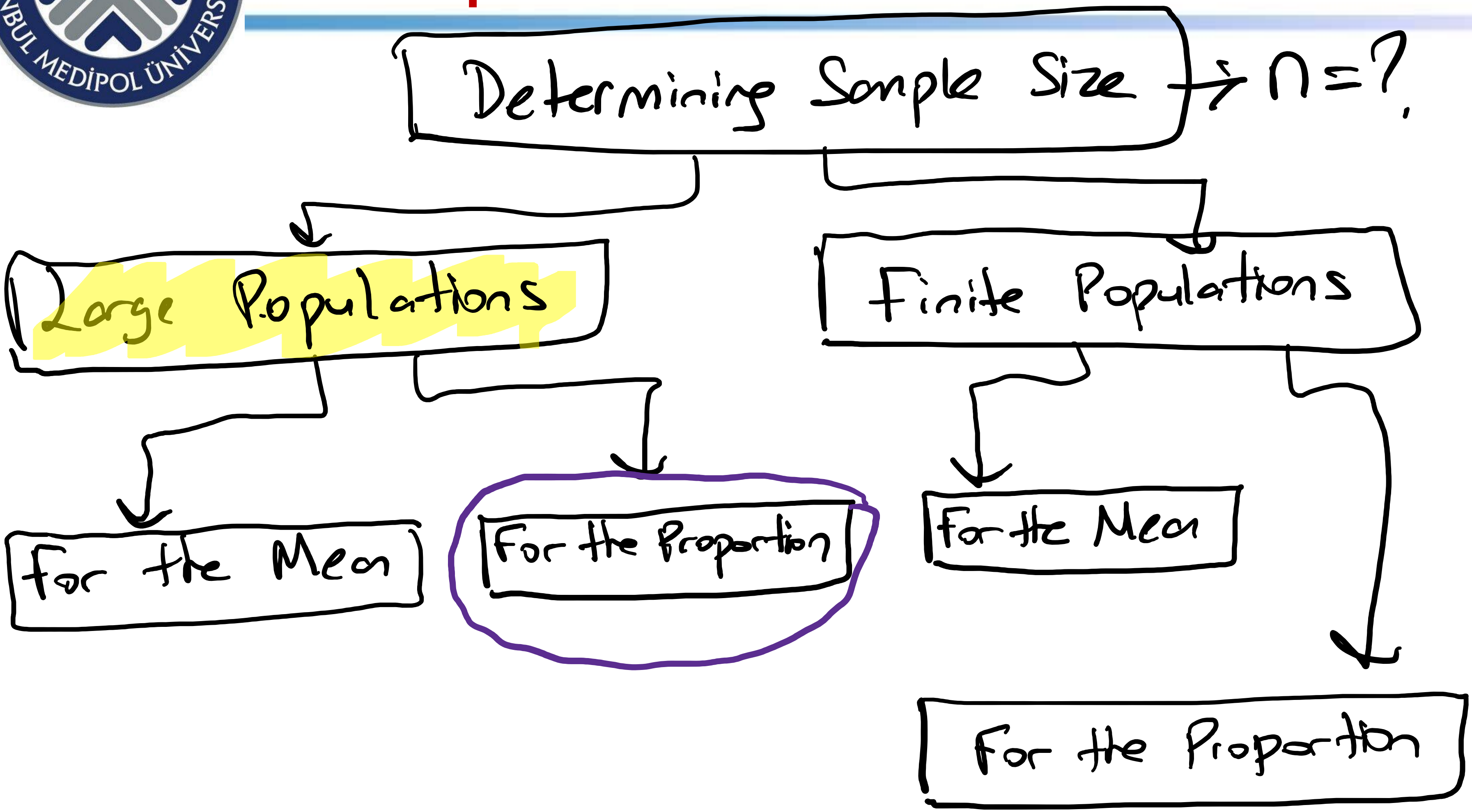
APPLIED STATISTICS

WEEK 4

OUTLINE

- Sample Size Determination

Sample Size Determination



Margin of Error (ME)

* The required somple size ca be fond to reach a desired margin of error (ME) with a specified level of confidence $(1-\alpha)$

* The margin of error is also called "sampling error".
  - the amount of imprecision in the estimate of population parameter.
  - the amount added and subtracted to the point estimate to form confidence interval.

$$\bar{X} \mp ME$$

to construct our confidence interval

LCL , UCL

Margin of Error

$$\bar{x} \pm ME \qquad x \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$ME = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Sampling Error

$$n = ?$$

Sample Size Determination

Sample Size Determination

Determining Sample Size for Large Populations

- To determine the required sample size (n) for the mean, you must know:

① The desired level of confidence $(1-\alpha)$, which determines $z_{\alpha/2}$ value (critical value)

② The acceptable margin of error (ME)

③ The population standard deviation, σ

Sample Size Determination

$$ME = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{ME}$$

$$n = \frac{(z_{\alpha/2})^2 \cdot \sigma^2}{ME^2}$$

"Isolate n"

"sample size"

Sample Size Determination

A note about σ:

* Solving for the sample size requires the population standard deviation, σ. Most often we do not know it so we have to use an estimate or "planning value" in its place.

There a few options:

1. Estimate σ from previous studies using the same population of interest
2. Conduct a pilot study to select a preliminary sample. Use the sample standard deviation from the pilot study.

Sample Size Determination

3 Use a "judgement" or "best guess" for σ. A common guess is the data range.

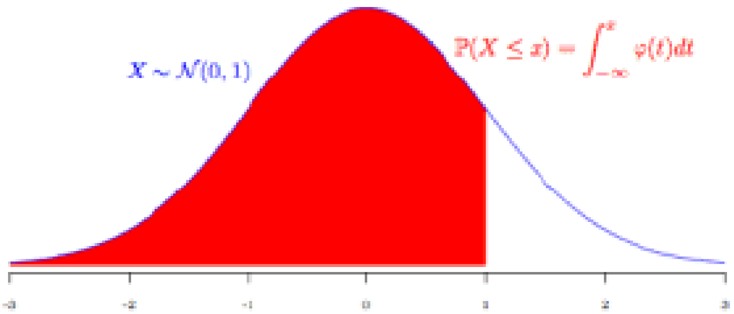
$$\rightarrow \frac{(\text{high} - \text{low})}{4} = \sigma$$

Common Levels of Confidence

- Commonly used confidence levels are 90%, 95%, and 99%

z-scores
"critical" values

Confidence Level	*Confidence Coefficient,* $1-\alpha$	$Z_{\alpha/2}$ *value*
80%	**.80**	**1.28**
90%	**.90**	**1.645**
95%	**.95**	**1.96**
98%	**.98**	**2.33**
99%	**.99**	**2.58**
99.8%	**.998**	**3.08**
99.9%	**.999**	**3.27**



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
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
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Standard Normal Tables

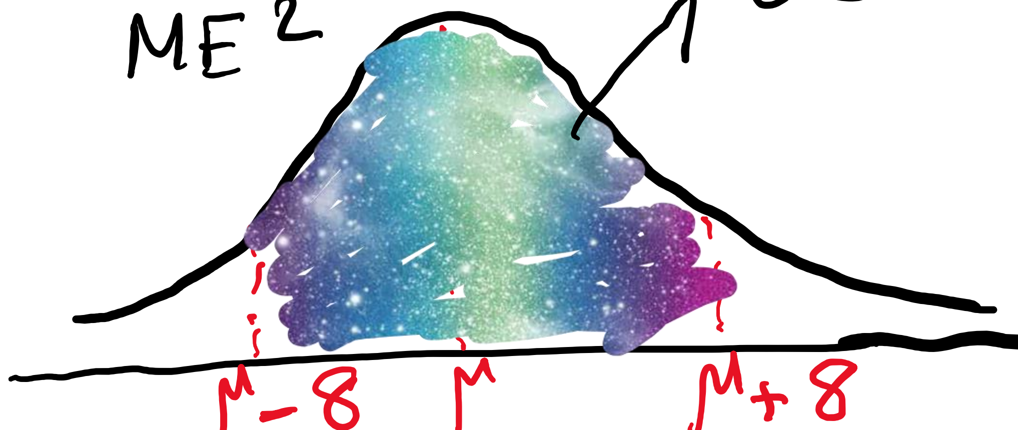
Required Sample Size Example

Question 1:

How large a sample should be selected to provide a 95% confidence interval with a margin of error (ME) of 8? Assuming that the population standard deviation, $\sigma = 36$.

a. 50
b. 57
c. 77
d. $\boxed{78}$ *

$$n = \frac{Z_{\alpha/2}^{2} \cdot \sigma^{2}}{ME^{2}}$$



Required Sample Size Example

$$n = \frac{z_{\alpha/2}^2 \cdot \sigma^2}{ME^2} = \frac{(1.96)^2 \cdot (36)^2}{8^2}$$

$(1-\alpha) = 0.95$

$\alpha = 0.05$

$\boxed{\alpha/2 = 0.025}$

$= 77.8$

(round up to next integer value)

$\cong \boxed{78}$

Interpretation:

– To have 95% of our "sample mean contain μ, we need a sample size of 78.

Question 2:

Required Sample Size Example

(2) A quick look at production data over the past month revealed that a global auto company produced a low of 782 and a high of 1303 (minimum and maximum) per hour. A 95% confidence interval estimate for vehicle production is desired.

a. What's the planing value for the population standard deviation, σ?

(b) How large a sample of production plant hours should be taken with the following MEs

→ ± 100 cars → ± 50 cars → ± 30 cars

Required Sample Size Example

a) Planning Value $= \frac{\text{Range}}{4} = \frac{1303 - 782}{4}$

Planning Value for σ $= \frac{521}{4} = \boxed{130.25} \approx 130$

b) $n = \frac{(z_{\alpha/2})^2 \cdot \sigma^2}{ME^2}$

$n_1 = \frac{(1.96)^2 \cdot (130)^2}{100^2}$, $n_2 = \frac{(1.96)^2 (130)^2}{50^2} = \boxed{26}$

$\boxed{n_1 = 7}$

$n_3 = \frac{(1.96)^2 (130)^2}{30^2} = \boxed{73}$

Question 1 :

Required Sample Size Example

(3) A survey by the US Chamber of Commerce asked upper-level managers about their retirement investment practices. Based on a previous study, the chamber assumed a planning value for the standard deviation of $1500 and they wish to use $200 as the desired margin of error for the interval estimate of the population mean. Determine the sample size for the following

sampling error amount

1. A 90 % Confidence Level → n = ?
2. A 95 % Confidence Level → n = ?
3. A 99 % Confidence Level → n = ?

Required Sample Size Example

$$n_1 = \frac{(1.645)^2 \cdot 1500^2}{200^2} \Rightarrow 153 \text{ managers to interview}$$

90%

$$n_2 = \frac{(1.96)^2 \cdot 1500^2}{200^2} = 217 \text{ managers to interview}$$

95%

$$n_3 = \frac{(2,58)^2 \cdot 1500^2}{200^2} = 374 \text{ managers to interview}$$

99%

Sample size increases

When the level of confidence is increased a larger sample size is required

Sample Size Determination: Population Proportion

* The sample and population proportions, $\hat{p}$ and p are generally not known (since no sample has taken yet!)

* $\boxed{p(1-p) = 0.25}$ generates the largest possible margin of error (so guarantees that the resulting sample size will meet the desired level of confidence)

Sample Size Determination: Population Proportion

$\hat{p} \pm ME$ → While conducting confidence interval for P

Margin of Error

$$\boxed{ME = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

$$n = \frac{\hat{p}(1-\hat{p}) \cdot z^2_{\alpha/2}}{ME^2}$$

0.25

$\hat{p}(1-\hat{p})$ can not be larger than 0.25 when $\hat{p} = 0.5$ substitute 0.25 for $\hat{p}(1-\hat{p})$ and solve for n.

Sample Size Determination: Population Proportion

$\hat{p} * (1-\hat{p})$

$0.1\,(0.9) = 0.09$

$0.2\,(0.8) = 0.16$

$0.3\,(0.7) = 0.21$

$0.4\,(0.6) = 0.24$

$0.5\,(0.5) = \boxed{0.25}$

$$n = \frac{\hat{p}(1-\hat{p}) \cdot Z^2_{\alpha/2}}{ME^2}$$

$$n = \frac{\boxed{0.25} \cdot Z^2_{\alpha/2}}{ME^2}$$

$$\boxed{n = \frac{0.25 \cdot Z^2_{\alpha/2}}{ME^2}}$$

Sample Size Determination: Population Proportion

(1) How large a sample would be necessary to estimate the true proportion defective in a large population within $\pm$ 3% with 95% confidence?

For 95% confidence $\rightarrow Z_{\alpha/2} = Z_{0.025} = 1.96$

ME = Margin of Error = Sampling Error = 0.03

Estimate $\Rightarrow \hat{p}(1-\hat{p}) = 0.25$

$$n = \frac{0.25 \cdot Z_{\alpha/2}^2}{ME^2} = \frac{0.25(1.96)^2}{(0.03)^2} = 1067.11 \uparrow$$

$$\boxed{n = 1068}$$

Required Sample Size Example: Population Proportion

2) If a retailer would like to estimate the proportion of their customers who bought an item after viewing their website on a certain day with a 95% confidence level and 5% margin of error. How many customers do they have to monitor?

$$n = \frac{0.25 \; z_{\alpha/2}^2}{ME^2} = ?$$

$$n = \frac{0.25 \, (1.96)^2}{(0.05)^2} = 384.16$$

always round up to next integer

$\boxed{385}$

Sample Size Tables

- Several tables are available to calculated the required sample size!.

Required Sample Size

Population Size	Confidence = 95%				Confidence = 99%			
	Margin of error				Margin of Error			
	5.0%	3.5%	2.5%	1.0%	5.0%	3.5%	2.5%	1.0%
10	10	10	10	10	10	10	10	10
20	19	20	20	20	19	20	20	20
30	28	29	29	30	29	29	30	30
50	44	47	48	50	47	48	49	50
75	63	69	72	74	67	71	73	75
100	80	89	94	99	87	93	96	99
150	108	126	137	148	122	135	142	149
200	132	160	177	196	154	174	186	198
250	152	190	215	244	182	211	229	246
300	169	217	251	291	207	246	270	295
400	146	265	318	384	250	309	348	391
500	217	306	377	475	285	365	421	485
600	234	340	432	565	315	416	490	579
700	248	370	481	653	341	462	554	672
800	260	396	526	739	363	503	615	763
1,000	278	440	606	906	399	575	727	943
1,200	291	474	674	1,067	427	636	827	1,119
1,500	306	515	759	1,297	460	712	959	1,376
2,000	322	563	869	1,655	498	808	1,141	1,785
2,500	333	597	952	1,984	524	879	1,288	2,173
3,500	346	641	1,068	2,565	558	977	1,510	2,890
5,000	357	678	1,176	3,288	586	1,066	1,734	3,842
7,500	365	710	1,275	4,211	610	1,147	1,960	5,165
10,000	370	727	1,332	4,899	622	1,193	2,098	6,239
25,000	378	760	1,448	6,939	646	1,285	2,399	9,972
50,000	381	772	1,491	8,056	655	1,318	2,520	12,455
75,000	382	776	1,506	8,514	658	1,330	2,563	13,583
100,000	383	778	1,513	8,762	659	1,336	2,585	14,227
250,000	384	782	1,527	9,248	662	1,347	2,626	15,555
500,000	384	783	1,532	9,423	663	1,350	2,640	16,055
1,000,000	384	783	1,534	9,512	663	1,352	2,647	16,317
2,500,000	384	783	1,536	9,567	663	1,353	2,651	16,478
10,000,000	384	784	1,536	9,594	663	1,354	2,653	16,560
100,000,000	384	784	1,537	9,603	663	1,354	2,654	16,584
300,000,000	384	784	1,537	9,603	663	1,354	2,654	16,586

Designing Survey or Developing Questionnaire

$n = 217$

$n = ?$

Sample size

Population is Finite?