8.1 Population Mean when σ is known

100

- Recall that the sample mean $\bar{x}$ is the point estimator for the population mean μ.
- $\bar{x}$ can not be expected to provide the exact estimate of μ

$\Rightarrow$ An interval Estimat of a population Mean when σ is known is used as

Point estimate $\pm$ Margin of error

$$\bar{x} \pm z_{\alpha/2}\, \sigma_{\bar{x}} = \left[\bar{x} - z_{\alpha/2}\sigma_{\bar{x}} ,\ \bar{x} + z_{\alpha/2}\, \sigma_{\bar{x}}\right]$$

confidence interval

where
- $(1-\alpha)$ is the confidence coefficient
- $z_{\alpha/2}$ is the z value providing an area of $\frac{\alpha}{2}$ in the upper tail of the standard normal prob. distribution
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ is the standard error.

- Values of $z_{\alpha/2}$ for common used confidence levels:

Confidence Level	α	$\alpha/2$	$z_{\alpha/2}$
90 %	0.10	0.05	1.645
95 %	0.05	0.025	1.96
99 %	0.01	0.005	$\boxed{2.576}$ → 2,575

Q1 page 299

Example: A simple random sample of 40 items resulted in a sample mean of 25. The population standard deviation is $\sigma = 5$

$n = 40$, $\bar{x} = 25$
$\sigma = 5$

[a] what is the standard error of the mean?

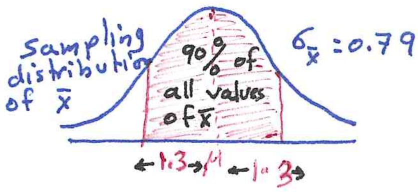
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{40}} = 0.79$$

[b] At 95 % confidence, what is the margin of error?

The margin of error $= z_{\alpha/2}\, \sigma_{\bar{x}} = (1.96)(0.79) = 1.55$

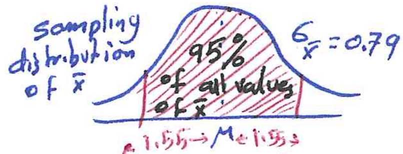
[c] Provide a 90 % confidence interval for the population mean?

$$\bar{x} \pm z_{\alpha/2}\, \sigma_{\bar{x}} = 25 \pm (1.645)(0.79) = 25 \pm 1.3 = [23.7, 26.3]$$



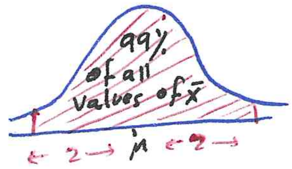
d) Provide a 95% confidence interval for the population mean?

$$\bar{X} \pm Z_{\alpha/2}\, \sigma_{\bar{X}} = 25 \pm (1.96)(0.79) = 25 \pm 1.55 = [23.45, 26.55]$$



e) Provide a 99% confidence interval for the population mean?

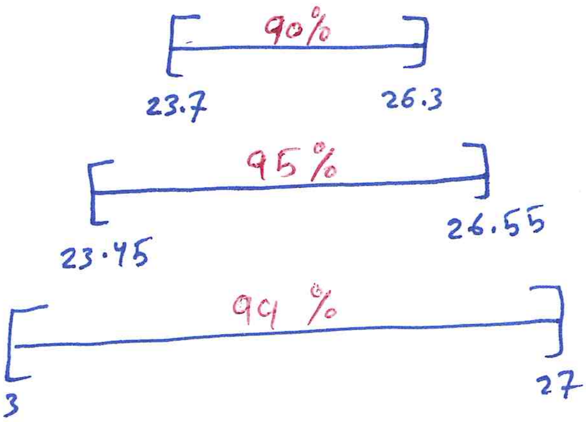
$$\bar{x} \pm Z_{\alpha/2}\, \sigma_{\bar{x}} = 25 \pm (2.576)(0.79) = 25 \pm 2 = [23, 27]$$



f) what is the relation between the confidence level and the confidence interval?

positive relation shipe.

- As the confidence level increases, the confidence interval increases,
- As the confidence level decreases, the confidence level decreases.



g) what is the relationship between the confidence level and the margin of error?

Confidence level	Marginal of error
90%	1.3
95%	1.55
99%	2

✓ Positive relationship

- As the level of confidence ↑ the margin of error ↑
- As " " " " ↓ " " " " ↓

h) what is the relationship between the sample size and the confidence interval?

As $n\uparrow \Rightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \downarrow \Rightarrow$ the margin of error $= Z_{\alpha/2}\, \sigma_{\bar{x}} \downarrow$

$\Rightarrow$ the confidence interval ↓

As $n\downarrow \Rightarrow$ the confidence interval ↑

- Interval estimate (or Confidence interval): is an estimate of a population parameter that provides an interval believed to contain the value of the parameter. It has the form: point estimate $\pm$ margin of error.

- Margin of error: The $\pm$ value added and subtracted from a point estimate in order to develop an interval estimate of a population parameter.

- σ known: The case when historical data or other information provides a good value for the population standard deviation σ prior taking a sample. The interval estimation uses this known value of σ in computing the margin of error.

- Confidence level: The confidence associated with an interval estimate. 90%, 95%, 99%.

Confidence coefficient: The confidence level expressed as decimal value 0.90, 0.95, 0.99

8.2 Population Mean when σ is unknown

(σ unknown): The more common case when no good basis exists for estimating the population standard deviation prior taking the sample.

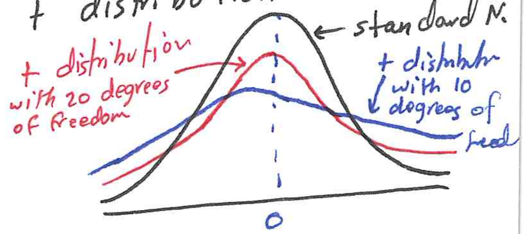
$\Rightarrow$ So we use $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ to estimat σ in order to compute the margin of error.

In this case, the margin of error and the interval estimate for the population mean are based on a prob. distribution called t distribution.

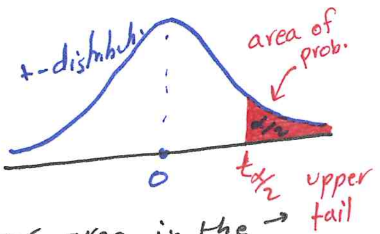
t distribution: A family of prob. distributions that can be used to develop an interval estimate of a population mean whenever the population standard deviation σ is unknown and is estimated by the sample standard deviation s.

Degrees of freedom: A parameter of the t distribution. when the t distribution is used in the computation of an interval estimate of a population mean, the appropriate t distribution has n-1 degrees of freedom, where n is the sample size.

- As the number of degrees of freedom increases, t distribution becomes closely to the standard normal distribution.



- The mean of the t-distribution is 0
- We use the notation $t_{\alpha/2}$ to represent a t value with an area of $\frac{\alpha}{2}$ in the upper tail of the t distribution (see table 303)



- For example when $\alpha = 0.05$, then $t_{0.025}$ indicates 0.025 area in the → upper tail
- A t distribution with 5 degrees of freedom is $t_{0.025} = 2.571$
- A t distribution with 100 " " " " $t_{0.025} = 1.984 \approx 1.96 = z_{0.025}$

If the number of degrees of freedom > 100, then the standard normal z value provides a good approximation to the t value.

An interval Estimate of a population Mean when (σ is unknown) is

$$* \cdots \quad \bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = \left[\bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}},\ \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$$

where
- $(1-\alpha)$ is the confidence coefficient
- $t_{\alpha/2}$ is the t value providing an area of $\frac{\alpha}{2}$ in the upper tail of the t distribution with $n-1$ degrees of freedom.
- $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ is the sample standard deviation.

  since $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$, it is enough to know $n-1$ independent terms = degree of freedom

---

Example (Q13 page 308) The following sample data are from a normal population

10, 8, 12, 15, 13, 11, 6, 5

[a] What is the point estimate of the population mean?

$$\bar{x} = \frac{\sum x_i}{n} = \frac{80}{8} = 10$$

[b] what is the point estimate of the population standard deviation?

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{84}{7}} = 3.464$$

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
10	0	0
8	-2	4
12	2	4
15	5	25
13	3	9
11	1	1
6	-4	16
5	-5	25
80		84

[c] with 95% confidence, what is the margin of error for the estimation of the population mean?

The margin of error $= t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

degrees of freedom $= n-1 = 7$

$$= t_{0.025} \frac{3.464}{\sqrt{8}} = \underline{2.365}\,(1.225) = 2.9$$

[d] what is the 95% confidence interval for the population mean?

$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 10 \pm 2.9 = [10-2.9,\ 10+2.9] = [7.1,\ 12.9]$$

---

[e] What is the 90% confidence interval interval for the population mean?

$$\alpha = 10\% \Rightarrow \frac{\alpha}{2} = 5\% \Rightarrow \bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 10 \pm t_{0.05}\,(1.225) = 10 \pm 1.86\,(1.225)$$

$$= 10 \pm 2.29 = [7.71,\ 12.29]$$

* If the population follows a normal distribution, then the confidence interval $\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ is exact and can be used for any n.

* If " " does not " " " " " " approximate. Thus, the quality of " " " " the approximation depens on the distribution of the population and on n.

* Larger sample sizes n are needed if the distribution of the population is highly skewed or includes outliers.

* If $n \geq 30$, then we can use $\bar{X} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ to estimate the population mean.

* If the population is highly skewed or contains outliers, then $n \geq 50$ need to use $\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$.

* If the population is not normally distributed but is roughly symmetric, then small sample sizes "$n \geq 15$" can be used to provid good approximate confidence interval $\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

* If $n < 15$, then $\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ can be used only if we assume that the population is at least approximately normal.

Summary:

Two approaches to develop an interval estimate of μ

σ is known

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

The margin of error $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ is fixed for all n

σ is unknown

$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

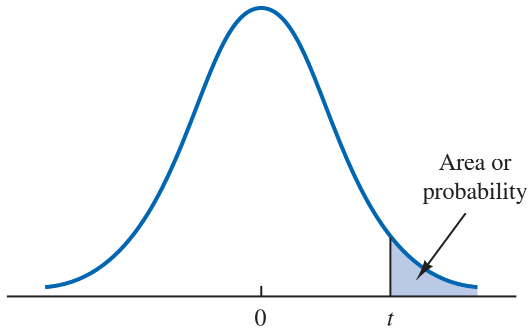
where $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

The margin of error $t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ varies from sample to sample.

as $s \uparrow \Rightarrow$ margin of error $\uparrow$

$s \downarrow \Rightarrow$ " " " $\downarrow$

TABLE 2 *t* DISTRIBUTION



Entries in the table give t values for an area or probability in the upper tail of the t distribution. For example, with 10 degrees of freedom and a .05 area in the upper tail, $t_{.05} = 1.812$.

Degrees of Freedom	Area in Upper Tail					
	.20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
10	.879	1.372	1.812	2.228	2.764	3.169
11	.876	1.363	1.796	2.201	2.718	3.106
12	.873	1.356	1.782	2.179	2.681	3.055
13	.870	1.350	1.771	2.160	2.650	3.012
14	.868	1.345	1.761	2.145	2.624	2.977
15	.866	1.341	1.753	2.131	2.602	2.947
16	.865	1.337	1.746	2.120	2.583	2.921
17	.863	1.333	1.740	2.110	2.567	2.898
18	.862	1.330	1.734	2.101	2.552	2.878
19	.861	1.328	1.729	2.093	2.539	2.861
20	.860	1.325	1.725	2.086	2.528	2.845
21	.859	1.323	1.721	2.080	2.518	2.831
22	.858	1.321	1.717	2.074	2.508	2.819
23	.858	1.319	1.714	2.069	2.500	2.807
24	.857	1.318	1.711	2.064	2.492	2.797
25	.856	1.316	1.708	2.060	2.485	2.787
26	.856	1.315	1.706	2.056	2.479	2.779
27	.855	1.314	1.703	2.052	2.473	2.771
28	.855	1.313	1.701	2.048	2.467	2.763
29	.854	1.311	1.699	2.045	2.462	2.756
30	.854	1.310	1.697	2.042	2.457	2.750
31	.853	1.309	1.696	2.040	2.453	2.744
32	.853	1.309	1.694	2.037	2.449	2.738
33	.853	1.308	1.692	2.035	2.445	2.733
34	.852	1.307	1.691	2.032	2.441	2.728

TABLE 2 *t* DISTRIBUTION (*Continued*)

Degrees of Freedom	Area in Upper Tail .20	.10	.05	.025	.01	.005
35	.852	1.306	1.690	2.030	2.438	2.724
36	.852	1.306	1.688	2.028	2.434	2.719
37	.851	1.305	1.687	2.026	2.431	2.715
38	.851	1.304	1.686	2.024	2.429	2.712
39	.851	1.304	1.685	2.023	2.426	2.708
40	.851	1.303	1.684	2.021	2.423	2.704
41	.850	1.303	1.683	2.020	2.421	2.701
42	.850	1.302	1.682	2.018	2.418	2.698
43	.850	1.302	1.681	2.017	2.416	2.695
44	.850	1.301	1.680	2.015	2.414	2.692
45	.850	1.301	1.679	2.014	2.412	2.690
46	.850	1.300	1.679	2.013	2.410	2.687
47	.849	1.300	1.678	2.012	2.408	2.685
48	.849	1.299	1.677	2.011	2.407	2.682
49	.849	1.299	1.677	2.010	2.405	2.680
50	.849	1.299	1.676	2.009	2.403	2.678
51	.849	1.298	1.675	2.008	2.402	2.676
52	.849	1.298	1.675	2.007	2.400	2.674
53	.848	1.298	1.674	2.006	2.399	2.672
54	.848	1.297	1.674	2.005	2.397	2.670
55	.848	1.297	1.673	2.004	2.396	2.668
56	.848	1.297	1.673	2.003	2.395	2.667
57	.848	1.297	1.672	2.002	2.394	2.665
58	.848	1.296	1.672	2.002	2.392	2.663
59	.848	1.296	1.671	2.001	2.391	2.662
60	.848	1.296	1.671	2.000	2.390	2.660
61	.848	1.296	1.670	2.000	2.389	2.659
62	.847	1.295	1.670	1.999	2.388	2.657
63	.847	1.295	1.669	1.998	2.387	2.656
64	.847	1.295	1.669	1.998	2.386	2.655
65	.847	1.295	1.669	1.997	2.385	2.654
66	.847	1.295	1.668	1.997	2.384	2.652
67	.847	1.294	1.668	1.996	2.383	2.651
68	.847	1.294	1.668	1.995	2.382	2.650
69	.847	1.294	1.667	1.995	2.382	2.649
70	.847	1.294	1.667	1.994	2.381	2.648
71	.847	1.294	1.667	1.994	2.380	2.647
72	.847	1.293	1.666	1.993	2.379	2.646
73	.847	1.293	1.666	1.993	2.379	2.645
74	.847	1.293	1.666	1.993	2.378	2.644
75	.846	1.293	1.665	1.992	2.377	2.643
76	.846	1.293	1.665	1.992	2.376	2.642
77	.846	1.293	1.665	1.991	2.376	2.641
78	.846	1.292	1.665	1.991	2.375	2.640
79	.846	1.292	1.664	1.990	2.374	2.639

TABLE 2 *t* DISTRIBUTION (*Continued*)

Degrees of Freedom	Area in Upper Tail					
	.20	**.10**	**.05**	**.025**	**.01**	**.005**
80	.846	1.292	1.664	1.990	2.374	2.639
81	.846	1.292	1.664	1.990	2.373	2.638
82	.846	1.292	1.664	1.989	2.373	2.637
83	.846	1.292	1.663	1.989	2.372	2.636
84	.846	1.292	1.663	1.989	2.372	2.636
85	.846	1.292	1.663	1.988	2.371	2.635
86	.846	1.291	1.663	1.988	2.370	2.634
87	.846	1.291	1.663	1.988	2.370	2.634
88	.846	1.291	1.662	1.987	2.369	2.633
89	.846	1.291	1.662	1.987	2.369	2.632
90	.846	1.291	1.662	1.987	2.368	2.632
91	.846	1.291	1.662	1.986	2.368	2.631
92	.846	1.291	1.662	1.986	2.368	2.630
93	.846	1.291	1.661	1.986	2.367	2.630
94	.845	1.291	1.661	1.986	2.367	2.629
95	.845	1.291	1.661	1.985	2.366	2.629
96	.845	1.290	1.661	1.985	2.366	2.628
97	.845	1.290	1.661	1.985	2.365	2.627
98	.845	1.290	1.661	1.984	2.365	2.627
99	.845	1.290	1.660	1.984	2.364	2.626
100	.845	1.290	1.660	1.984	2.364	2.626
∞	.842	1.282	1.645	1.960	2.326	2.576

8.3 Determining the sample size

How to choose a sample size $\underline{n}$ to provide a desired margin of error?

- Recall that the margin of Error $= Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$$E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \iff {}^{*} n = \left(\frac{Z_{\alpha/2}\, \sigma}{E} \right)^2 ,$$

n is rounded up because n will be the min sample size satisfies the E.

- $Z_{\alpha/2}$ is known as far as we choose the confidence level.
  " If we choose 95% confidence level, then $Z_{\alpha/2} = Z_{0.025} = 1.96$ "

- If σ is known, then we use * directely.
- If σ is unknown, then we use * also by estimating the planing value for σ as follows:

1) Use s as the planing value for σ, where $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$, when the data are available in sample.

2) Select a sample and estimate s as the planing value for σ, when there is no sample.

3) Use judgment or "best guess" for σ. For example, if we know the largest value and the smallest value in the population, then the planning value for $\sigma = \frac{\text{Range}}{4} = \frac{\text{Max Value} - \text{Min Value}}{4}$.

Example (Q 23 page 312) How large a sample should be selected to provide 95% confidence level interval with margin of error of 10?

1 Assume that the population standard deviation is 40?

$Z_{\alpha/2} = Z_{0.025} = 1.96$ and $\sigma = 40$ and $E = 10$

$$n = \left(\frac{Z_{\alpha/2}\, \sigma}{E} \right)^2 = \left(\frac{1.96\,(40)}{10} \right)^2 = 61.46 \approx 62$$

2 Assume that the range of data is estimated to be 124?

The planning value of $\sigma = \frac{\text{Range}}{4} = \frac{124}{4} = 31$

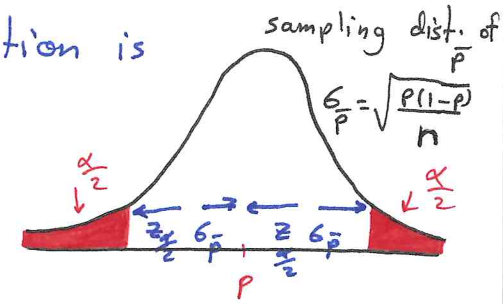
$$n = \left(\frac{Z_{\alpha/2}\, \sigma}{E} \right)^2 = \left(\frac{1.96 \times 31}{10} \right)^2 = 36.9 \approx 37$$

8.4 Population Proportion

- Recall that: the sample proportion $\bar{p}$ is the point estimator for the population proportion p.
- $\bar{p}$ can not be expected to provide the exact estimate of p
- An interval estimate of a population proportion is

Point estimate ± Margin of error

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$



where • $(1-\alpha)$ is the confidence coefficient

• $z_{\alpha/2}$ is the z value providing an area of $\frac{\alpha}{2}$ in the upper tail of the standard normal distribution.

• Recall that the sampling distribution $\bar{p}$ can be approximated by a normal distribution if $np \geq 5$ and $n(1-p) \geq 5$.

Example (Φ31 page 316) A simple random sample of 400 individuals provides 100 Yes responses.

[a] what is the point estimate of the proportion of the population that would provide Yes responses?

$$\bar{p} = \frac{100}{400} = 0.25$$

[b] What is your estimate of the standard error of the proportion?

$$\sigma_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{0.25 \times 0.75}{400}} = 0.0217$$

[c] Compute the 95% confidence interval for the population proportion?

$$\bar{p} \pm z_{0.025}\,\sigma_{\bar{p}} = 0.25 \pm (1.96)(0.0217) = 0.25 \pm 0.0424$$
$$= [0.2076, 0.2924]$$

How Large the sample size should be to obtain an estimate of a population proportion for a desired margin of error?

• Recall that the margin of error $= z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

$$E = z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \iff \boxed{n = \frac{z_{\alpha/2}^2\, p^*(1-p^*)}{E^2}}$$, where p^* is

the planning value for $\bar{p}$ that can be determine in the following ways:

1) use $p^* = \bar{p}$ the sample proportion as the planning value for p when data are available in sample.

2) Select a sample and estimate $\bar{p} = p^*$ as the planning value for p when there is no sample.

3) Use judgment or "best guess" for p^*

4) If 1), 2), 3) don't apply or if data are not available, then use planning value $p^* = 0.5$

Example (Q33 page 317) In a survey, the planning value for the population is $p^* = 0.35$. How large a sample should be taken to provide a 95% confidence interval with a margin of error 0.05?

$$n = \frac{z_{\alpha/2}^2 \, p^*(1-p^*)}{E^2} = \frac{(1.96)^2(0.35)(0.65)}{(0.05)^2} = 349.6 \approx 350$$

Example How large a sample should be taken to provid a 99% confidence interval with a margin of error 0.03 if the past data are not available for developing a planning value for p^*?

$z_{\alpha/2} = z_{0.005} = 2.576$

$p^* = 0.5$

$$n = \frac{z_{\alpha/2}^2 \, p^*(1-p^*)}{E^2} = \frac{(2.576)^2(0.5)(0.5)}{(0.03)^2} = 1067.1 \approx 1068$$

The desired margin of error for estimating a population proportion is almost always ≤ 0.1

$\Rightarrow$ This provides a sample that is large enough to satisfy $np \geq 5$ and $n(1-p) \geq 5$ so that

---

we use the normal distribution to approximate the sampling distribution of $\bar{p}$.