STAT 400 — Stepanov
UIUC — Dalpiaz

Answers for 7.3, 7.4

The **sample proportion**:

$$\hat{p} = \frac{x}{n}$$

where x is the number of elements in the sample found to belong to the category of interest (the number of "successes"), and n is the sample size.

$\mathrm{E}(\hat{P}) = p$, $\qquad \mathrm{Var}(\hat{P}) = \dfrac{p \cdot (1 - p)}{n}$, $\qquad \mathrm{SD}(\hat{P}) = \sqrt{\dfrac{p \cdot (1 - p)}{n}}$.

A large-sample confidence interval for the population proportion p is

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

0. Let X have a Binomial distribution with parameters n and p. Recall that

$$\frac{\mathrm{X} - n\,p}{\sqrt{n\,p\,(1 - p)}}$$

has an approximate Standard Normal $\mathrm{N}(0, 1)$ distribution, provided that n is large enough, and

$$\mathrm{P}\left[-z_{\alpha/2} < \frac{\mathrm{X} - n\,p}{\sqrt{n\,p\,(1 - p)}} < z_{\alpha/2} \right] \approx 1 - \alpha .$$

Show that an approximate $100(1 - \alpha)\%$ confidence interval for p is

$$\frac{\hat{p} + \dfrac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1 - \hat{p})}{n} + \dfrac{z_{\alpha/2}^2}{4n^2}}}{1 + \dfrac{z_{\alpha/2}^2}{n}}, \qquad \text{where } \hat{p} = \frac{\mathrm{X}}{n}.$$

This interval is called the Wilson interval. Note that for large n, this interval is approximately equal to

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

$$\left| \frac{X - np}{\sqrt{np(1-p)}} \right| < z_{\alpha/2} \qquad \Leftrightarrow \qquad \frac{(X - np)^2}{np(1-p)} < z_{\alpha/2}^2$$

$$\Leftrightarrow \qquad X^2 - 2nXp + n^2 p^2 < n z_{\alpha/2}^2 p - n z_{\alpha/2}^2 p^2$$

$$\Leftrightarrow \qquad \hat{p}^2 - 2 \hat{p} p + p^2 < \frac{z_{\alpha/2}^2}{n} p - \frac{z_{\alpha/2}^2}{n} p^2$$

$$\Leftrightarrow \qquad \left(1 + \frac{z_{\alpha/2}^2}{n} \right) p^2 - \left(2\hat{p} + \frac{z_{\alpha/2}^2}{n} \right) p + \hat{p}^2 < 0$$

$$a p^2 + b p + c = 0 \qquad \Rightarrow \qquad p_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$a p^2 + b p + c < 0 \qquad \Leftrightarrow \qquad p_1 < p < p_2 \qquad (a > 0)$$

$$p_{1,2} = \frac{\left(2\hat{p} + \frac{z_{\alpha/2}^2}{n} \right) \pm \sqrt{\left(2\hat{p} + \frac{z_{\alpha/2}^2}{n} \right)^2 - 4 \left(1 + \frac{z_{\alpha/2}^2}{n} \right) \hat{p}^2}}{2 \left(1 + \frac{z_{\alpha/2}^2}{n} \right)}$$

$$= \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm \sqrt{\left(\hat{p} + \frac{z_{\alpha/2}^2}{2n} \right)^2 - \left(1 + \frac{z_{\alpha/2}^2}{n} \right) \hat{p}^2}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

$$= \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm \sqrt{\hat{p}^2 + 2\hat{p} \frac{z_{\alpha/2}^2}{2n} + \frac{z_{\alpha/2}^4}{4n^2} - \hat{p}^2 - \frac{z_{\alpha/2}^2}{n} \hat{p}^2}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

$$= \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm \sqrt{\hat{p}\frac{z_{\alpha/2}^2}{n} - \frac{z_{\alpha/2}^2}{n}\hat{p}^2 + \frac{z_{\alpha/2}^4}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

$$= \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

Wilson interval: $$\frac{X + \frac{z_{\alpha/2}^2}{2}}{n + z_{\alpha/2}^2} \pm z_{\alpha/2} \;\cdots$$

1. Just prior to an important election, in a random sample of 749 voters, 397 preferred Candidate Y over Candidate Z. Construct a 90% confidence interval for the overall proportion of voters who prefer Candidate Y over Candidate Z.

X = 397. n = 749. $\hat{p} = \frac{X}{n} = \frac{397}{749} = 0.53.$

The confidence interval : $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}$.

90% confidence level α = 0.10 $\alpha/2 = 0.05.$ $z_{\alpha/2} = 1.645.$

$0.53 \pm 1.645 \cdot \sqrt{\frac{(0.53)(0.47)}{749}}$ **0.53 ± 0.03** **(0.50 , 0.56)**

	Upper-tail probability											
	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
z	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level											

2. An article on secretaries' salaries in the *Wall Street Journal* reports: "Three-fourth of surveyed secretaries said they make less than $25,000 a year." Suppose that the *Journal* based its results on a random sample of 460 secretaries drawn from every category of business. Give a 95% confidence interval for the proportion of secretaries earning less than $25,000 a year.

$n = 460.$ $\hat{p} = 0.75.$

The confidence interval : $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p} \cdot (1-\hat{p})}{n}}.$

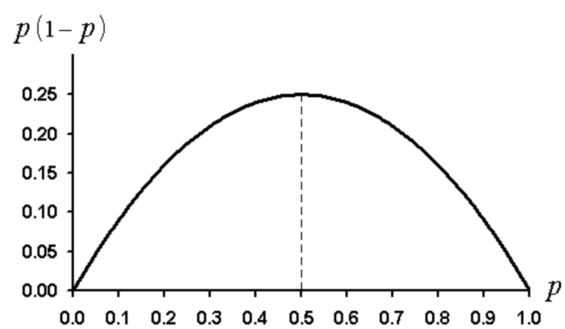
95% confidence level $\alpha = 0.05$ $\alpha/2 = 0.025.$ $z_{\alpha/2} = 1.960.$

$0.75 \pm 1.960 \cdot \sqrt{\dfrac{(0.75)(0.25)}{460}}$ $\mathbf{0.75 \pm 0.04}$ **(0.71 , 0.79)**

==

The sample size required to obtain a confidence interval for the population proportion p with specified margin of error ε is

$$n = \left(\frac{z_{\alpha/2}}{\varepsilon} \right)^2 p^* \left(1 - p^* \right).$$



Always round n up.

Conservative Approach: $p^* = 0.50.$

- If it is possible that $p = 0.50$, use $p^* = 0.50.$
- If it is not possible that $p = 0.50$, use p^* = the closest to 0.50 possible value of p.

3. Find the minimum sample size required for the overall proportion of voters who prefer Candidate Y over Candidate Z to within 2% with 90% confidence. (Assume that no guess as to what that proportion might be is available.)

Use $p^* = 0.50$. $\quad\varepsilon = 0.02$.

90% confidence level, $\quad\alpha = 0.10$, $\quad\alpha/2 = 0.05$, $\quad z_{\alpha/2} = z_{0.05} = 1.645$.

$$n = \left(\frac{z_{\alpha/2}}{\varepsilon}\right)^2 \cdot p^* \cdot (1 - p^*) = \left(\frac{1.645}{0.02}\right)^2 \cdot 0.50 \cdot 0.50 = 1691.266.$$

Round **up**. $\quad n =$ **1692**.

4. A television station wants to estimate the proportion of the viewing audience in its area that watch its evening news. Find the minimum sample size required to estimate that proportion to within 3% with 95% confidence if …

a) no guess as to the value of that proportion is available.

$\varepsilon = 0.03$.

95% confidence level, $\quad\alpha = 0.05$, $\quad\alpha/2 = 0.025$, $\quad z_{\alpha/2} = z_{0.025} = 1.960$.

Use $p^* = 0.50$.

$$n = \left(\frac{z_{\alpha/2}}{\varepsilon}\right)^2 \cdot p^* \cdot (1 - p^*) = \left(\frac{1.960}{0.03}\right)^2 \cdot 0.50 \cdot 0.50 = 1067.1111.$$

Round **up**. $\quad n =$ **1068**.

b) it is known that the station's evening news reaches at most 30% of the viewing audience.

Use $p^* = 0.30$.

$$n = \left(\frac{z_{\alpha/2}}{\varepsilon}\right)^2 \cdot p^* \cdot (1 - p^*) = \left(\frac{1.960}{0.03}\right)^2 \cdot 0.30 \cdot 0.70 = 896.3733.$$

Round **up**. $\quad n =$ **897**.

4½. If a fair 6-sided die is rolled "many" times, then the proportion of times when **6** comes up will "eventually" get very close to $\frac{1}{6}$. How many times should the die be rolled for the probability to be 0.9544 that the proportion of times when **6** comes up is between $\frac{4}{30}$ and $\frac{6}{30}$ (i.e., within $\frac{1}{30}$ of $\frac{1}{6}$)?

Use $p^* = \frac{1}{6}$. $\varepsilon = \frac{1}{30}$.

95.44% confidence level, $\alpha = 0.0456$, $\alpha/2 = 0.0228$, $z_{\alpha/2} = z_{0.0228} = 2.00$.

$$n = \left(\frac{z_{\alpha/2}}{\varepsilon}\right)^2 \cdot p^* \cdot (1 - p^*) = \left(\frac{2.00}{1/30}\right)^2 \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right) = \mathbf{500}.$$

At least 500 times.