$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\}$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n - np}{n}\right| > \epsilon\right\}$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n - np}{n}\right| > \epsilon\right\}$$

$$= \mathbf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n - np}{n}\right| > \epsilon\right\}$$

$$= \mathbf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\left\{\left|S_n^*\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$
\begin{aligned}
\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} &= \mathbf{P}\left\{\left|\frac{S_n - np}{n}\right| > \epsilon\right\} \\
&= \mathbf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\} \\
&= \mathbf{P}\left\{\left|S_n^*\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}
\end{aligned}
$$

$$
z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}
$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n - np}{n}\right| > \epsilon\right\}$$

$$= \mathbf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\left\{|S_n^*| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\{|S_n^*| > z\}$$

$$z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n - np}{n}\right| > \epsilon\right\}$$

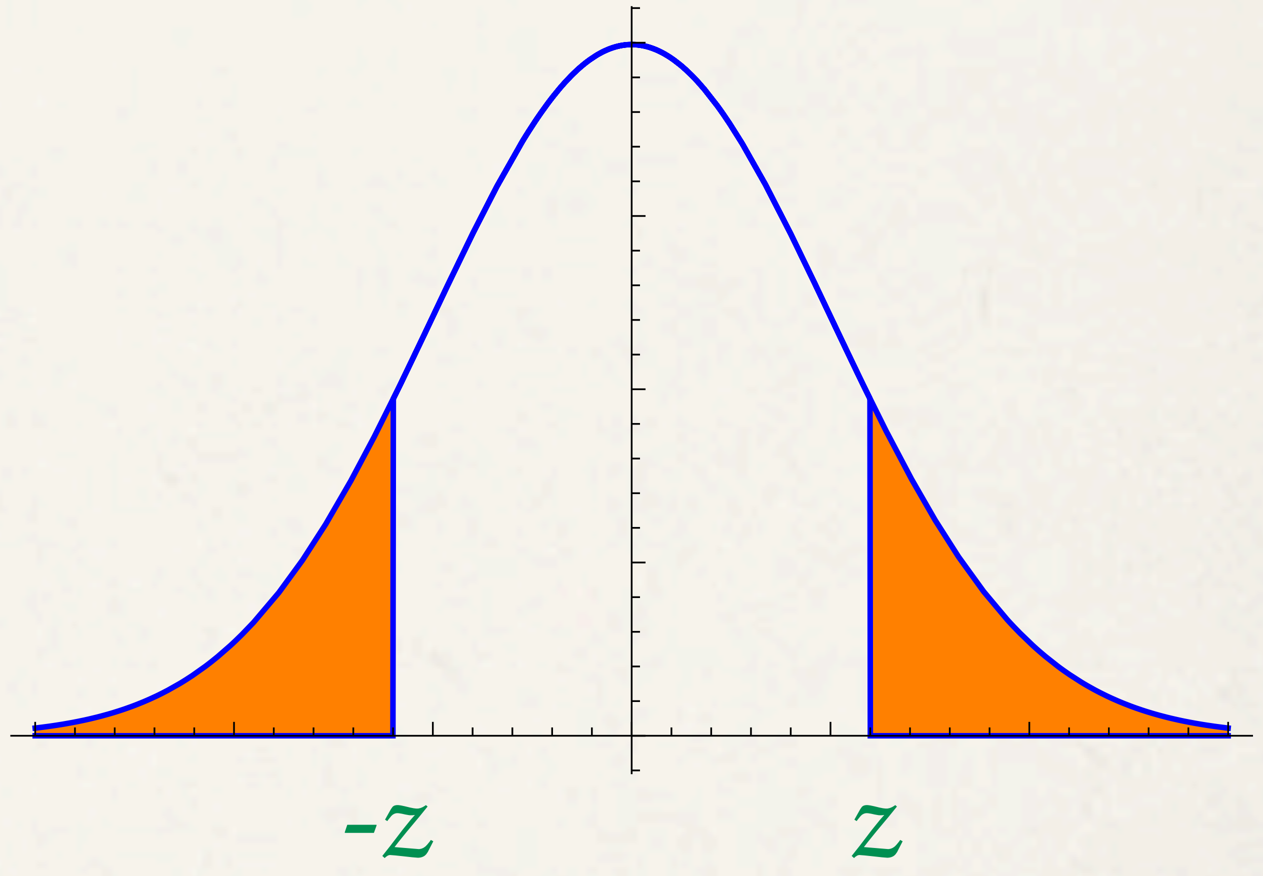
$$= \mathbf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\left\{|S_n^*| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\{|S_n^*| > z\}$$

$$z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

Normal approximation!



$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n - np}{n}\right| > \epsilon\right\}$$

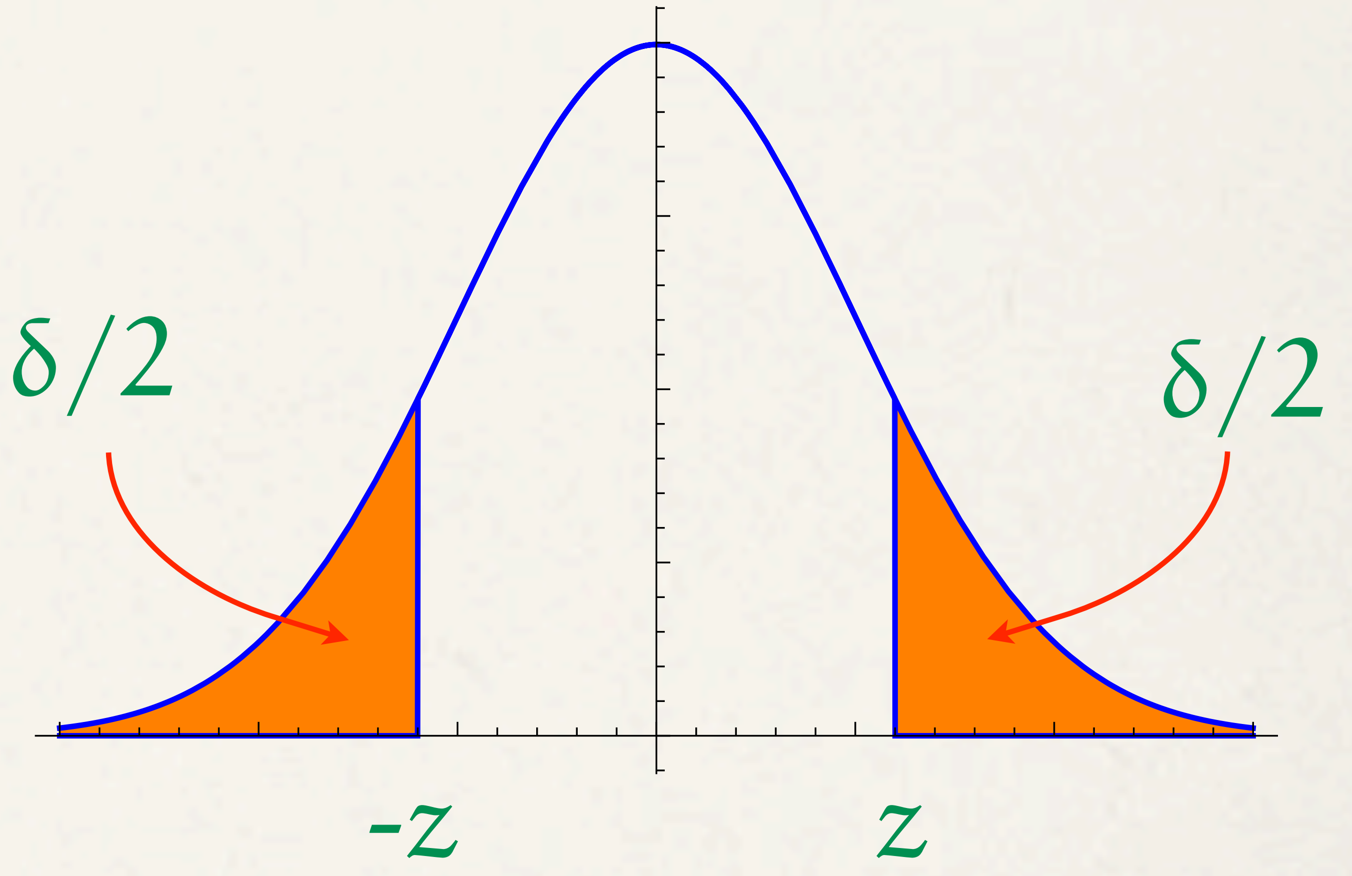
$$= \mathbf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\left\{|S_n^*| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\{|S_n^*| > z\}$$

$$z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

Normal approximation!



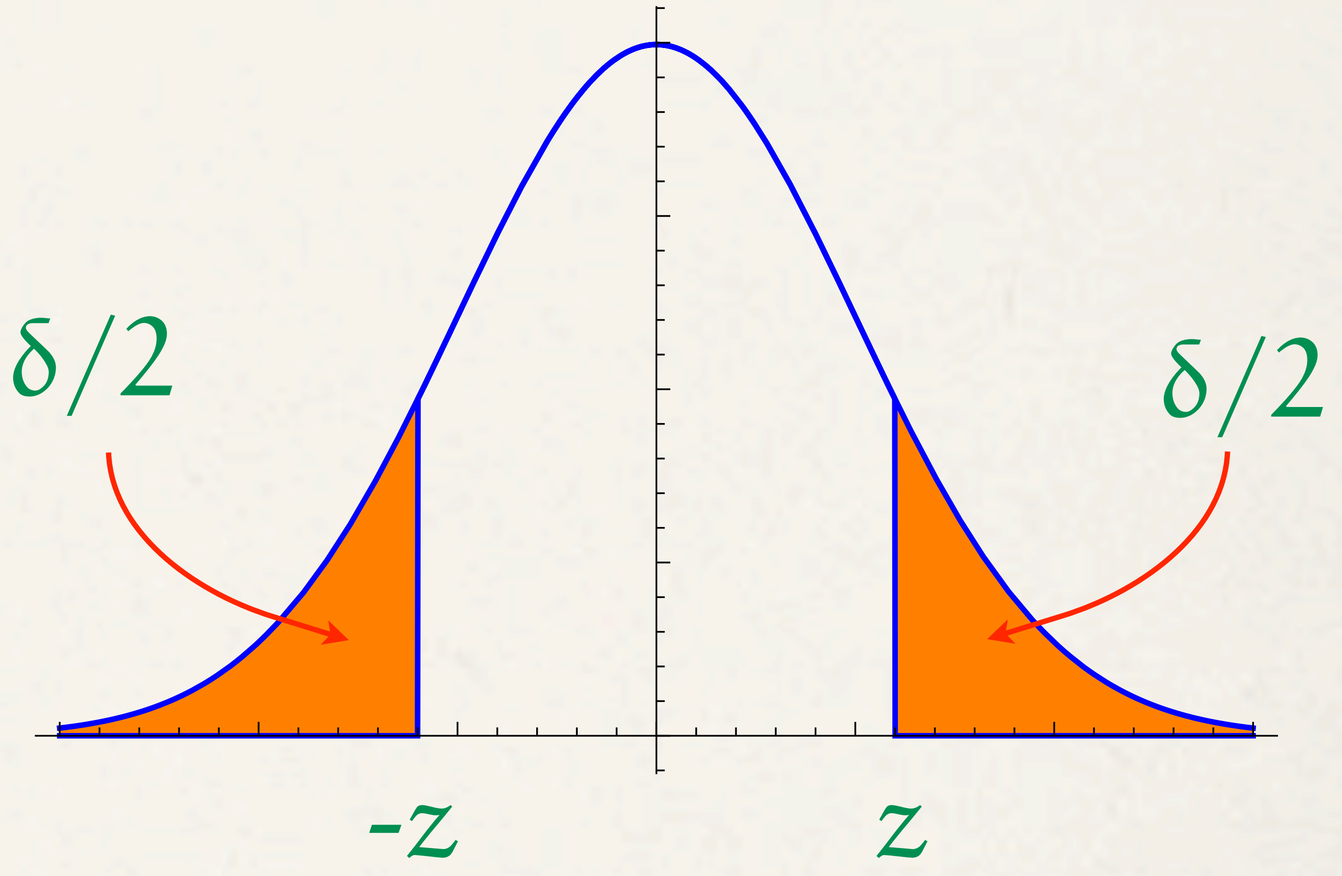
$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n - np}{n}\right| > \epsilon\right\}$$

$$= \mathbf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\left\{|S_n^*| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\{|S_n^*| > z\}$$

Normal approximation!



$$z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

The operating principle!

Select the sample size n so that the quantile $z = z(\delta)$ is the unique real number for which the area under the right tail of the bell curve is $\delta/2$.

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n - np}{n}\right| > \epsilon\right\}$$

$$= \mathbf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

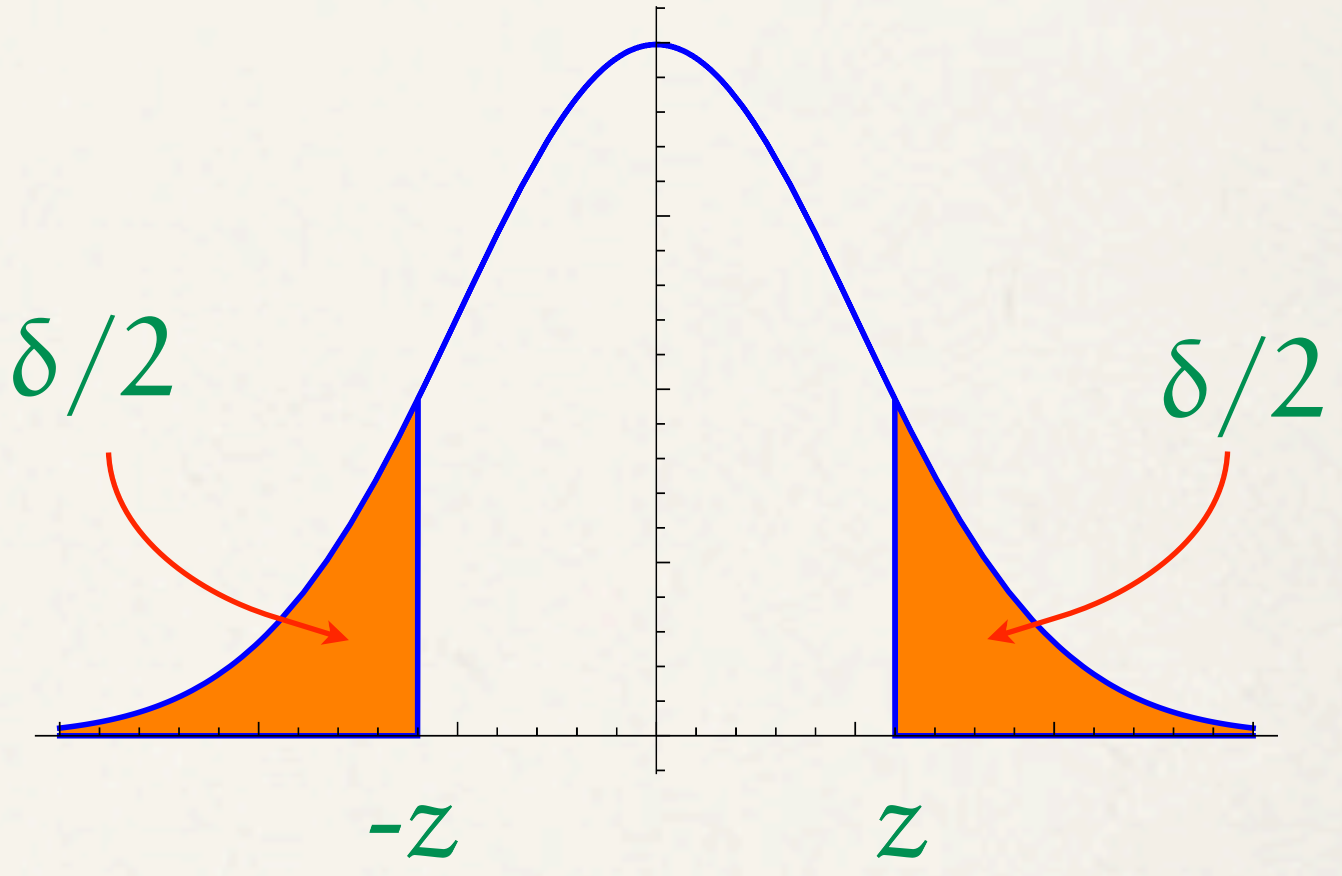
$$= \mathbf{P}\left\{|S_n^*| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$= \mathbf{P}\{|S_n^*| > z\}$$

$$\approx 2\int_z^\infty \phi(x)\,dx$$

$$z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

Normal approximation!



The operating principle!

Select the sample size n so that the quantile $z = z(\delta)$ is the unique real number for which the area under the right tail of the bell curve is $\delta/2$.

$$\mathbf{P}\left\{\left|\frac{S_n}{n}-p\right|>\epsilon\right\}=\mathbf{P}\left\{\left|\frac{S_n-np}{n}\right|>\epsilon\right\}$$

$$=\mathbf{P}\left\{\left|\frac{S_n-np}{\sqrt{npq}}\right|>\frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

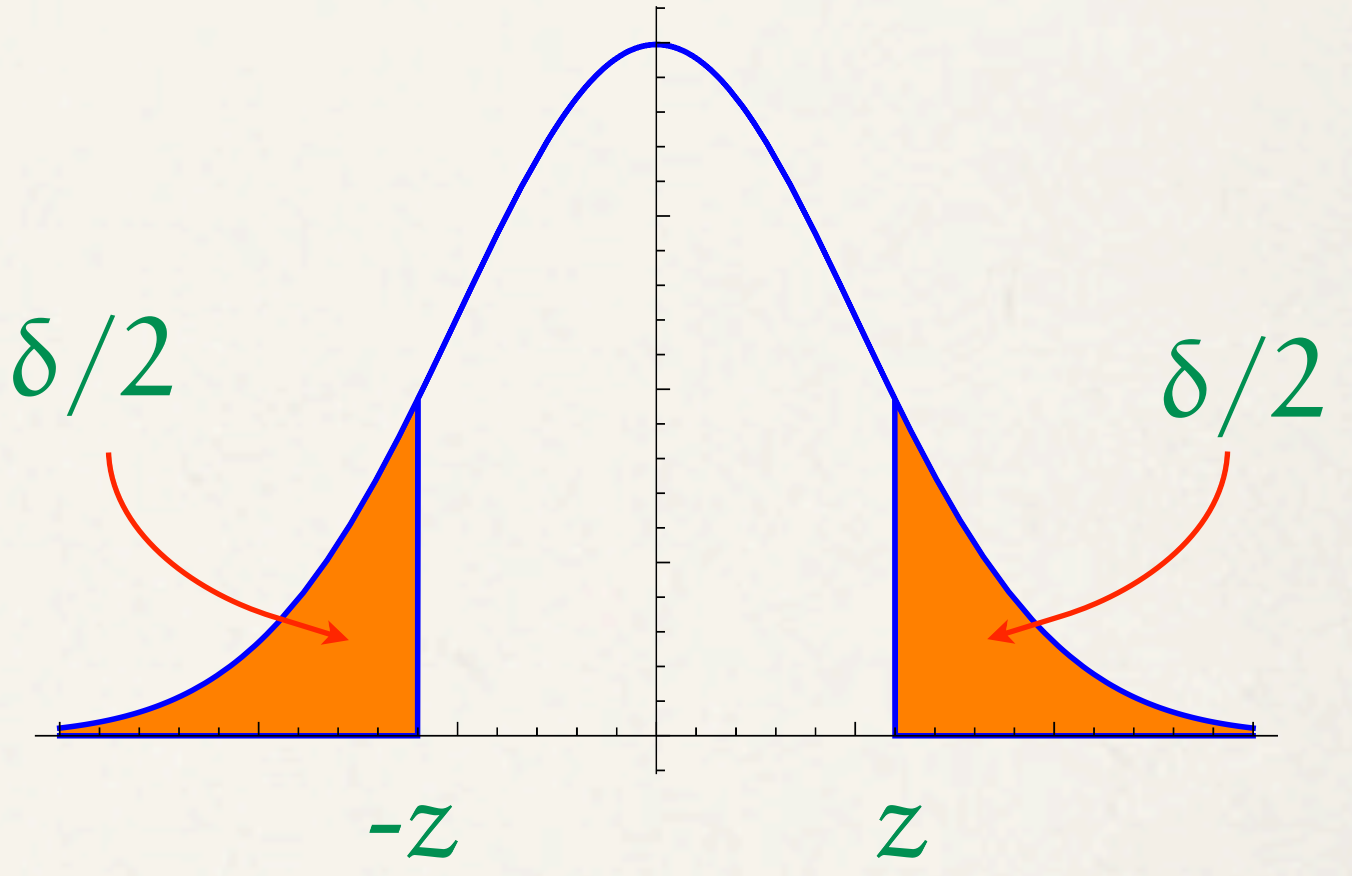
$$=\mathbf{P}\left\{|S_n^*|>\frac{\epsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$=\mathbf{P}\{|S_n^*|>z\}$$

$$\approx 2\int_z^\infty \phi(x)\,dx$$

$$=\delta$$

Normal approximation!

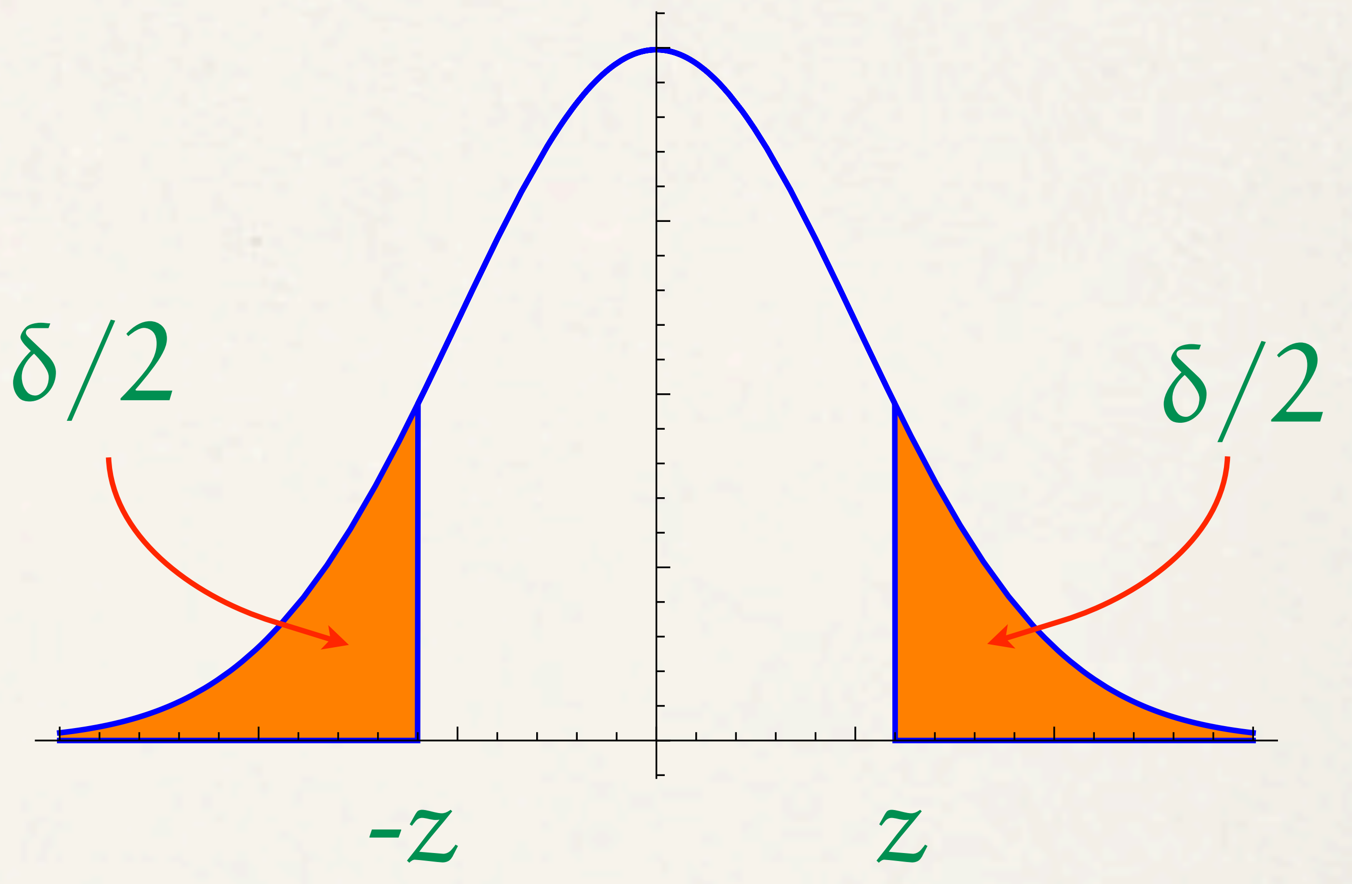


$$z=\frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

The operating principle!
Select the sample size n so that the quantile $z = z(\delta)$ is the unique real number for which the area under the right tail of the bell curve is $\delta/2$.

Normal approximation!

$$z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$



The operating principle!

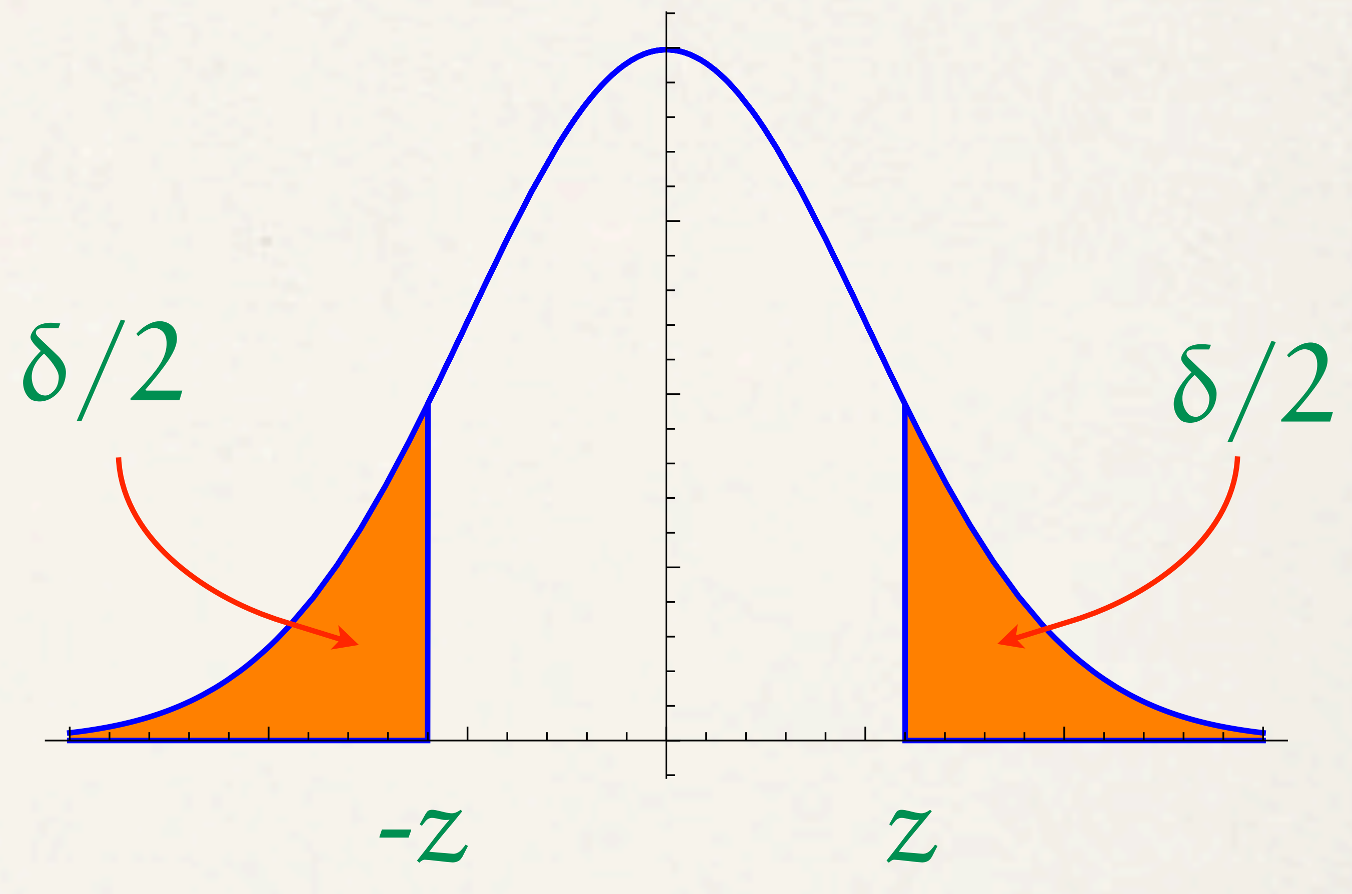
Select the sample size n so that the quantile $z = z(\delta)$ is the unique real number for which the area under the right tail of the bell curve is $\delta/2$.

The requisite statistical guarantee:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \leq \delta$$

$$z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

Normal approximation!



The operating principle!

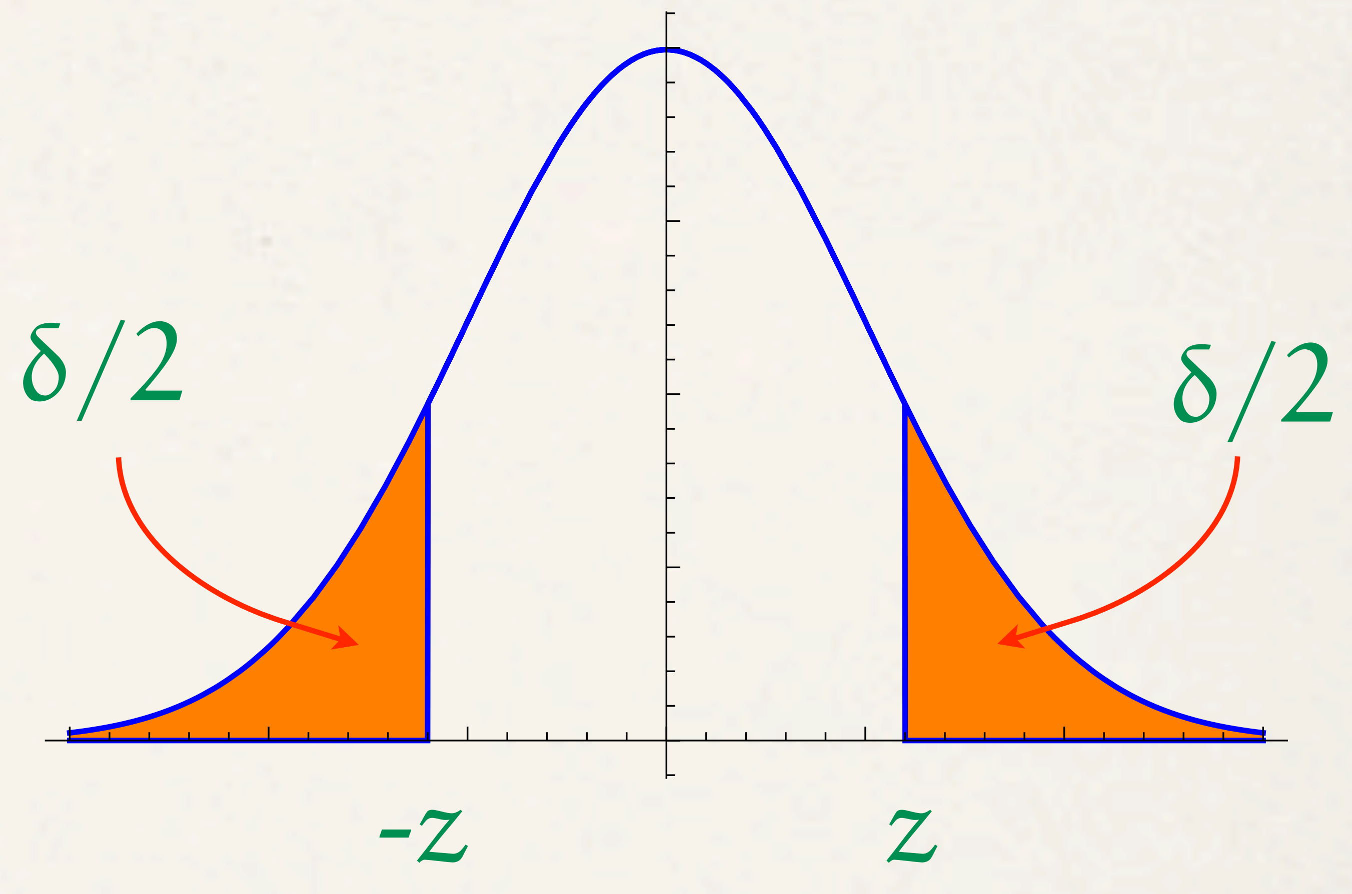
Select the sample size n so that the quantile $z = z(\delta)$ is the unique real number for which the area under the right tail of the bell curve is $\delta/2$.

The requisite statistical guarantee:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \le \delta$$

$$z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

Normal approximation!



The operating principle!

Select the sample size n so that the quantile $z = z(\delta)$ is the unique real number for which the area under the right tail of the bell curve is $\delta/2$.

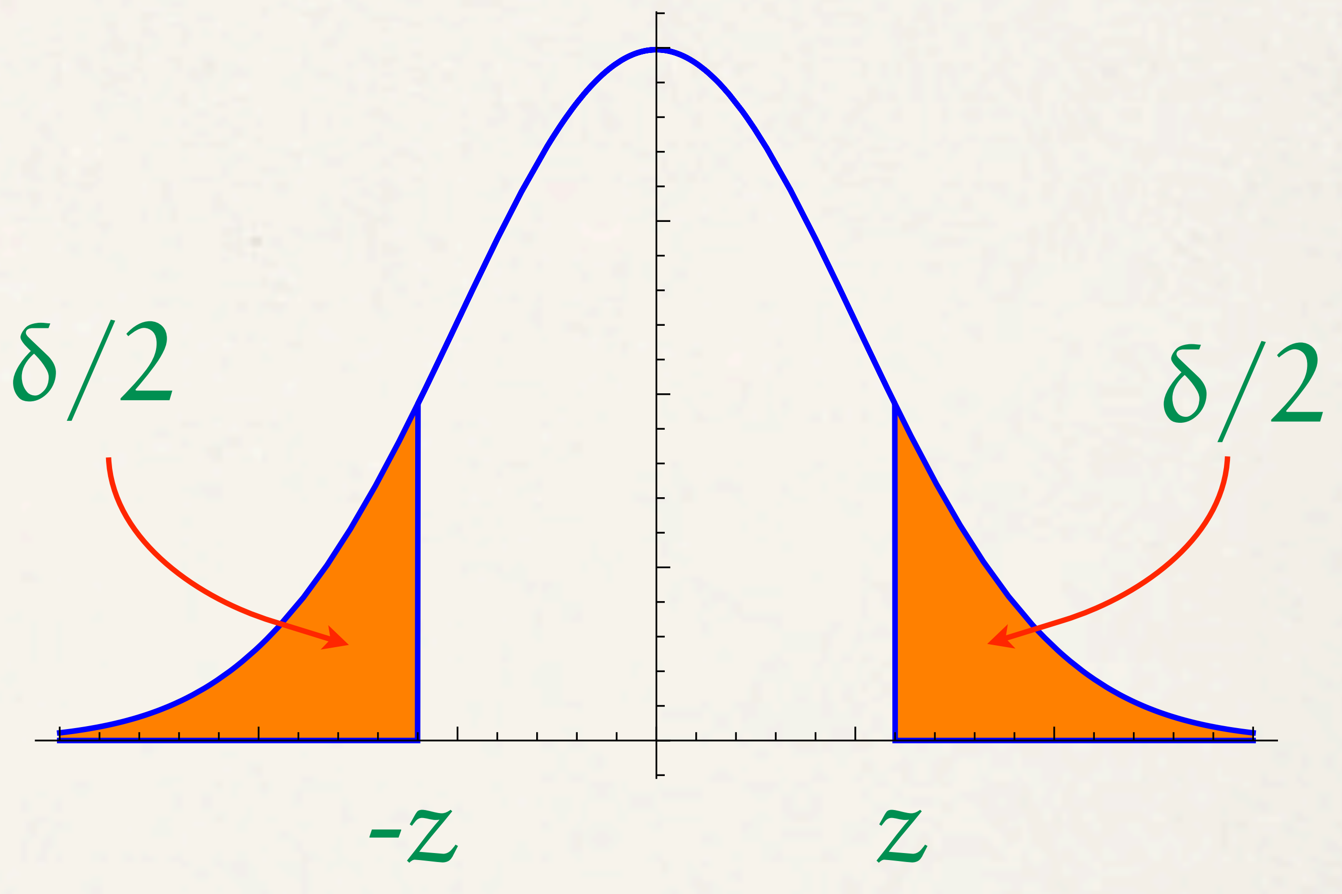
We require: $$n \ge \frac{z(\delta)^2 pq}{\epsilon^2}$$

The requisite statistical guarantee:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \leq \delta$$

$$z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

Normal approximation!



The operating principle!

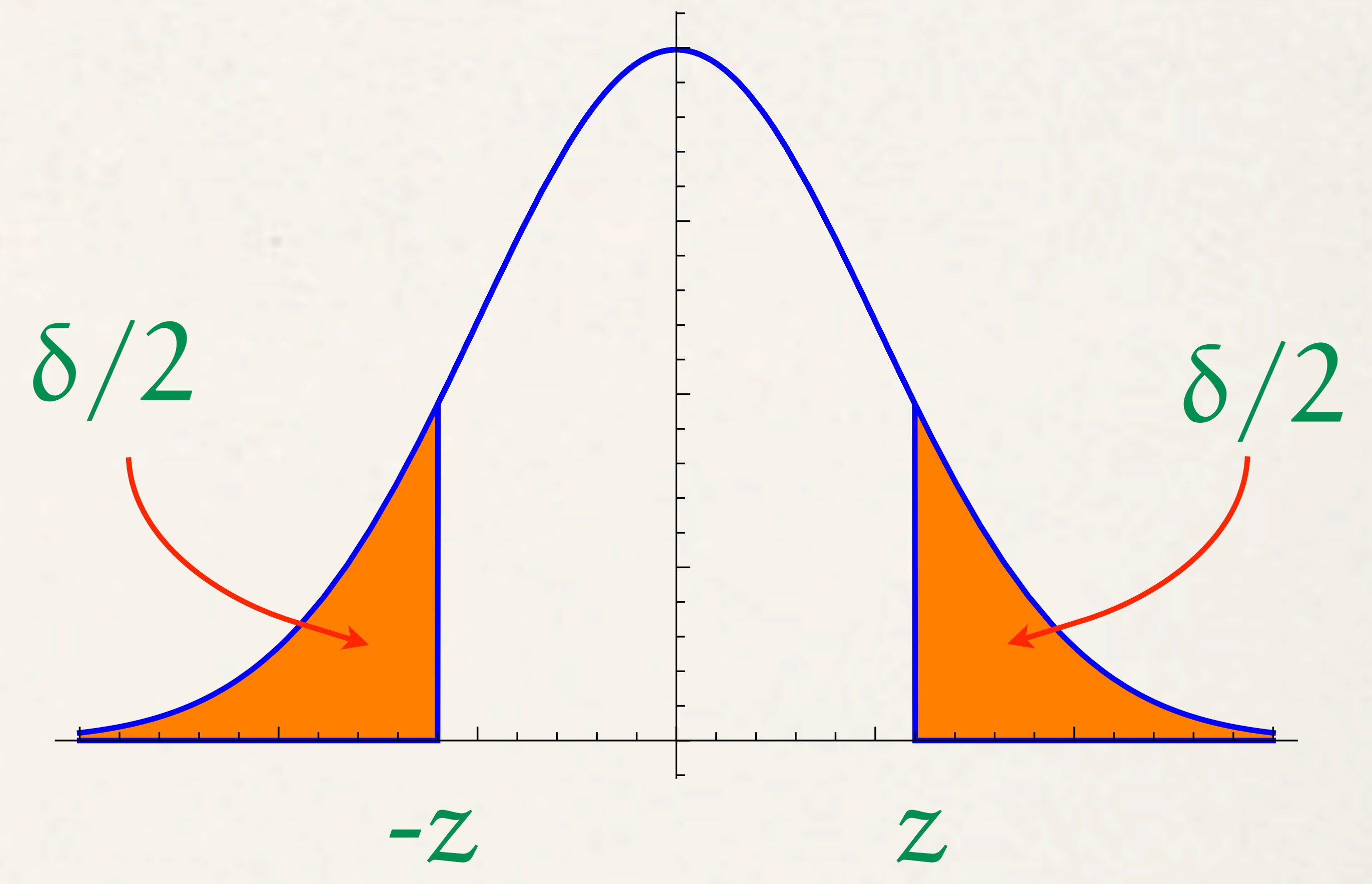
Select the sample size n so that the quantile $z = z(\delta)$ is the unique real number for which the area under the right tail of the bell curve is $\delta/2$.

We require: $$n \geq \frac{z(\delta)^2 pq}{\epsilon^2}$$ (recall $pq \leq 1/4$)

The requisite statistical guarantee:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \leq \delta \qquad z = \frac{\epsilon\sqrt{n}}{\sqrt{pq}}$$

Normal approximation!



The operating principle!

Select the sample size n so that the quantile $z = z(\delta)$ is the unique real number for which the area under the right tail of the bell curve is $\delta/2$.

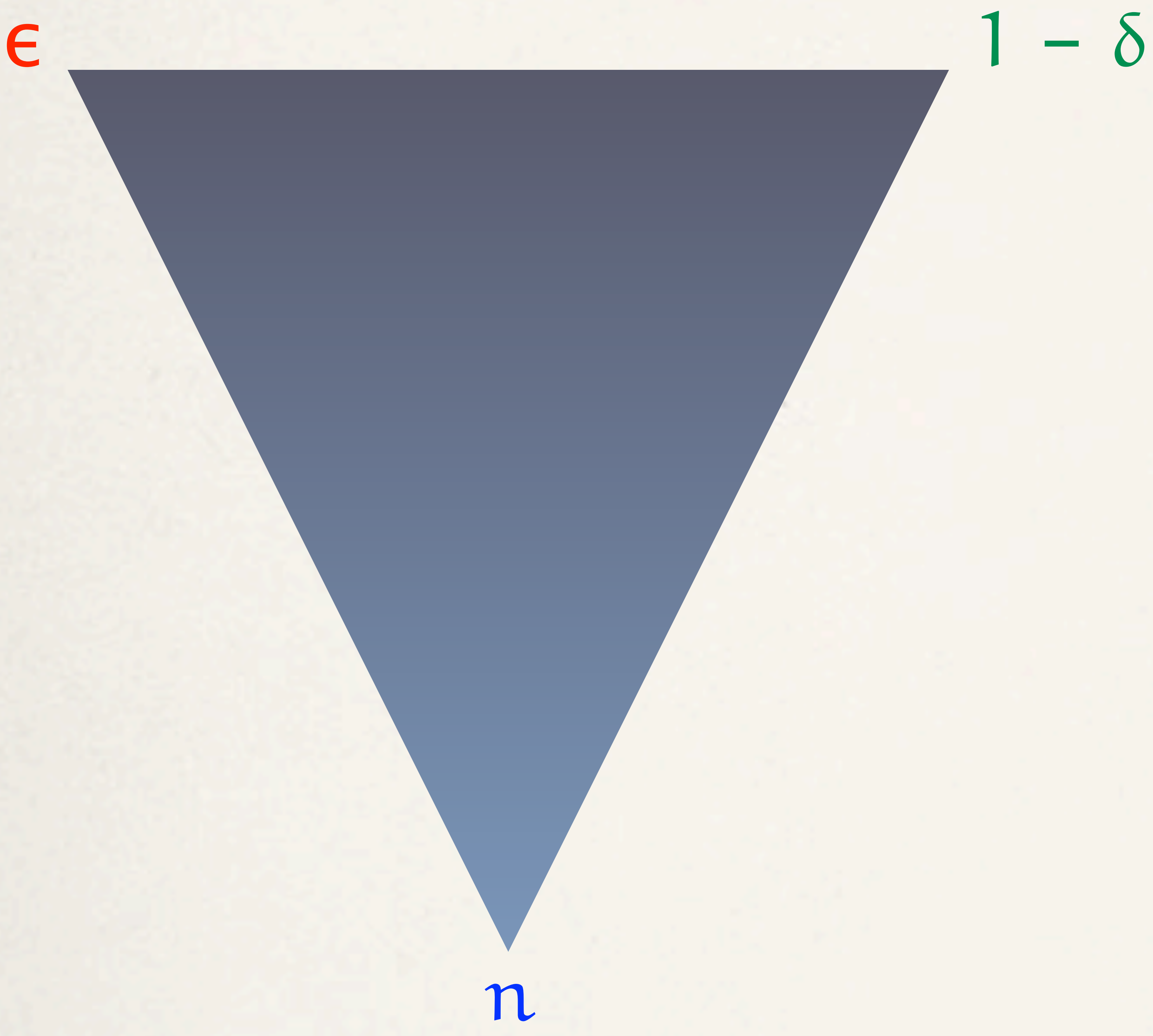
We require: $$n \geq \frac{z(\delta)^2 pq}{\epsilon^2}$$ (recall $pq \leq 1/4$)

An estimate of the requisite sample size under a normal approximation:

To guarantee that the estimate error is no larger than ϵ with confidence at least $1 - \delta$, for any value p, we will require a sample size of at least

$$n \geq \frac{z(\delta)^2}{4\epsilon^2}$$

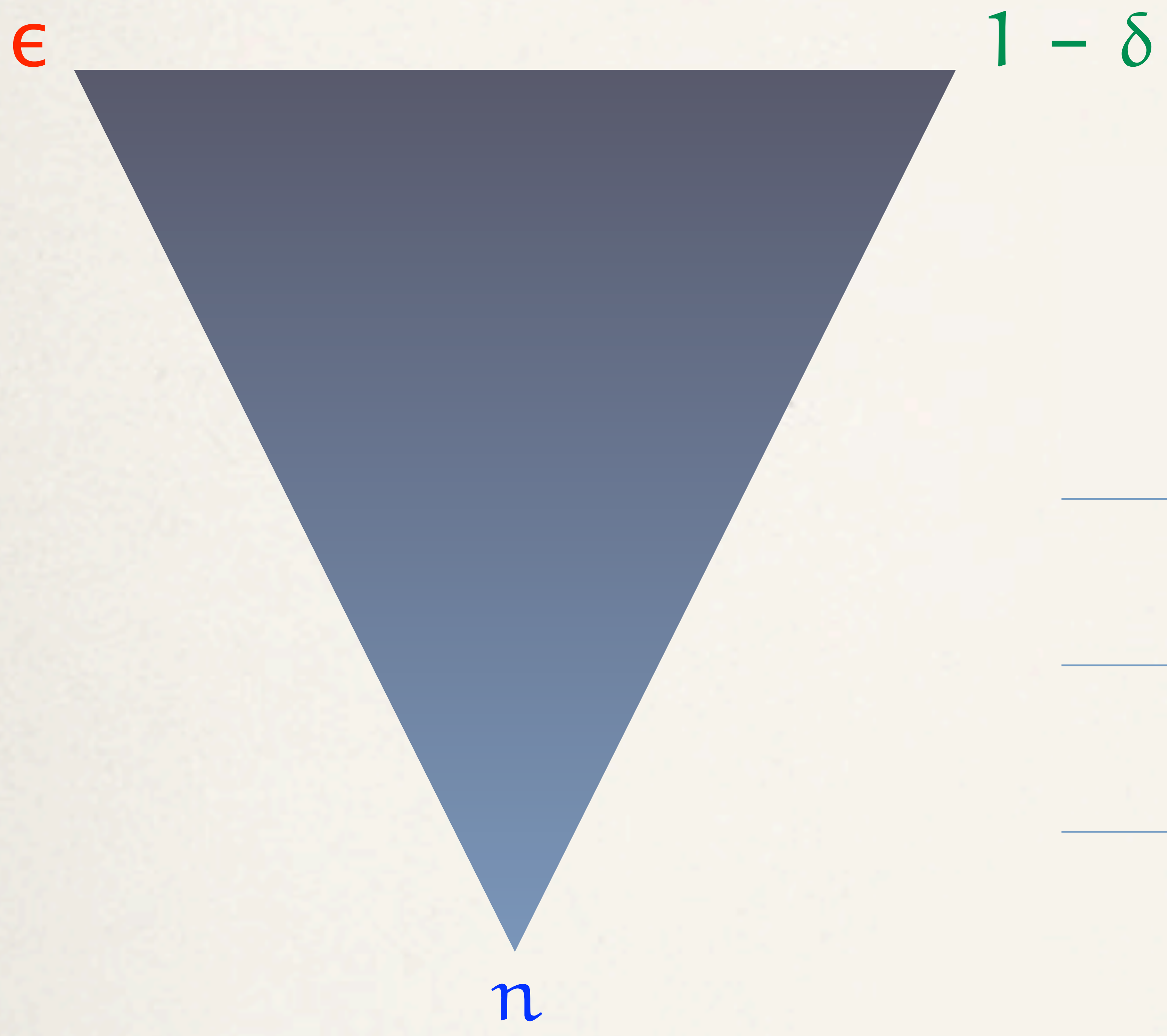
How are the error ϵ, the confidence $1 - \delta$, and the sample size n related? $n \geq \frac{z(\delta)^2}{4\epsilon^2}$



The requisite statistical guarantee:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \leq \delta$$

How are the error ϵ, the confidence $1 - \delta$, and the sample size n related? $\quad n \geq \dfrac{z(\delta)^2}{4\epsilon^2}$



Error ϵ	Confidence $1 - \delta$	Quantile $z(\delta)$	Sample size (Chebyshev) n	Sample size (normal approximation) n
0.10	0.90	1.65	250	68
0.05	0.95	1.96	2000	385
0.03	0.95	1.96	5556	1068

The requisite statistical guarantee:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \leq \delta$$

Slogan

A relatively small, honest, random sample *whose size does not depend upon the size of the underlying population or its composition* gives a good estimate of the underlying population proportions (sentiments).