Overview
- Polls and Statistical Inference
- Confidence Levels and Critical z-values
- Standard Error and Margin of Error
- Confidence Intervals

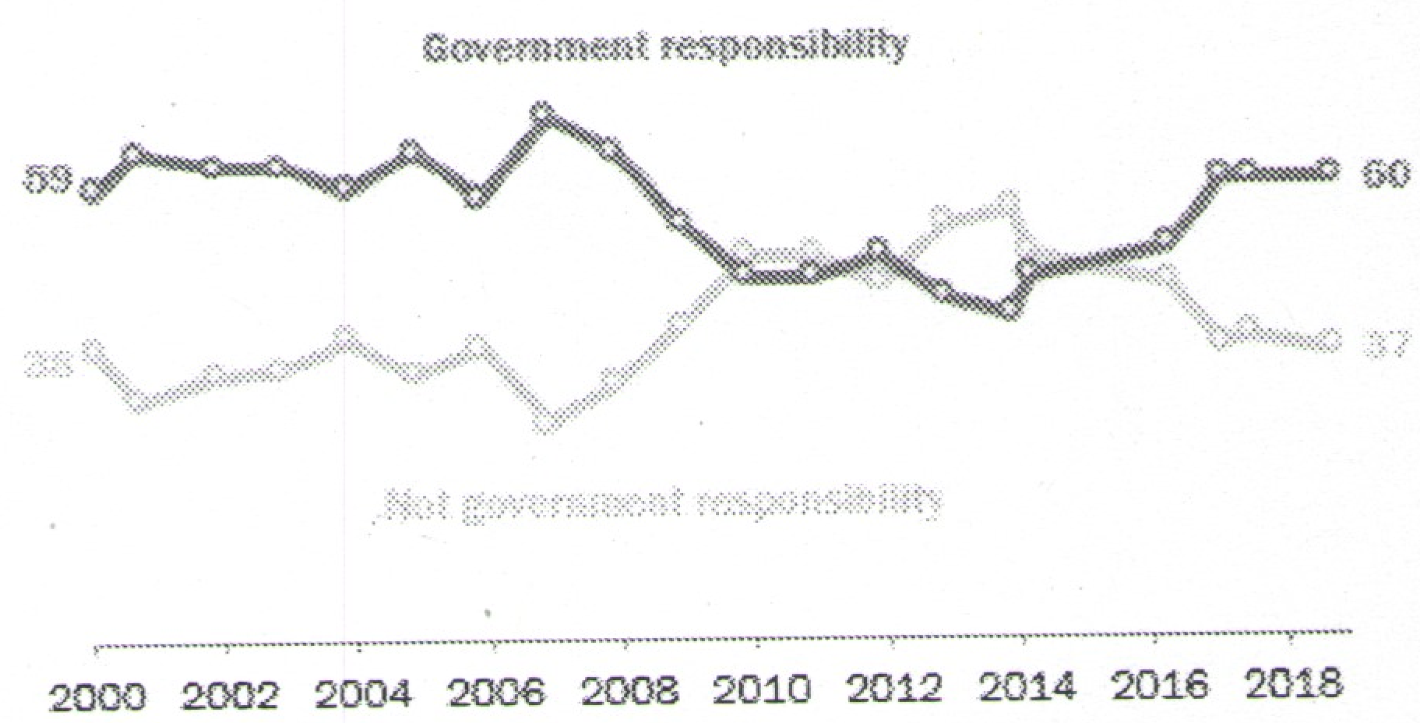
Statistical Polls

Go to http://www.pewresearch.org/fact-tank/2018/10/03/most-continue-to-say-ensuring-health-care-coverage-is-governments-responsibility/ to find the poll on U.S. healthcare.

Find the sample proportion who think ensuring healthcare is the government's responsibility. Click on the methodology link at the bottom of the article to find the sample size, and the margin of error. Then use the margin of error to write the confidence interval.

Majority continues to say ensuring health care coverage is a government responsibility

Is it the responsibility of the federal government to make sure that all Americans have health care coverage? (%)



Notes: 2000-2013 data from Gallup.
Don't know responses not shown.
Source: Survey conducted Sept. 18-24, 2018.
PEW RESEARCH CENTER

$\frac{6}{10}$ 60%

Sample Proportion, $\hat{p}$: .60 (point estimate) Margin of Error: ±2.7% ±.027

Sample Size: 1754 adults Confidence Level: 95%

A **confidence interval** is the point estimate ± the margin of error for the given confidence level. It is more accurate to give a range for the population proportion rather than a point estimate.

Confidence Interval: (.6 − .027, .6 + .027)
(.573, .627) or (57.3%, 62.7%)

Statistical Inference

Now that we have studied the sampling distribution of a proportion, $\hat{p}$, we can begin to look at **inferential statistics**. That is, we want to take a single sample and make an estimate of the population parameter, which we do not know.

Confidence Levels and Critical z-values – How many standard deviations from the mean?

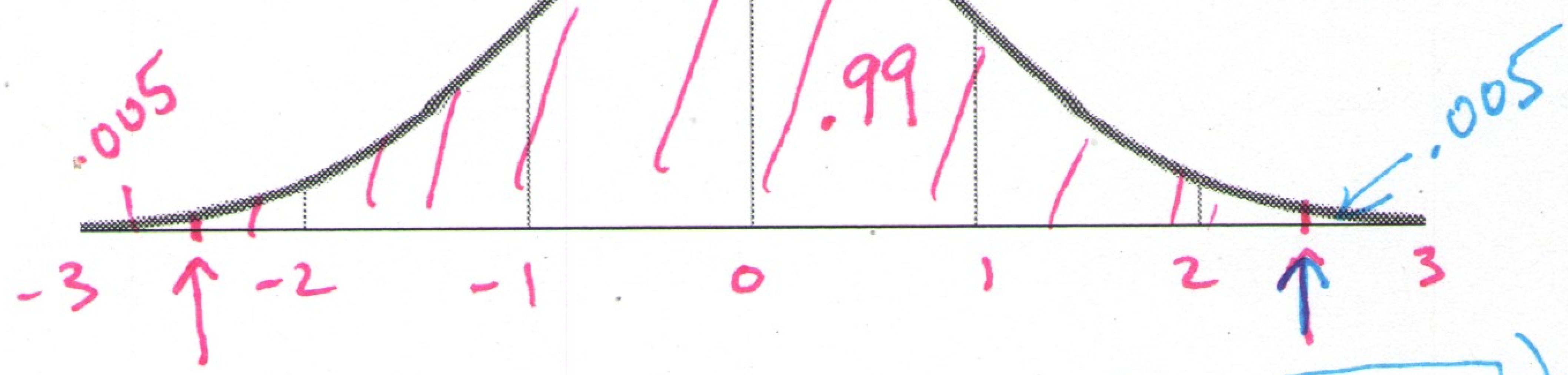
Identify the critical z-scores for 99%, 95%, 90% and 80% confidence levels.

99% Confidence

z* = 2.5758

1-.99 = .01

.01÷2



-2.5758

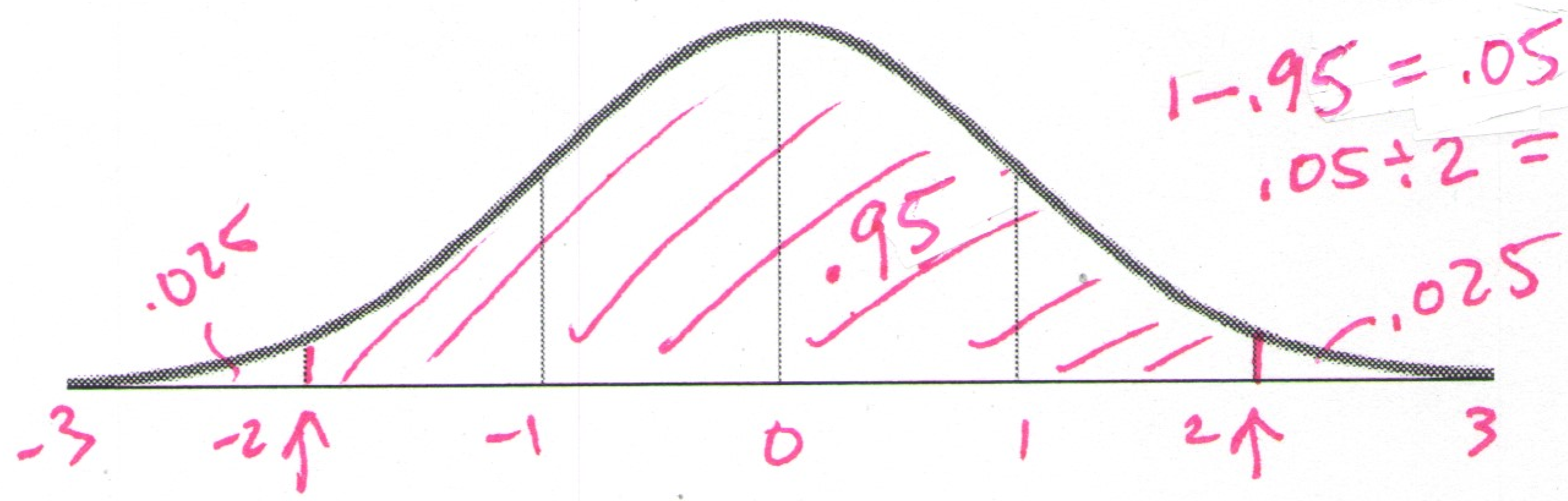
$P(Z > \boxed{2.5758}) = .005$

95% Confidence

z* = 1.96

1-.95 = .05

.05÷2 =



-1.96

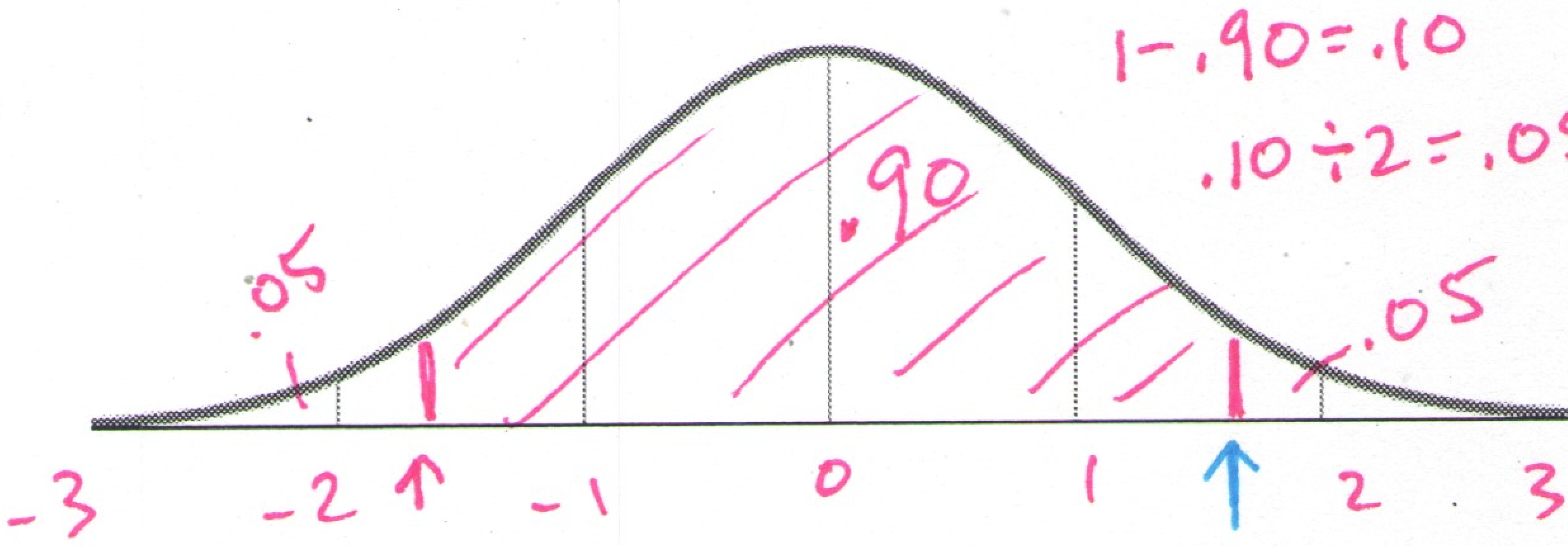
$P(Z > \boxed{1.96}) = .025$

90% Confidence

z* = 1.6449

1-.90 = .10

.10÷2 = .05



-1.6449

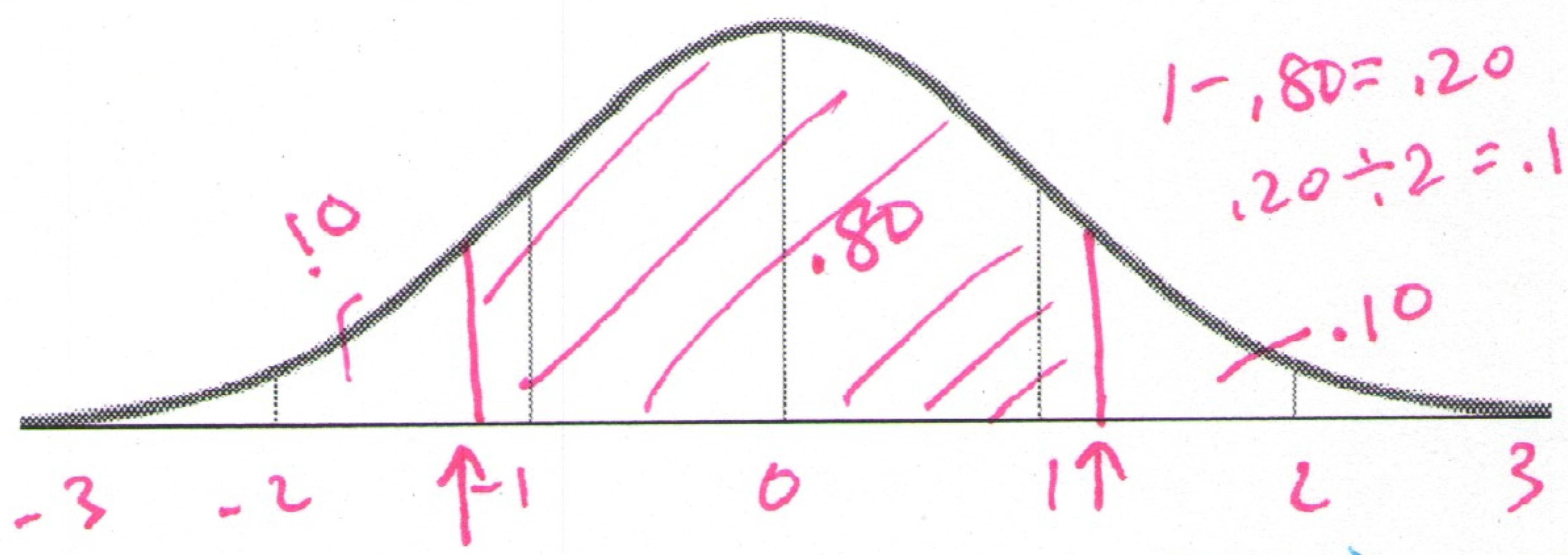
$P(Z > \boxed{1.6449}) = .05$

80% Confidence

z* = 1.2816

1-.80 = .20

.20÷2 = .10



$P(Z > \boxed{1.2816}) = .10$

Standard Error

The **Standard Error (SE)** is an estimate of the standard deviation of the sampling distribution of a proportion. It's used when we don't know the value of p and are not able to determine the true standard deviation.

$\hat{p}$ is the sample proportion or the proportion who answered "yes" in the sample.

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Example 1. A survey of 2,000 hiring managers showed that 1,200 use social media sites to research job applicants.

a. State the sample proportion and calculate the standard error for this sample proportion.

$\hat{p} = \frac{1200}{2000} = .60$ $\quad$ $SE = \sqrt{\frac{.60(.40)}{2000}} = .0110$

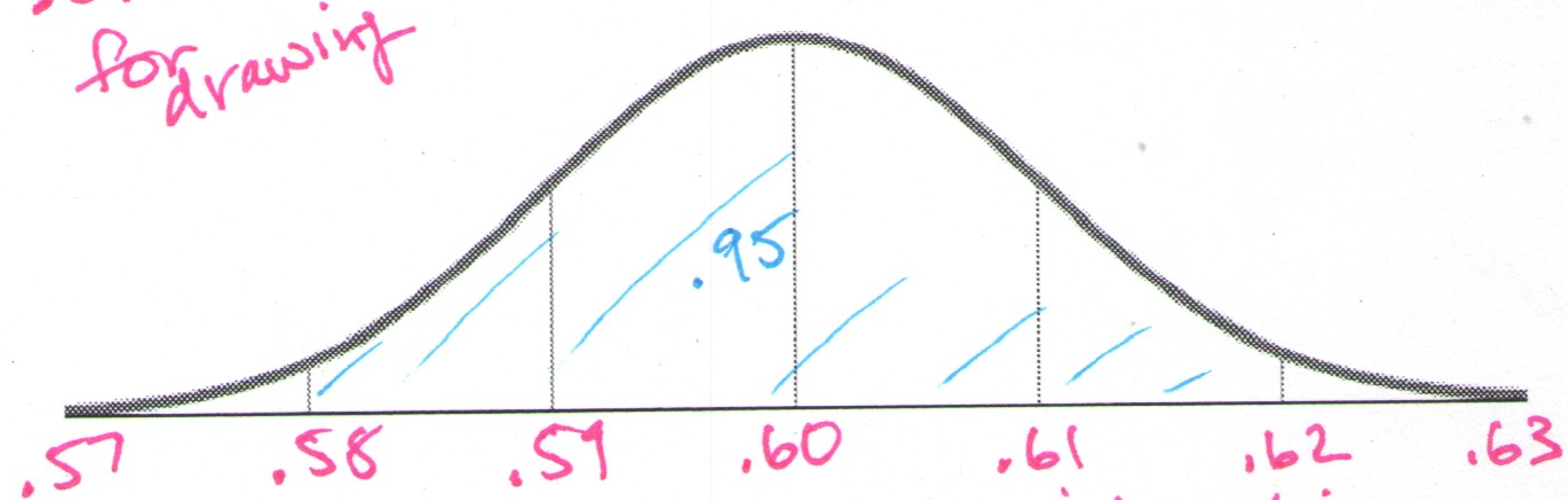
4 decimal places

b. Check the four conditions needed to use a sampling distribution for $\hat{p}$. Then draw and label the distribution.

1. independence – the hiring managers must be independent of each other in hiring practices
2. randomization – take a random sample
3. 10% condition – 2000 is less than 10% of all hiring managers
4. success/failure – $np = 2000(.6) = 1200 \geq 10$ ✓
   $nq = 2000(.4) = 800 \geq 10$ ✓

$\hat{p} \sim N(.60, .0110)$

.01 for drawing



proportion who use social media for hiring in samples of 2000.

Based on the confidence levels we found above, we would expect 95% of random samples to fall within 1.96 standard deviations of the mean. Our margin of error is 1.96 times the standard error.

about (.58, .62)

Margin of Error ± part

The **Margin of Error (ME)** is an estimate that expresses the amount of sampling error in the results of a survey. When you see something like ±3 percentage points, that is the margin of error.

$$ME_{\hat{p}} = \pm z^* \cdot SE_{\hat{p}} \quad \text{or} \quad ME_{\hat{p}} = \pm z^* \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Where z* is the critical z-score value that corresponds to the desired confidence level.

c. Continuing the previous example, calculate the margin of error at the 95% confidence level.

$$ME = \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$= \pm 1.96 \sqrt{\frac{(.60)(.40)}{2000}}$$

$$= \pm .0215$$

d. Write the confidence interval for the sample proportion.

$(.60 - .0215, .60 + .0215)$

$(.5785, .6215)$

e. Interpret the confidence interval:

We are 95 % confident that the true proportion of hiring managers that use social media (insert context here)

is between 57.85 % and 62.15 %.

Confidence Intervals

If a point estimate follows the normal model with standard error SE, then a **Confidence Interval** for the population parameter is

$$\hat{p} \pm z^* \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

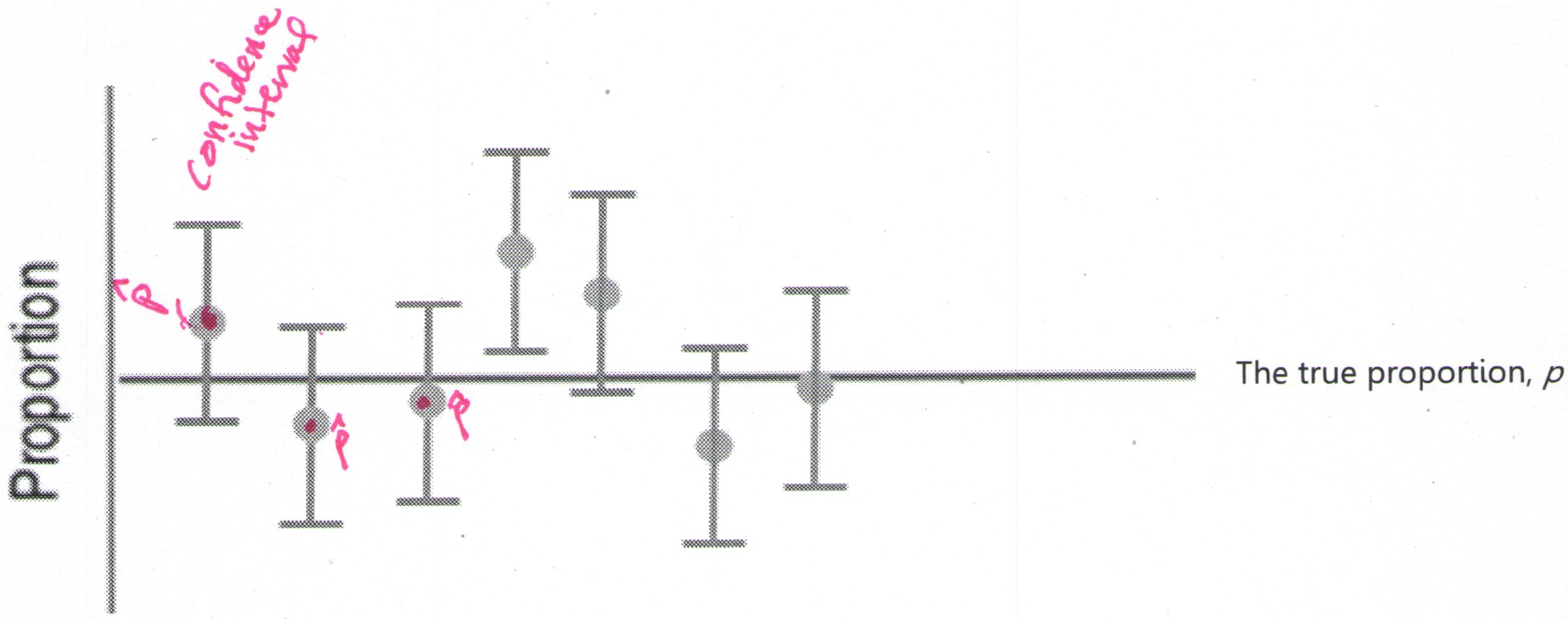
Where z* is the critical z-score value that corresponds to the desired confidence level.

Critical z-scores:		
	99% Confidence	$z^* =$ 2.5758
	95% Confidence	$z^* =$ 1.96
	90% Confidence	$z^* =$ 1.6449
	80% Confidence	$z^* =$ 1.2816

What does 95% confidence mean?

If we collected random samples over and over, with the same sample size:

- Each time we would get a different sample proportion, $\hat{p}$.
- From each $\hat{p}$, a different Standard Error, Margin of Error, and confidence interval would be computed.
- About 95% of these confidence intervals would capture the true proportion.
- About 5% would miss the true proportion.



Interpretation of Confidence Intervals

We need to convey that the uncertainty is in the interval, not in the true proportion. That is why we use confidence rather than probability.

Technically Correct:

> We are 95% confident that the interval from 58% to 62% captures the true proportion of hiring managers who use social media to research job applicants.

More Casual, But Fine

> We are 95% confident that between 58% and 62% hiring managers use social media to research job applicants.

Incorrect

> There is a 95% probability that the true proportion is between 58% and 62%.

The true proportion is either in the interval or not. The randomness is not in the true proportion but in the confidence interval.

Example 2. A 2012 poll asked 166 adults whether they were baseball fans; 48% said that they were.

a. Construct a 99% confidence interval for the true proportion of US adults that are baseball fans.

$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$.48 \pm 2.5758 \sqrt{\frac{.48(.52)}{166}}$

$.48 \pm .0999$

$(.48 - .0999,\ .48 + .0999)$

$(.3801,\ .5799)$

$(38\%,\ 58\%)$

b. Construct a 95% confidence interval for the true proportion of US adults that are baseball fans.

$.48 \pm 1.96 \sqrt{\frac{.48(.52)}{166}}$

$.48 \pm \underbrace{.0760}_{\text{margin of error}}$

$(.48 - .0760,\ .48 + .0760)$

$(.404,\ .556)$

c. Construct a 90% confidence interval for the true proportion of US adults that are baseball fans.

$.48 \pm 1.6449 \sqrt{\frac{.48(.52)}{166}}$

$.48 \pm .0638$

$(.4162,\ .5438)$

d. Construct an 80% confidence interval for the true proportion of US adults that are baseball fans.

$.48 \pm 1.2816 \sqrt{\frac{.48(.52)}{166}}$

$.48 \pm .0497$

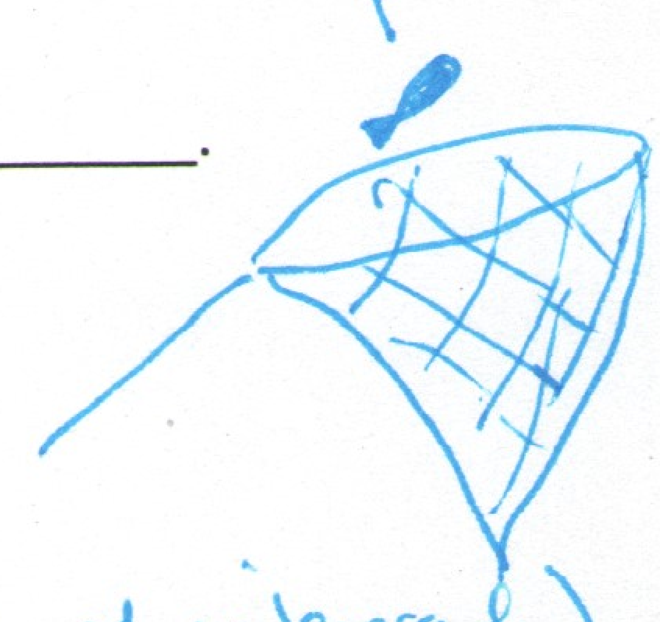
$(.4303,\ .5297)$

We are 80% confident that the true proportion of baseball fans is between 43% and 53%.

Certainty vs. Precision

As the confidence level increases, the interval gets wider.

$\hat{p}$



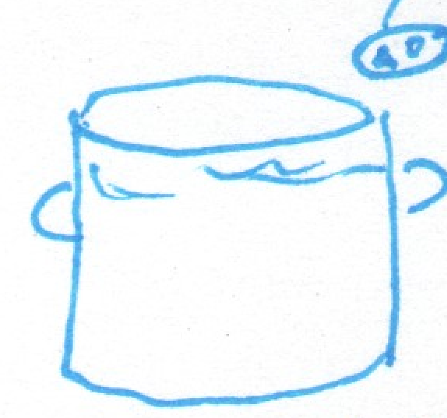
Why? To be more confident we need a wider interval

Trade-off between precision (size of interval) and confidence level.

Determining Sample Size for a desired Margin of Error

Sample size

representative sample



doesn't need to be large

Using Algebra we can solve the margin of error formula for n:

$$ME_{\hat{p}} = \pm z^* \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\frac{ME}{z^*} = \frac{\cancel{z^*}}{\cancel{z^*}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\left(\frac{ME}{z^*}\right)^2 = \left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)^2$$

$$n\cdot\left(\frac{ME}{z^*}\right)^2 = \frac{\hat{p}\hat{q}}{\cancel{n}}\cdot\cancel{n}$$

$$\left(\frac{z^*}{ME}\right)^2 n\left(\frac{ME}{z^*}\right)^2 = \hat{p}\hat{q}\left(\frac{z^*}{ME}\right)^2$$

$$n = \hat{p}\hat{q}\left(\frac{z^*}{ME}\right)^2$$

The sample size needed to get a desired margin of error (ME) is given by the formula:

$$n = \hat{p}\hat{q}\left(\frac{z^*}{ME}\right)^2$$

decimal form

Example 3. It's believed that 25% ($\hat{p}$) of adults over 50 never graduated high school. We wish to see if the same is true among 25 to 30 year olds.

a. How many of this younger age group must we survey in order to estimate the proportion of non-grads to within 6% with 90% confidence?

±.06 margin of error

$$n = .25(.75)\left(\frac{1.6449}{.06}\right)^2$$

$$= 140.92$$

141 people

always round up

b. Suppose we want to cut the margin of error to 4% (.04) (again with 90% confidence). What's the necessary sample size?

$$n = .25(.75)\left(\frac{1.6449}{.04}\right)^2$$

$$= 317.07$$

$\approx$ 318 people

c. What is the relationship between the number of people sampled and the margin of error?

To get a smaller margin of error we need to make our sample size larger.

Practice 1. A 2016 Gallup poll asked 1021 U.S. adults whether they are satisfied with their current healthcare and 581 people said they are satisfied.

a. Give a 95% confidence interval for the true proportion of U.S. adults who are satisfied with their healthcare.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad \hat{p} = \frac{581}{1021} = .5690$$

$$.5690 \pm 1.96\sqrt{\frac{.5690(.431)}{1021}}$$

$$.569 \pm .0304$$

$$(.5386, .5994) \quad \text{or} \quad (54\%, 60\%)$$

b. Explain what your interval means.

we are 95% confident that the true proportion of people who are satisfied with their current healthcare is between 54% and 60%.

Practice 2. An article titled "Tongue Piercing May Speed Tooth Loss, Researchers Say" found that 18 out of 52 participants had receding gums, which can lead to tooth loss.

$$\hat{p} = \frac{18}{52} = .3462$$

a. How many people need to be surveyed in order to estimate the proportion of pierced-tongue people with receding gums to within 3% with 95% confidence?

1.96

$$n = \hat{p}\hat{q}\left(\frac{z^*}{ME}\right)^2$$

$$= (.3462)(1-.3462)\left(\frac{1.96}{.03}\right)^2$$

$$= 966.14$$

we need to survey 967 people to have a margin of error of 3%.

b. Suppose we decide that a margin of error of 8% would be sufficient (again with 95% confidence). What's the necessary sample size?

$$n = (.3462)(1-.3462)\left(\frac{1.96}{.08}\right)^2$$

$$= 135.86$$

we would only need to survey 136 people to get a margin of error of ±8