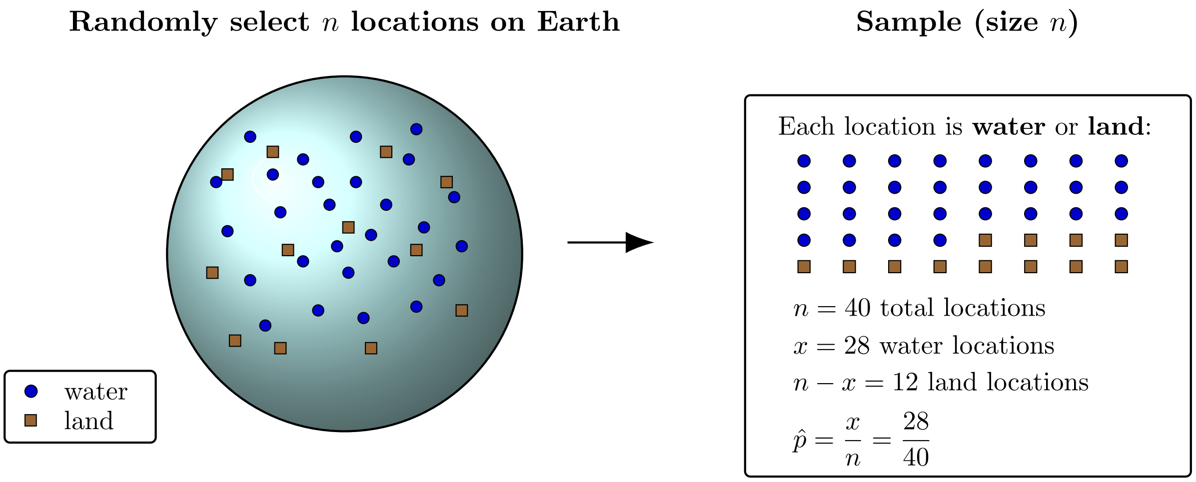
Estimating a Population Proportion with Confidence

AP Statistics · Mr. Merrick · February 1, 2026

What proportion of Earth's surface is covered by water? Here the *parameter* aim to study is p (a population proportion). In this context we assume p is a statistic theoretical value.

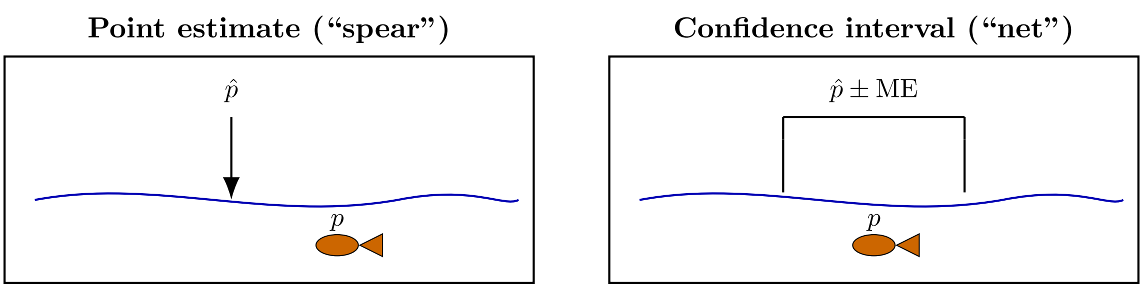
How might we estimate p using random sampling?

- **Step #1:** Randomly select n locations on Earth.
- **Step #2:** Classify each location as **water** or **land**.



Why is a point estimate flawed?

- A point estimate (statistics) reports only one number.
- Different random samples give different $\hat{p}$ values.
- It does not reflect sampling variability.



A confidence interval "casts a net" around $\hat{p}$. If the sampling conditions are satisfied, the interval gives a reliable range of plausible values for p.

Catching the fish with 95% confidence

Goal: Using our random sample of locations on Earth, we want to build a net that captures the true proportion of Earth covered by water, p, about **95% of the time**.

Step 1: Can we trust the method? (Check conditions)

Before constructing a confidence interval, we must verify the conditions for inference:

- **Random:** The n locations were selected randomly from Earth.
- **Independence (10% condition):** The sample size is less than 10% of all possible locations on Earth.
- **Normal (Large Counts):**
$$n\hat{p} \geq 10 \quad \text{and} \quad n(1-\hat{p}) \geq 10$$

When these conditions are satisfied, the sampling distribution of $\hat{p}$ is approximately Normal.

Step 2: What do we know about $\hat{p}$?

If the conditions are met, then:

$$\hat{p} \sim N\left(\mu_{\hat{p}} = p, \;\; \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\right)$$

Since p is unknown, we estimate the standard deviation using $\hat{p}$.

Step 3: Capturing the fish

For a Normal distribution,

$$P(-1.96 < Z < 1.96) = 0.95$$

This means that about 95% of standardized values fall within 1.96 standard deviations of the mean.

$$P\left(-1.96 < \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < 1.96\right) \approx 0.95$$

Rewriting gives:

$$P\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \approx 0.95$$

Step 4: Construct the interval for Earth

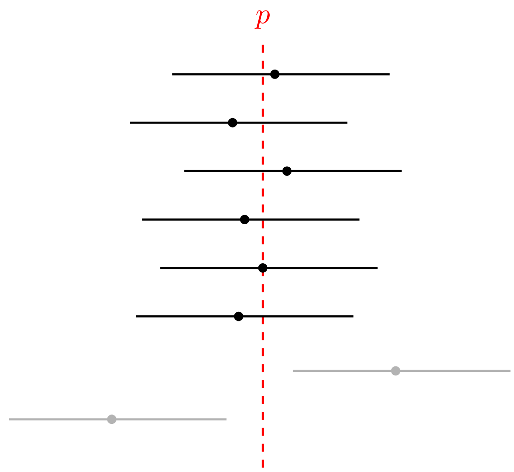
$$\frac{28}{40} \;\pm\; 1.96\sqrt{\frac{\left(\frac{28}{40}\right)\left(1-\frac{28}{40}\right)}{40}}$$

$$0.70 \pm 0.142 \;=\; (0.558,\; 0.842)$$

Step 5: Interpret the interval (conclusion)

We are 95% confident that the true proportion of Earth's surface covered by water is between 0.558 and 0.842 (55.8% to 84.2%).

If we repeated this random sampling process many times, about 95% of the intervals we construct would contain the true value of p.



The steps we used for Earth apply to *any* population proportion.

$$\text{point estimate} \;\pm\; \text{margin of error}$$

$$\text{statistic} \;\pm\; \text{(critical value)(standard error of statistic)}$$

$$\hat{p} \;\pm\; z^\star \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $z^\star$ depends on the chosen confidence level ($z^\star = 1.96$ for 95% confidence).

Common misconceptions about confidence

What 95% confidence does *not* mean:

- It does **not** mean there is a 95% probability that p is in this interval.
- It does **not** describe the variability of p.

What 95% confidence *does* mean: If we repeatedly take random samples and build intervals using this method, about 95% of them will contain the true value of p.