Inference

With mathematical models

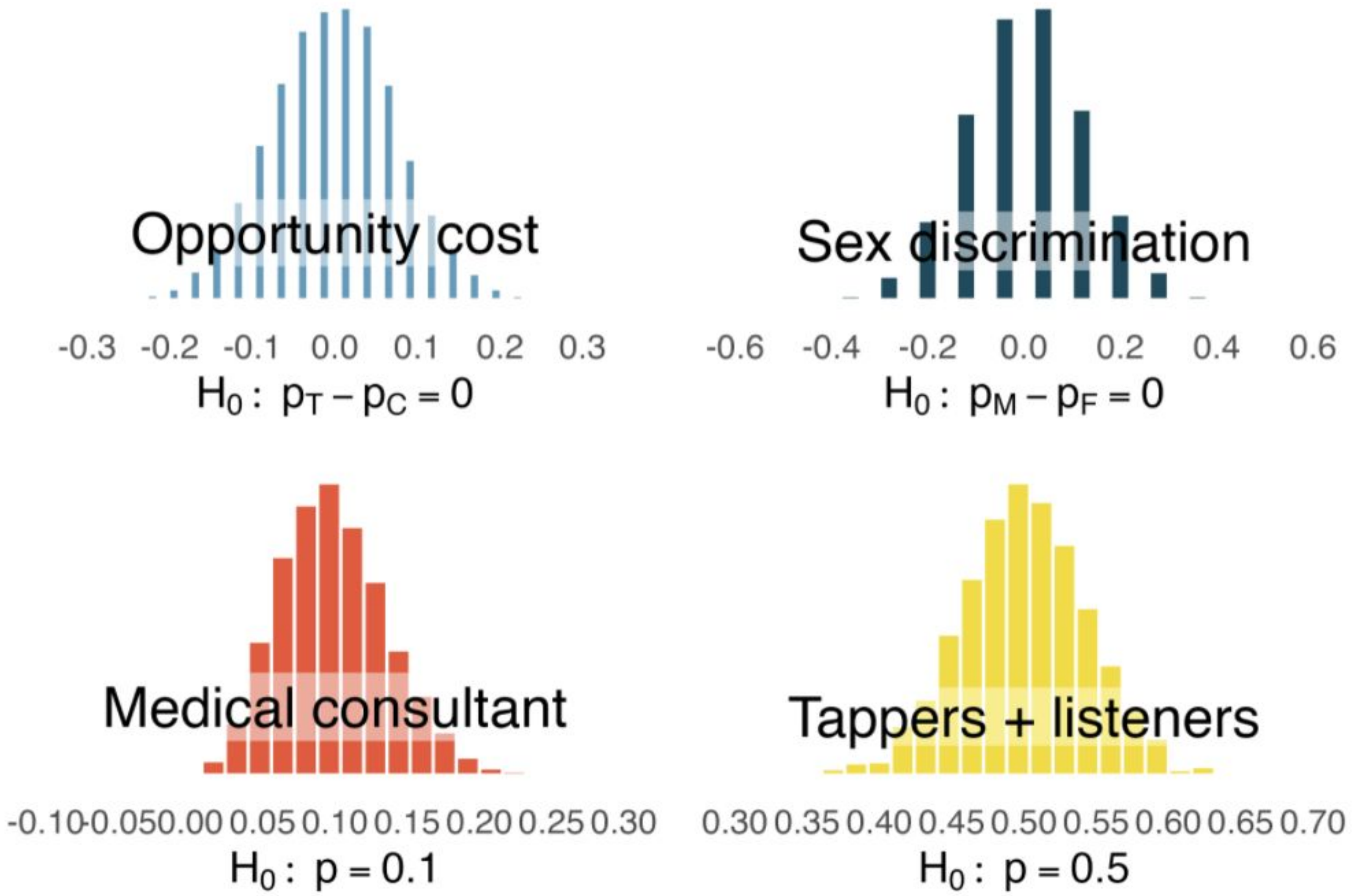
Prof. Dr. Jan Kirenz
HdM Stuttgart

Randomization and bootstrapping

With **randomization tests**, the data were permuted assuming the null hypothesis.

With **bootstrapping**, the data were resampled in order to measure the variability.

In many cases, the variability of the statistic can also be described by a **mathematical formula**



10000 simulations; Describe the shape of the distributions and note anything that you find interesting

Sampling distribution

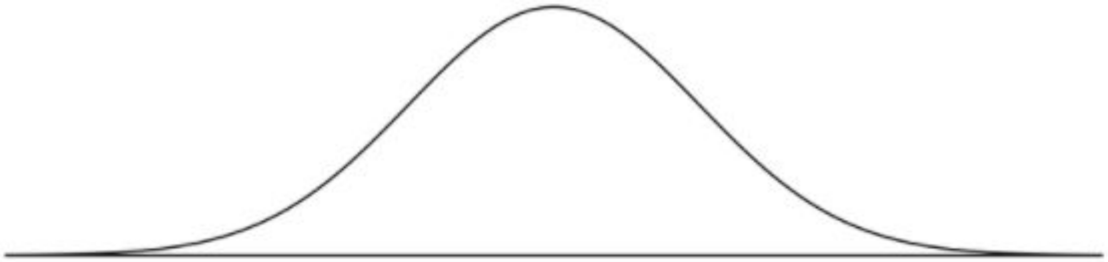
- Distribution of all possible values of a **sample statistic** from samples …
- … of a given **sample size** from a given **population**.

We can think about the sample distribution as describing as how sample statistics (e.g. the sample proportion or the sample mean) varies from one study to another.

Central Limit Theorem for proportions.

- If we look at a **proportion** (or difference in proportions)
- and the scenario satisfies certain technical **conditions**

- then the **sample proportion** (or difference in proportions) will appear to follow a
    - **bell-shaped curve** called the **normal distribution**.



Technical Conditions

Observations in the sample are **independent**.

Independence is guaranteed when we take a random sample from a population.

Independence can also be guaranteed if we randomly divide individuals into treatment and control groups.

The sample is **large enough**.

What qualifies as “large” differs from one context to the next, and we’ll provide suitable guidelines for proportions in a later lecture

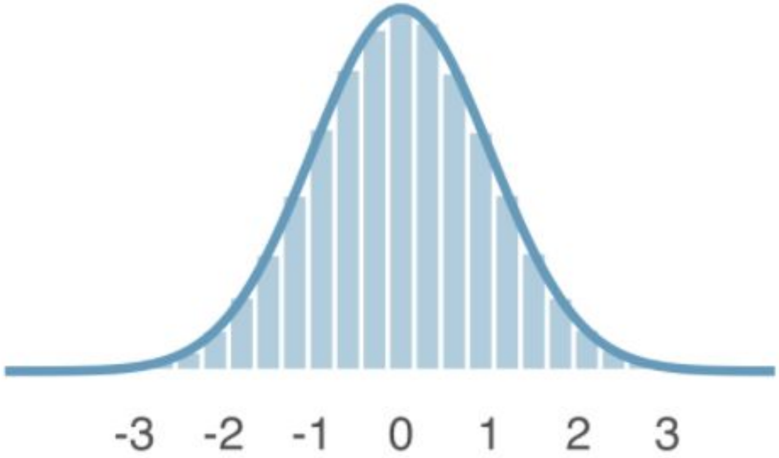
So far we’ve had no need for the **normal distribution**.

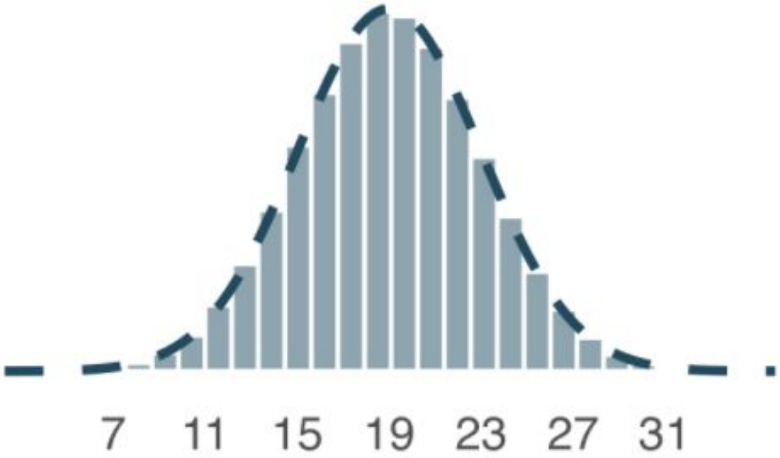
We’ve been able to answer our questions somewhat easily using **simulation techniques**.

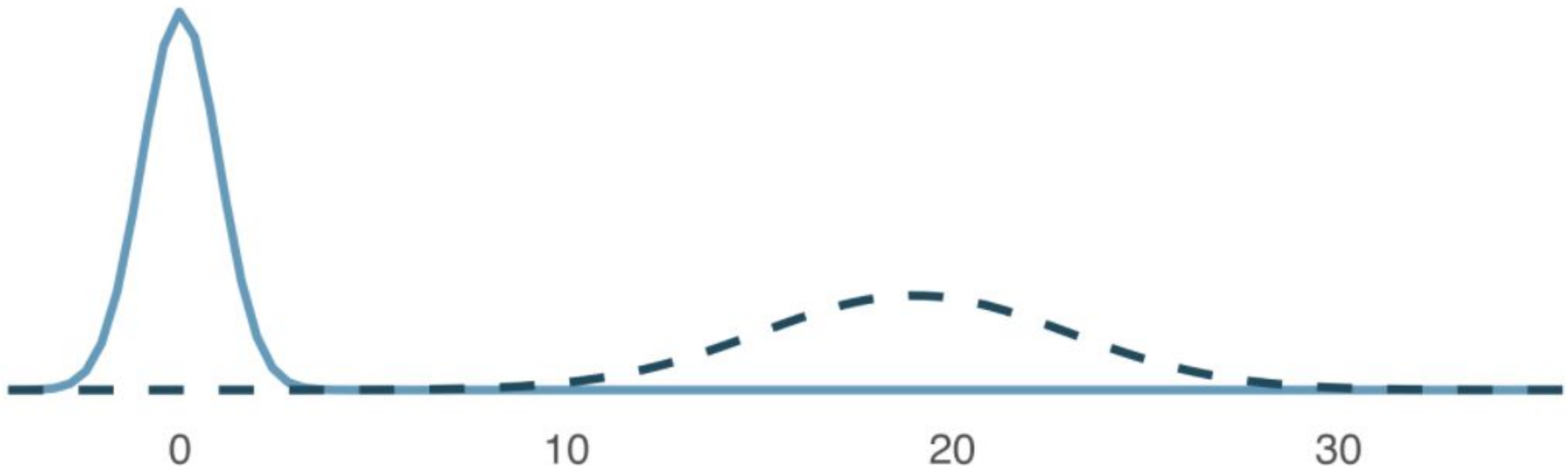
Normal distribution

- Many summary statistics and variables are nearly normal
- But none are exactly normal.

Standard normal distribution







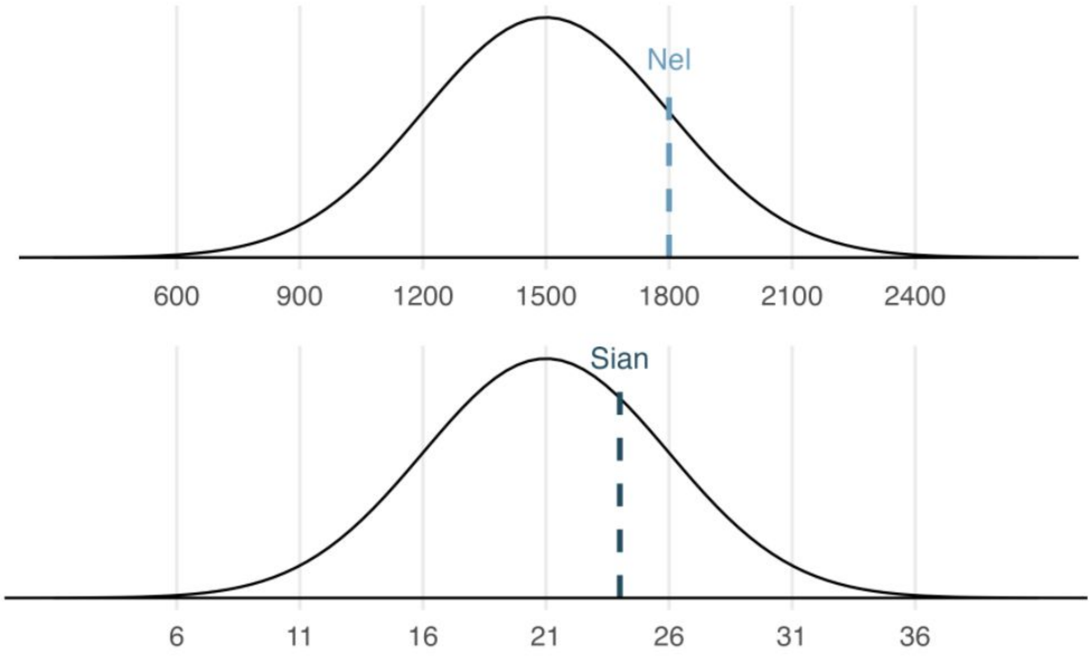
$$N(\mu = 0, \sigma = 1) \quad \text{and} \quad N(\mu = 19, \sigma = 4)$$

Practice

Write down the short-hand for a normal distribution with the following parameters.

- mean 5 and standard deviation 3
- mean -100 and standard deviation 10
- mean 2 and standard deviation 9

Who performed better?



mean SAT = 1500,
standard deviation SAT = 300;
Nel = 1800

mean ACT = 21,
standard deviation ACT = 5;
Sian = 24

Standardizing with Z scores

- Z score of an observation is the number of **standard deviations** it falls above or below the mean.

- If the observation x comes from a normal distribution centered at μ with standard deviation of σ, then the Z score will be distributed according to a normal distribution with a center of 0 and a standard deviation of 1

$$Z = \frac{x - \mu}{\sigma}$$

Z scores

- We can use Z scores to roughly identify which observations are **more unusual** than others.

- One observation x1is said to be more unusual than another observation x2 if the absolute value of its Z score is larger than the absolute value of the other observation's Z score

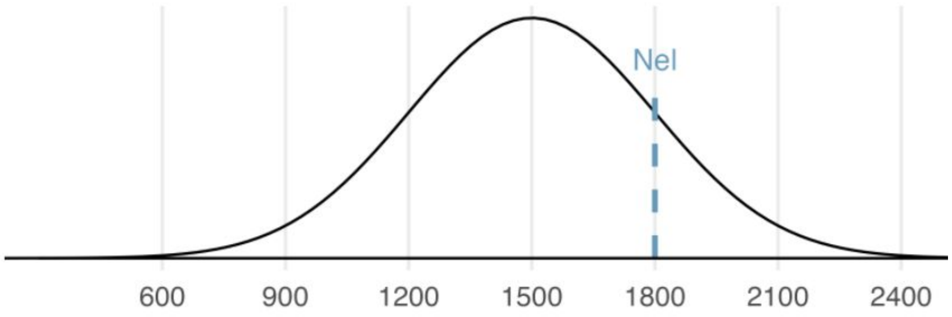
Standardizing with Z scores

$$Z = \frac{x - \mu}{\sigma}$$

Using $\mu_{SAT} = 1500$, $\sigma_{SAT} = 300$, and $x_{Nel} = 1800$, we find Nel's Z score:

$$Z_{Nel} = \frac{\boxed{}}{\boxed{}} = \frac{\boxed{}}{\boxed{}} = \boxed{}$$

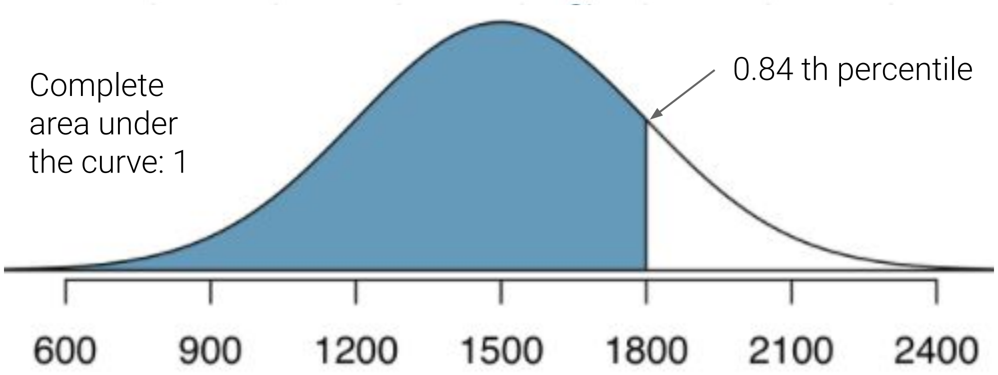
Normal probability calculations



We would like to know what **percentile** Nel falls in among all SAT test-takers.

Nel's percentile is the percentage of people who earned a lower SAT score than Nel.

Complete area under the curve: 1

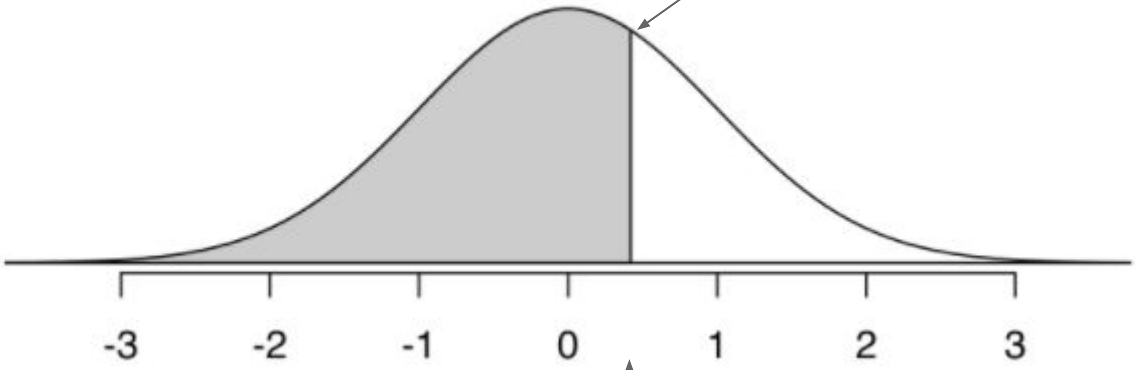


0.84 th percentile

Nel is in the 84th percentile of SAT takers.

The normal model for SAT scores, shading the area of those individuals who scored below Nel.

Step 2: obtain the percentile (with software): 0.66



Step 1: get the z-score: 0.43

$$Z = \frac{x - \mu}{\sigma}$$

Normal probability example

- Shannon is a randomly selected SAT taker, and nothing is known about Shannon's SAT aptitude.
- What is the probability that Shannon scores at least 1630 on their SATs?

- Cumulative SAT scores are approximated well by a normal model:
  - $N(\mu=1500, \sigma=300)$

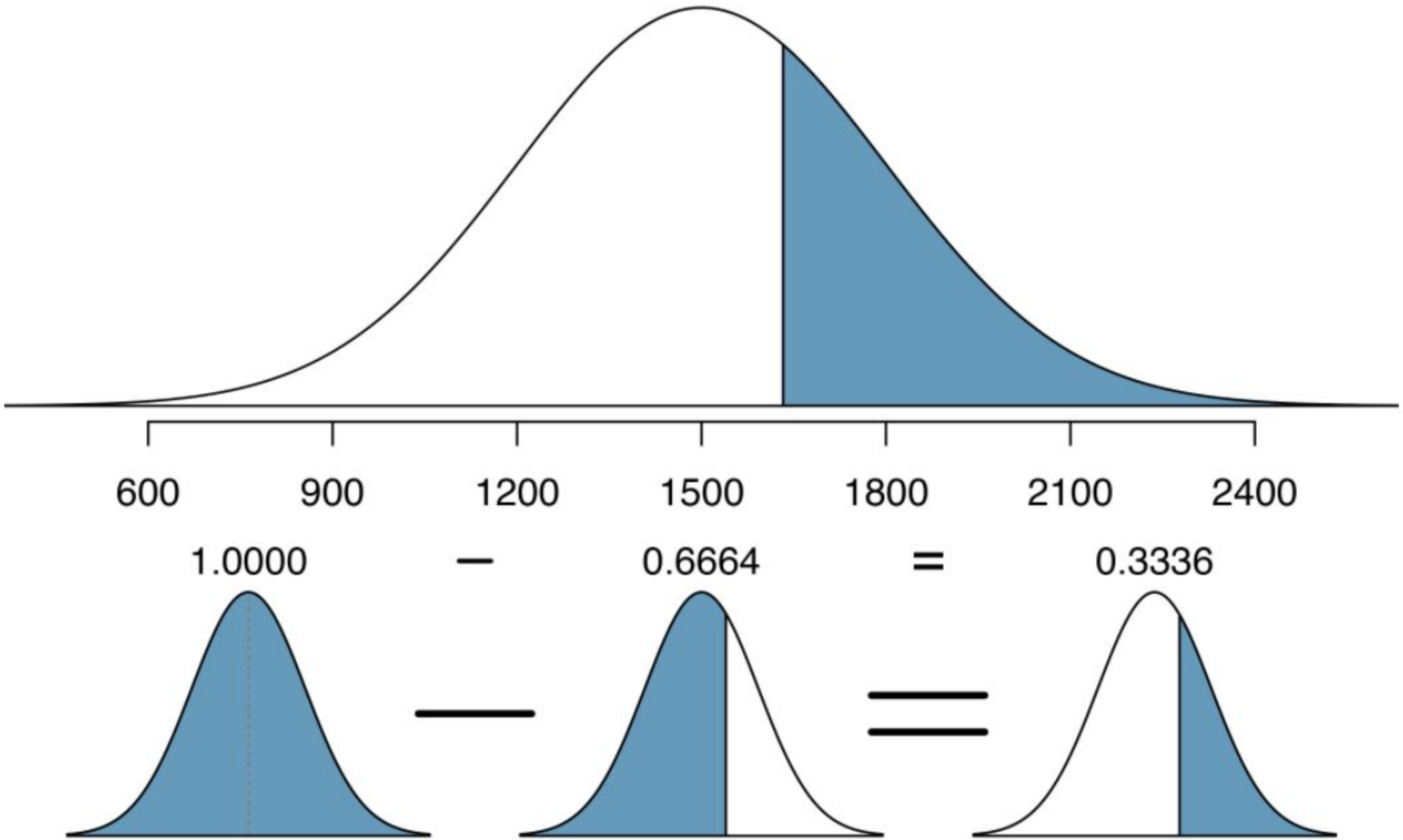
Example

$$Z = \frac{x - \mu}{\sigma} = \frac{\boxed{}}{\boxed{}} = \frac{\boxed{}}{\boxed{}} = \boxed{}$$

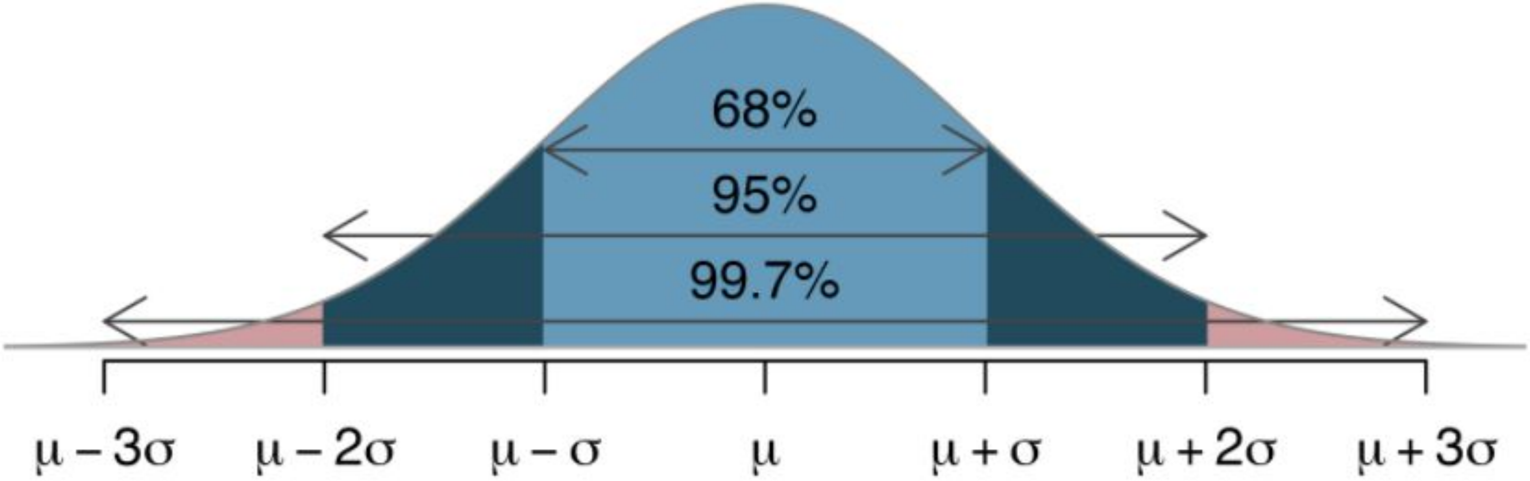
- We use software to find the percentile of Z=______
- … which yields 0.6664
- this is the percentile of those who had a Z score lower than our Z

To find the area above Z=_____,

we compute one minus the area of the lower tail, as seen below.


600
900
1200
1500
1800
2100
2400
1.0000
−
0.6664
=
0.3336

Quantifying the variability of a statistic


68%
95%
99.7%
μ − 3σ
μ − 2σ
μ − σ
μ
μ + σ
μ + 2σ
μ + 3σ

Standard error

Standard error

- Point estimates vary from **sample to sample**
- We quantify this variability **between samples** with what is called the **standard error** (SE)

- The standard error is equal to the **standard deviation** associated with the statistic (within one sample).

Margin of error

Margin of error

- Very related to the standard error is the **margin of error**.
- The margin of error describes how far away observations are from their mean.
- For example, to describe where most (i.e., 95%) observations lie, we say that the **margin of error** is approximately $2 \times SE$
- That is, **95%** of the observations are within one margin of error of the mean.

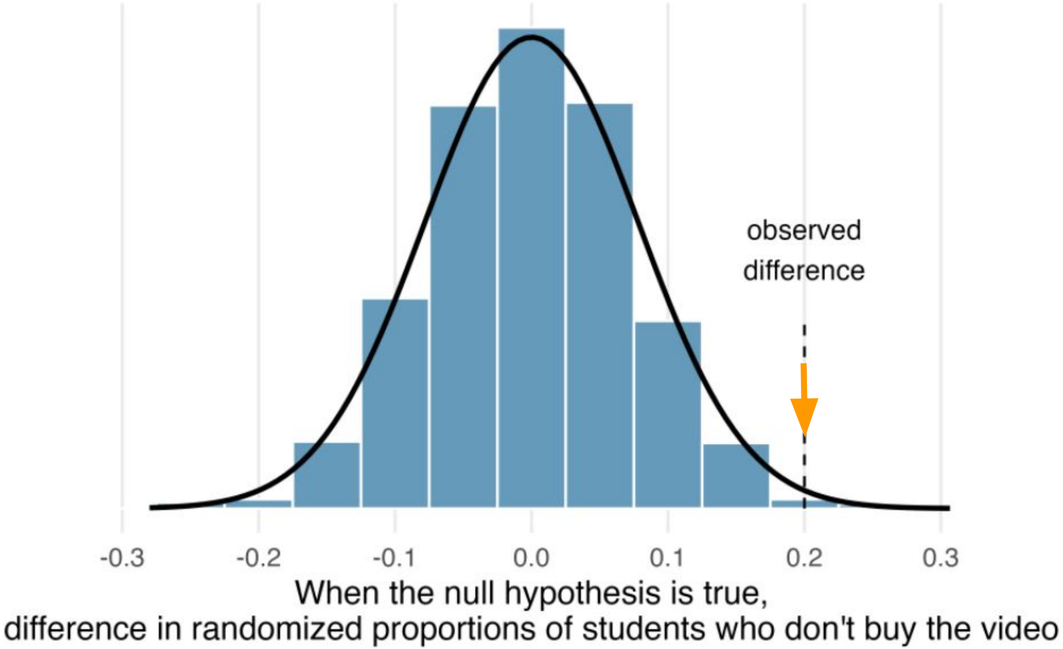
Margin of error for sample proportions.

The distance given by $z^\star \times SE$ is called the **margin of error**.

$z^\star$ is the cutoff value found on the normal distribution. The most common value of $z^\star$ is 1.96 (often approximated to be 2) indicating that the margin of error describes the variability associated with 95% of the sampled statistics.

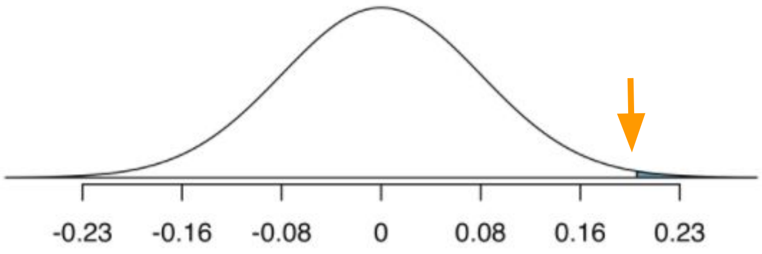
Case Study: Opportunity cost “money can be saved…”

Saving money experiment



Standard error (SE) = 0.078

Observed statistic vs. null statistics



We want to find the area of the tail beyond 0.2, representing the **p-value**.

$$Z = \frac{\text{observed difference} - 0}{SE} = \frac{0.20 - 0}{0.078} = 2.56$$

Z = 2.56 -> p-Value = 0.0052

Using this area as the p-value, we see that the p-value is less than 0.05, we conclude that the treatment did indeed impact students' spending.

Z score in a hypothesis test.

In the context of a hypothesis test, the Z score for a point estimate is

$$Z = \frac{\text{point estimate} - \text{null value}}{SE}$$

The standard error in this case is the equivalent of the standard deviation of the point estimate, and the null value comes from the claim made in the null hypothesis.

Case Study:
Medical consultant

Case study (test): Medical consultant

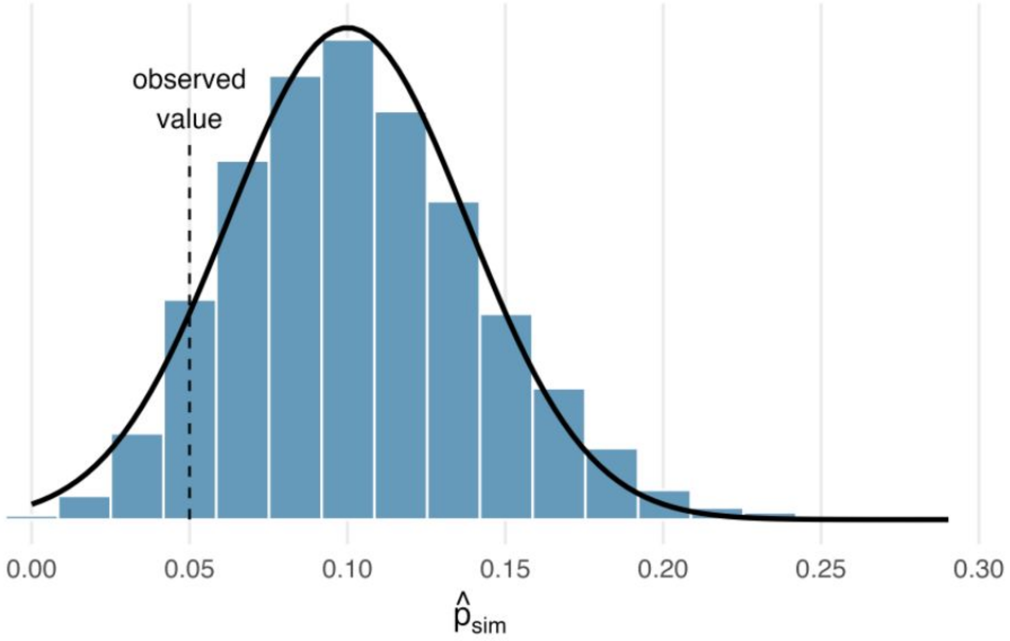
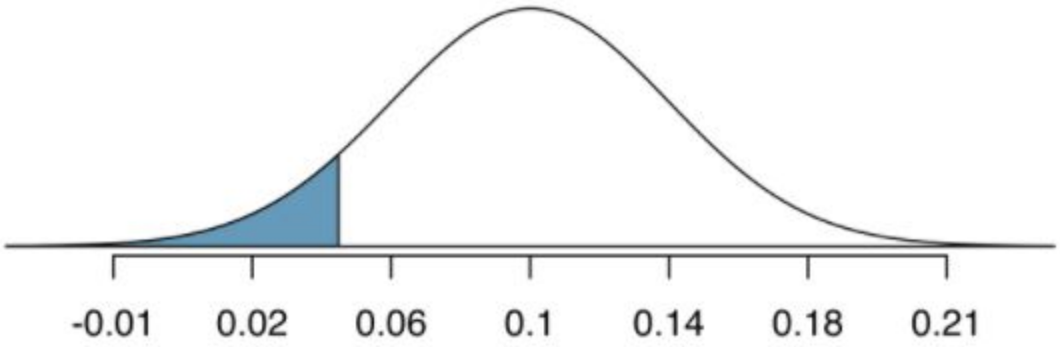


Figure 13.10: The null distribution for the sample proportion, created from 10,000 simulated studies, along with the best-fitting normal model.



$$Z = \frac{\hat{p} - p_0}{SE_{\hat{p}}} = \frac{0.048 - 0.10}{0.038} = -1.37$$

Z = -1.37 -> p-Value = 0.0853

Confidence interval

Confidence interval

- A **point estimate** is our best guess for the value of the parameter,
- It makes sense to build the **confidence interval** around that value.
- The **standard error**, which is a measure of the uncertainty associated with the point estimate, provides a guide for how large we should make the confidence interval.

- The 68-95-99.7 rule tells us that, in general, 95% of observations are within 2 standard errors of the mean.
- Here, we use the value 1.96 instead of 2 (to be slightly more precise).

Constructing a 95% confidence interval.

When the sampling distribution of a point estimate can reasonably be modeled as normal, the point estimate we observe will be within 1.96 standard errors of the true value of interest about 95% of the time. Thus, a **95% confidence interval** for such a point estimate can be constructed:

$$\text{point estimate} \pm 1.96 \times SE$$

We can be **95% confident** this interval captures the true value.

Correct confidence interval interpretation.

We are XX% confident that the population parameter is between *lower* and *upper* (where *lower* and *upper* are both numerical values).

Incorrect language might try to describe the confidence interval as capturing the population parameter with a certain probability.

This is one of the most common errors: while it might be useful to think of it as a probability, the confidence level only quantifies how plausible it is that the parameter is in the interval.

Summary

1. Frame the research question.
2. Collect data with an observational study or experiment.
3. **Model the randomness of the statistic.** In many cases, the normal distribution will be an excellent model for the randomness associated with the statistic of interest. The Central Limit Theorem tells us that if the sample size is large enough, sample averages (which can be calculated as either a proportion or a sample mean) will be approximately normally distributed when describing how the statistics change from sample to sample.
4. **Calculate the variability of the statistic.** Using formulas, come up with the standard deviation (or more typically, an estimate of the standard deviation called the standard error) of the statistic. The SE of the statistic will give information on how far the observed statistic is from the null hypothesized value (if performing a hypothesis test) or from the unknown population parameter (if creating a confidence interval).
5. **Use the normal distribution to quantify the variability.** The normal distribution will provide a probability which measures how likely it is for your observed and hypothesized (or observed and unknown) parameter to differ by the amount measured. The unusualness (or not) of the discrepancy will form the conclusion to the research question.
6. **Form a conclusion**. Using the p-value or the confidence interval from the analysis, report on the research question of interest. Also, be sure to write the conclusion in plain language so casual readers can understand the results.

Summary

Question	Answer
What does it do?	Uses theory (primarily the Central Limit Theorem) to describe the hypothetical variability resulting from either repeated randomized experiments or random samples
What is the random process described?	Randomized experiment or random sampling
What other random processes can be approximated?	Randomized experiment or random sampling
What is it best for?	Quick analyses through, for example, calculating a Z score.
What physical object represents the simulation process?	Not applicable

Terms you should know

95% confidence interval	normal distribution	percentile
95% confident	normal model	sampling distribution
Central Limit Theorem	normal probability table	standard error
margin of error	null distribution	standard normal distribution
normal curve	parameter	Z score