Review of Frequentist Inference

Setup for Inference

- There is an unknown parameter θ
- There is a ***random variable*** X defined on some sample space
- The sample data $x = (x_1, x_2, \ldots, x_n)$ are observations of the random variable,
  - x is a ***realization*** of the random variable X
- Example:
  - X is a random variable height
  - x contains the values of height in the sample data (x_1 is my height, x_2 is your height, etc.)
  - The ***random variable height*** is distributed following some parameter $X \sim p(X \mid \theta)$
  - The sample data values of height, x, is not distributed as anything (my height is not distributed, neither is yours, etc.)

Sampling Distribution of the Sample Mean Estimator

- Let X be a random variable, $p(X) = N(\mu, \sigma^2)$ where σ^2 is known
- Wish to make an inference about μ
- Let $\overline{X}$ be an ***estimator*** for μ
- Let $\bar{x}$ be the ***estimate*** from a particular sample x

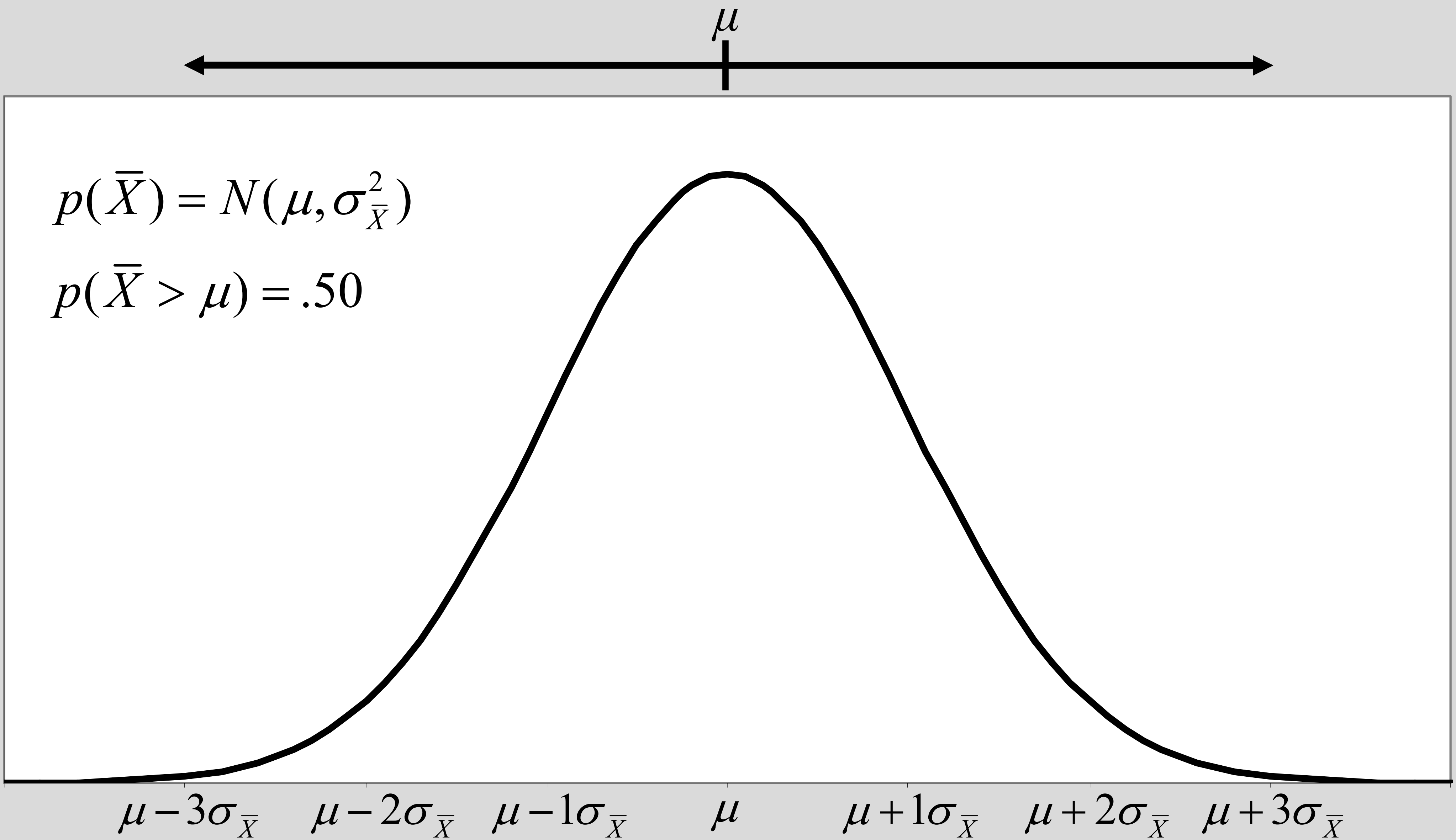
- Let $\theta = \mu$

$$\hat{\theta}(X) = \overline{X}$$

$$\hat{\theta}(x) = \bar{x}$$

- The sampling distribution of $\hat{\theta}(X) = \overline{X}$ is $p(\overline{X}) = N(\mu, \sigma_{\overline{X}}^2)$ where $\sigma_{\overline{X}} = \dfrac{\sigma_X}{\sqrt{n}}$

Sampling Distribution of the Sample Mean Estimator



Sampling Distribution for an Estimator, Not a Parameter

$\mu = 100$

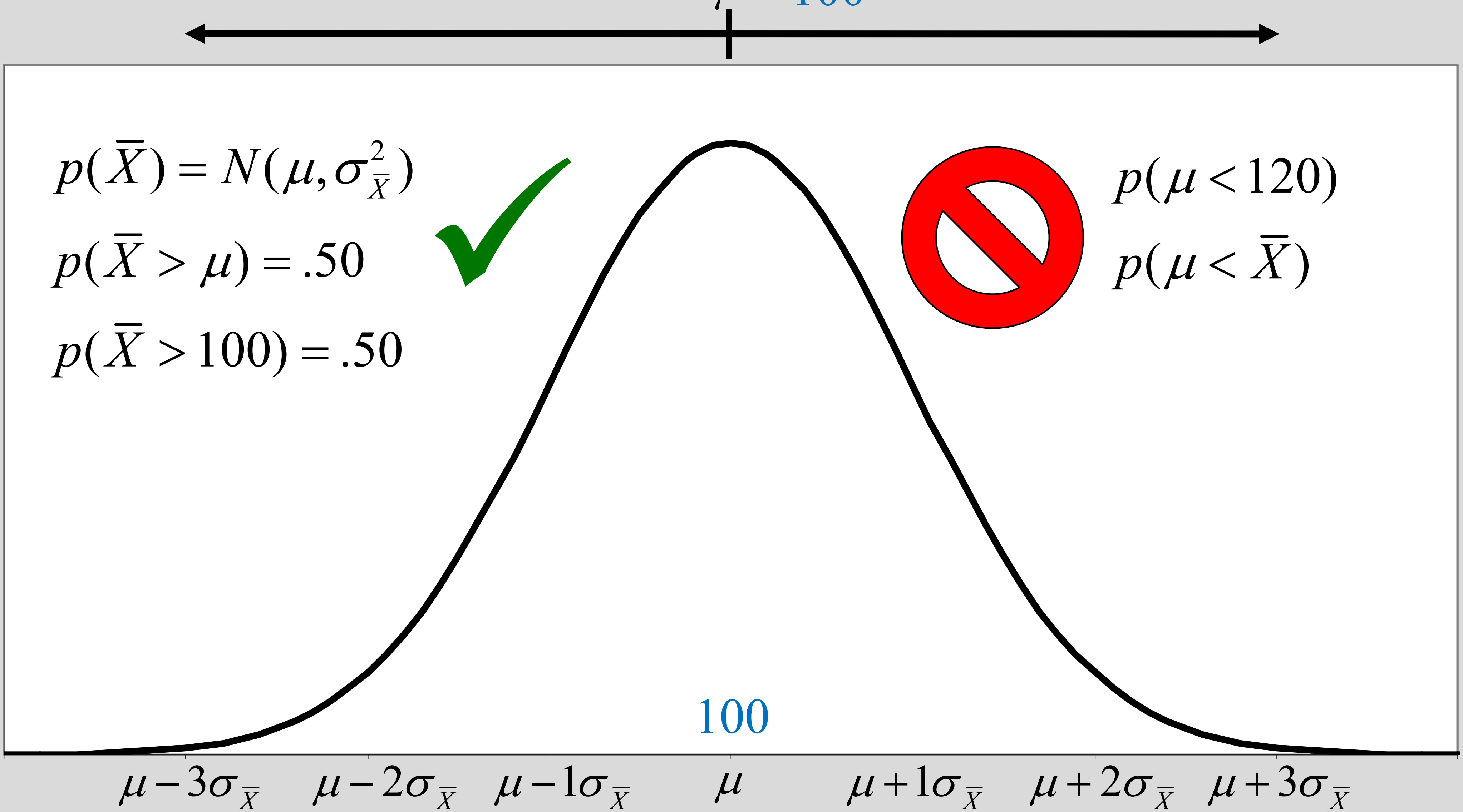
$$p(\bar{X}) = N(\mu, \sigma^2_{\bar{X}})$$

$$p(\bar{X} > \mu) = .50$$

$$p(\bar{X} > 100) = .50$$

$$p(\mu < 120)$$

$$p(\mu < \bar{X})$$



Estimator vs. ***Estimate***

- Statements like

$$p(\bar{X}) = N(\mu, \sigma_{\bar{X}}^2)$$

$$p(\bar{X} > \mu) = .50$$

$$p(\bar{X} > 100) = .50$$

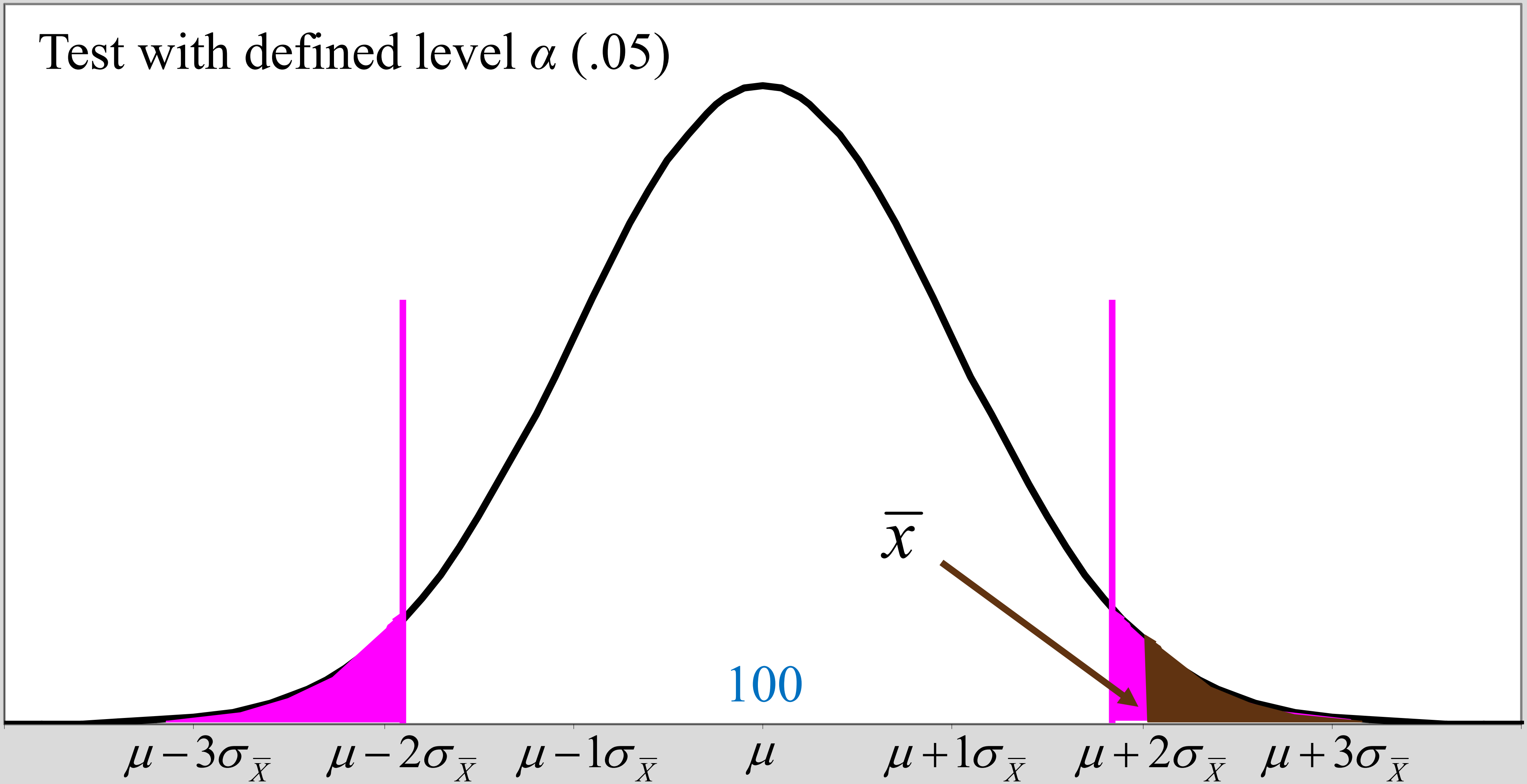
$$p(\mu - 1.96\sigma_{\bar{X}} < \bar{X} < \mu + 1.96\sigma_{\bar{X}})$$

  refer to properties of the ***estimator*** $\bar{X}$ and are ***prospective*** in the sense that they do not depend on the actual observed values x
- But for any ***estimate*** $\bar{x}$ from a particular sample, we have no way of knowing how close this estimate is to μ
- We cannot say if $\bar{x}$ is within 1.96 standard errors of μ, even with probability .95

Hypothesis Testing

H_0: $\mu = 100$

H_A: $\mu \neq 100$



Hypothesis Testing

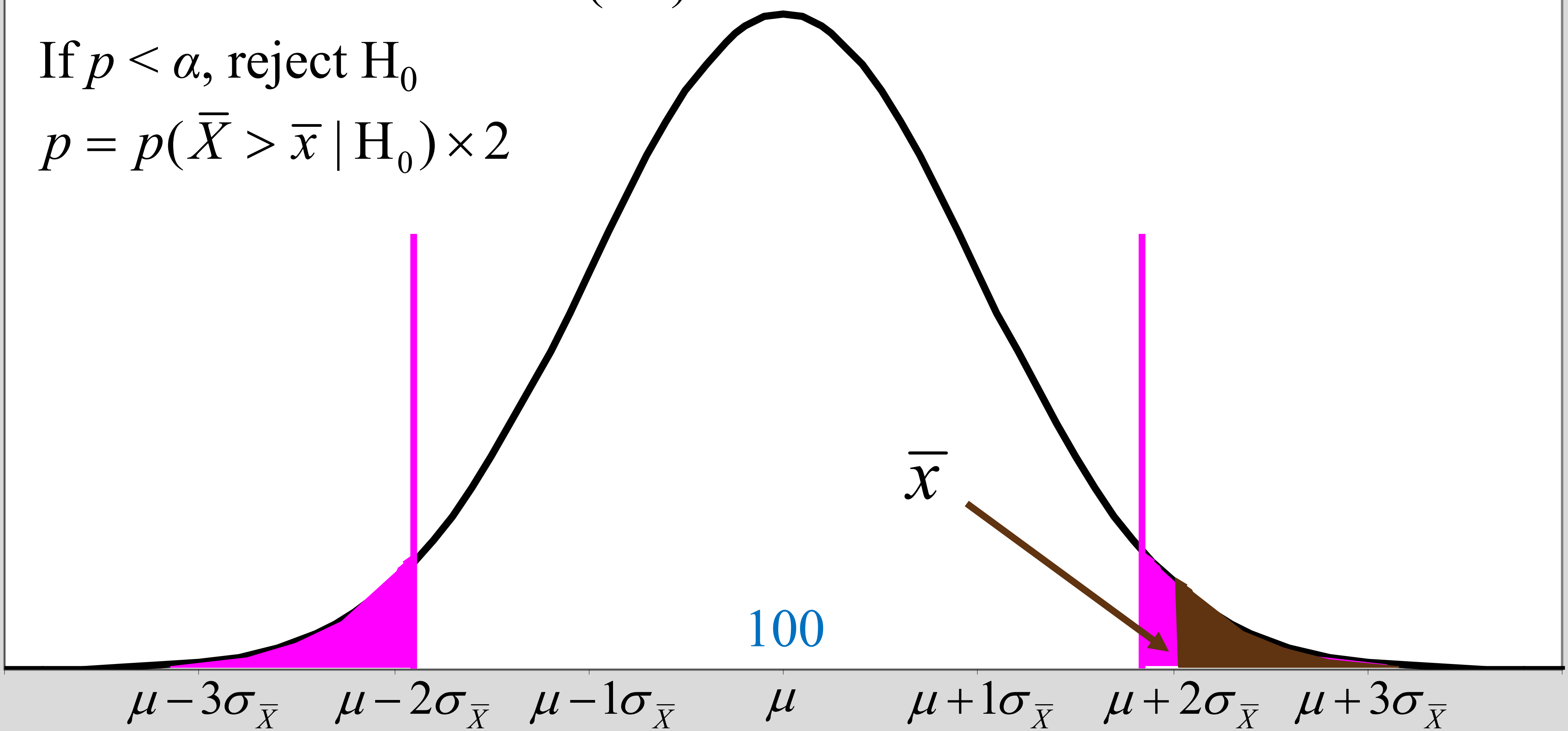
H_0: $\mu = 100$

H_A: $\mu \neq 100$

Test with defined level α (.05)

If $p < \alpha$, reject H_0

$$p = p(\bar{X} > \bar{x} \mid H_0) \times 2$$



Hypothesis Testing

H_0: $\mu = 100$

H_A: $\mu \neq 100$

Test with defined level α (.05)

If $p < \alpha$, reject H_0

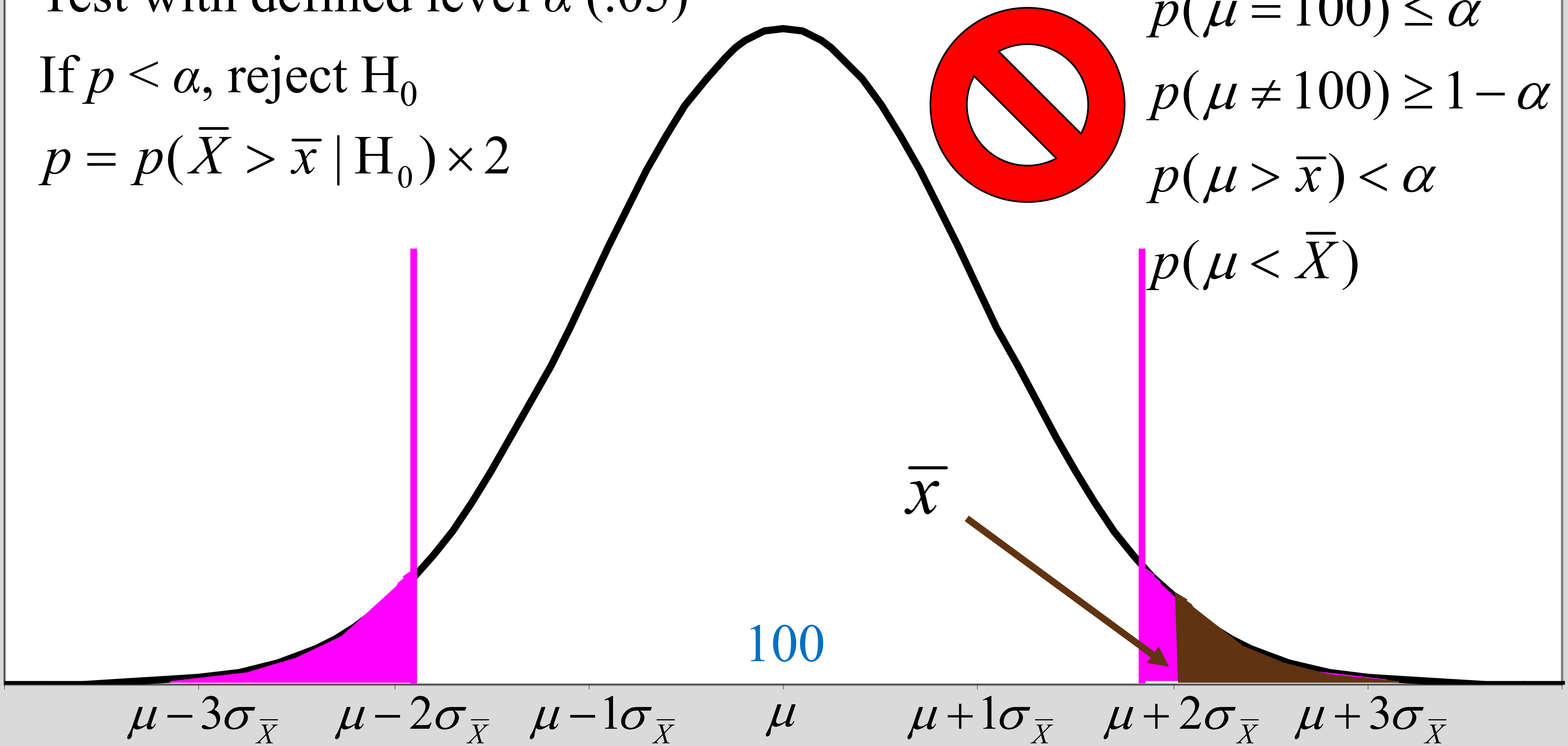
$p = p(\bar{X} > \bar{x} \mid H_0) \times 2$

$p(\mu = 100) \leq \alpha$

$p(\mu \neq 100) \geq 1 - \alpha$

$p(\mu > \bar{x}) < \alpha$

$p(\mu < \bar{X})$



Probability in Hypothesis Testing

What's wrong with NHST? Well, among many other things, it does not tell us what we want to know, and we so much want to know what we want to know that, out of desperation, we nevertheless believe that it does! What we want to know is "Given these data, what is the probability that H_0 is true?" But as most of us know, what it tells us is "Given that H_0 is true, what is the probability of these (or more extreme) data?" These are not the same, as has been pointed out many times over the years…

-- Cohen (1994, p. 997)

Probability in Hypothesis Testing

What's wrong with NHST? Well, among many other things, it does not tell us what we want to know, and we so much want to know what we want to know that, out of desperation, we nevertheless believe that it does! What we want to know is "Given these data, what is the probability that H_0 is true?" But as most of us know, what it tells us is "Given that H_0 is true, what is the probability of these (or more extreme) data?" These are not the same, as has been pointed out many times over the years…

-- Cohen (1994, p. 997)

Adopting an explicitly Bayesian approach would resolve a recurring source of confusion for these researchers, letting them say what they mean and mean what they say.

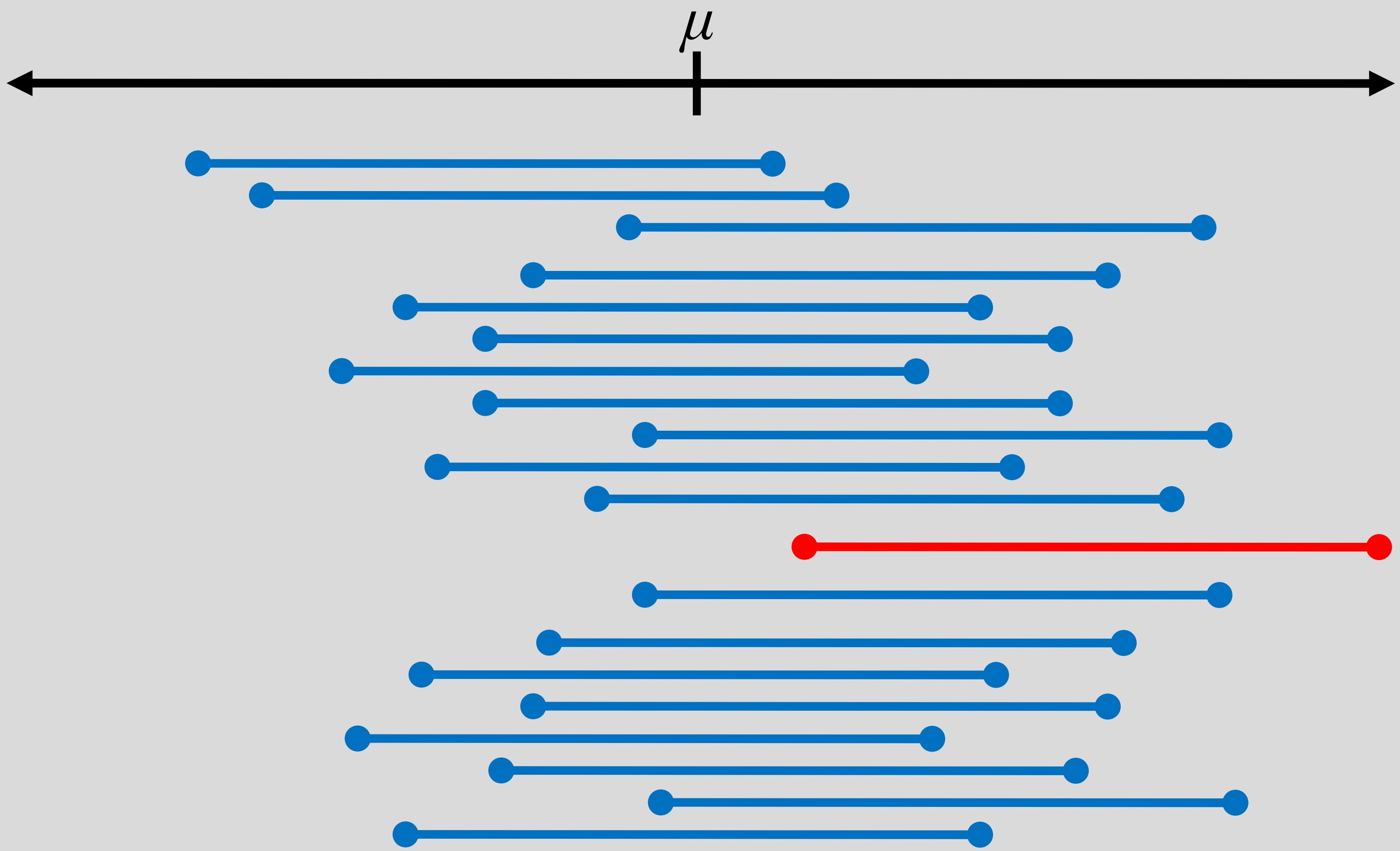
-- Jackman (2009, p. xxviii)

Confidence Intervals

A __% confidence interval for a parameter θ is an interval (L,U) generated by a procedure that in repeated sampling has a __% probability of containing the true value of θ, for all possible values of θ

- Define an interval for θ via the estimator (L, U) $= \underline{\overline{\theta}}(X)$, based on the desired level of confidence $1 - \alpha$
- Example: $p(X) = N(\mu, \sigma^2)$ where σ^2 is known
- The ***estimator*** is $\underline{\overline{\theta}}(X) = \overline{X} \pm Z_\alpha \sigma_{\overline{X}}$
- The ***estimate*** is $\underline{\overline{\theta}}(x) = \overline{x} \pm Z_\alpha \sigma_{\overline{X}}$
- The estimate may differ for different samples

Confidence Intervals for a Population Mean



Confidence Intervals

- Probability interpretation is for the ***estimator***

$$p(\theta \in \underline{\overline{\theta}}(X)) = 1 - \alpha$$

- Over repeated samples, the proportion of intervals that contain θ will be $1 - \alpha$
  - The proportion of times such intervals will contain θ is $1 - \alpha$
- Do not know ***whether a particular*** interval contains θ

Can we say that a particular interval contains θ with any probability?

From the Literature on Confidence Intervals

For instance, with respect to 95% confidence intervals, Masson and Loftus (2003) state that "in the absence of any other information, there is a 95% probability that the obtained confidence interval includes the population mean." Cumming (2014) writes that "[w]e can be 95% confident that our interval includes [the parameter] and can think of the lower and upper limits as likely lower and upper bounds for [the parameter].

-- Morey et al. (2016, p. 104)

From the Literature on Confidence Intervals

For instance, with respect to 95% confidence intervals, Masson and Loftus (2003) state that "in the absence of any other information, there is a 95% probability that the obtained confidence interval includes the population mean." Cumming (2014) writes that "[w]e can be 95% confident that our interval includes [the parameter] and can think of the lower and upper limits as likely lower and upper bounds for [the parameter].

These interpretations of confidence intervals are not correct. We call the mistake these authors have made the "Fundamental Confidence Fallacy"…

-- Morey et al. (2016, p. 104)

Fundamental Confidence Fallacy
Morey et al. (2016)

If the probability that a random interval contains the true value is __%, then the plausibility or probability that a particular observed interval contains the true value is also __%; or, alternatively, we can have __% confidence that the observed interval contains the true value.

Confuses

- What is known ***before*** observing the data
- What is known ***after*** observing the data

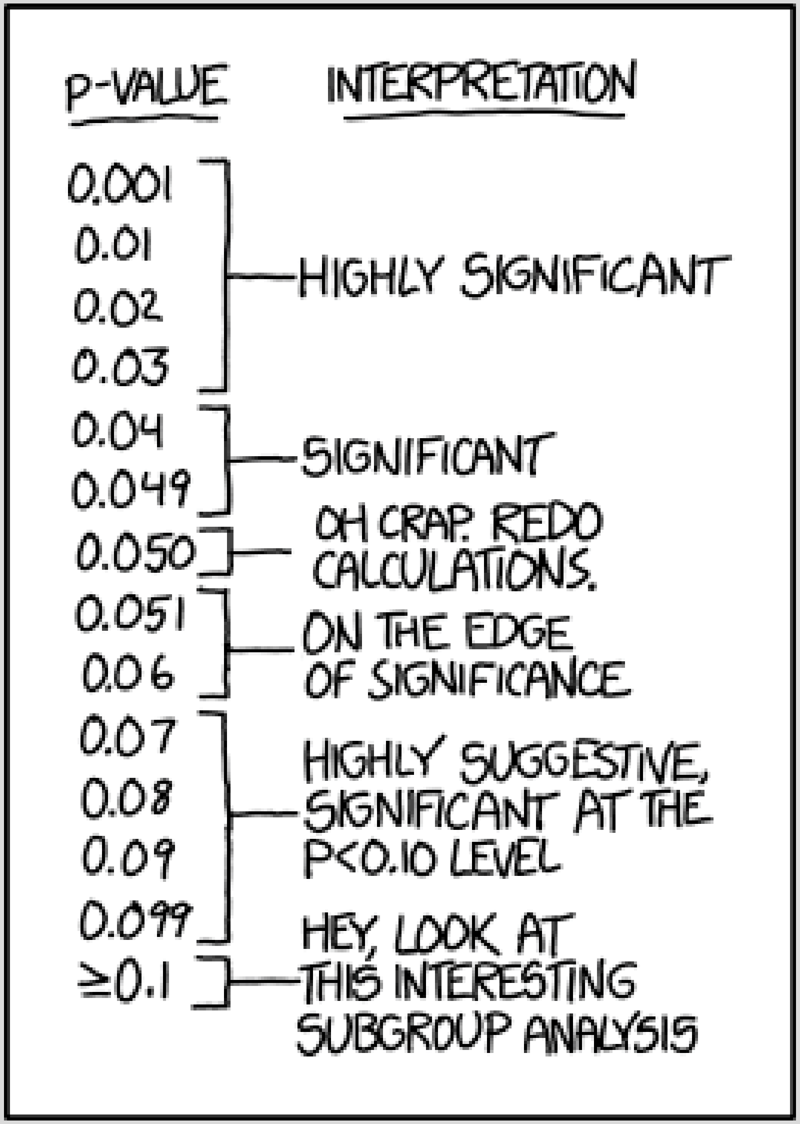
Frequentist CI theory says nothing at all about the probability that a particular, observed confidence interval contains the true value; it is either 0 (if the interval does not contain the parameter) or 1 (if the interval does contain the true value). --Morey et al. (2016, p. 105)

Summary of Probability in Frequentist Inference

- Grounded in frequentist conception of probability
  - Data vary upon repeated sampling, parameters don't
- In inference, probabilities formulated regarding estimators (point or interval) or stats/estimates upon repeated sampling
- Confidence interval is ***not*** a probability for a parameter
  - Probability for an interval, upon repeated sampling
- p-value is ***not*** a probability for a hypothesis
  - Small p-value evidence against a hypothesis…
  - …large p-value not evidence for a hypothesis
  - Magnitude of p-value not a good measure of magnitude of evidence

Summary

- Frequentist probability &inference
- Estimators vs. estimates
- Sampling distributions
- Hypothesis tests
- Confidence intervals
- Probability statements



https://xkcd.com/1478/