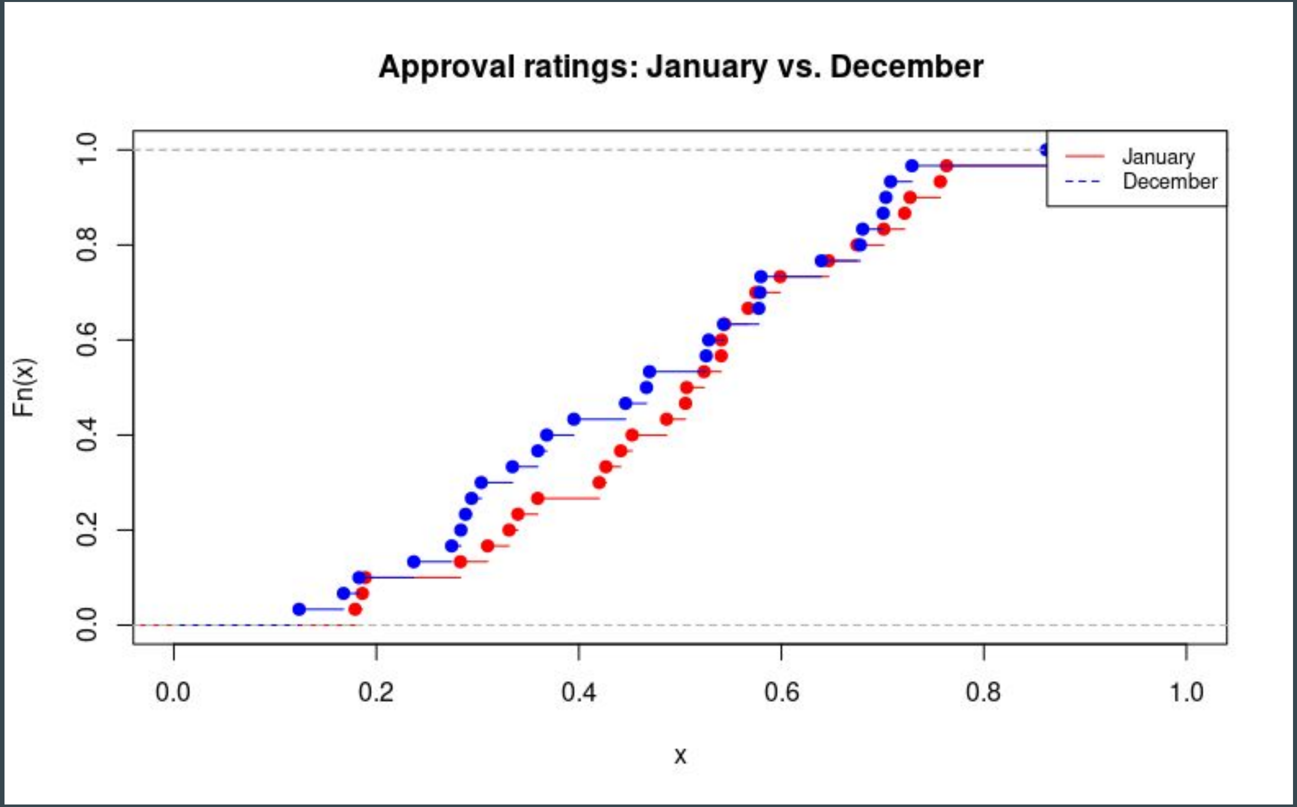
Foundations of Statistical Inference

•••

18 February 2019

PLSC 309

Review: signal or noise?



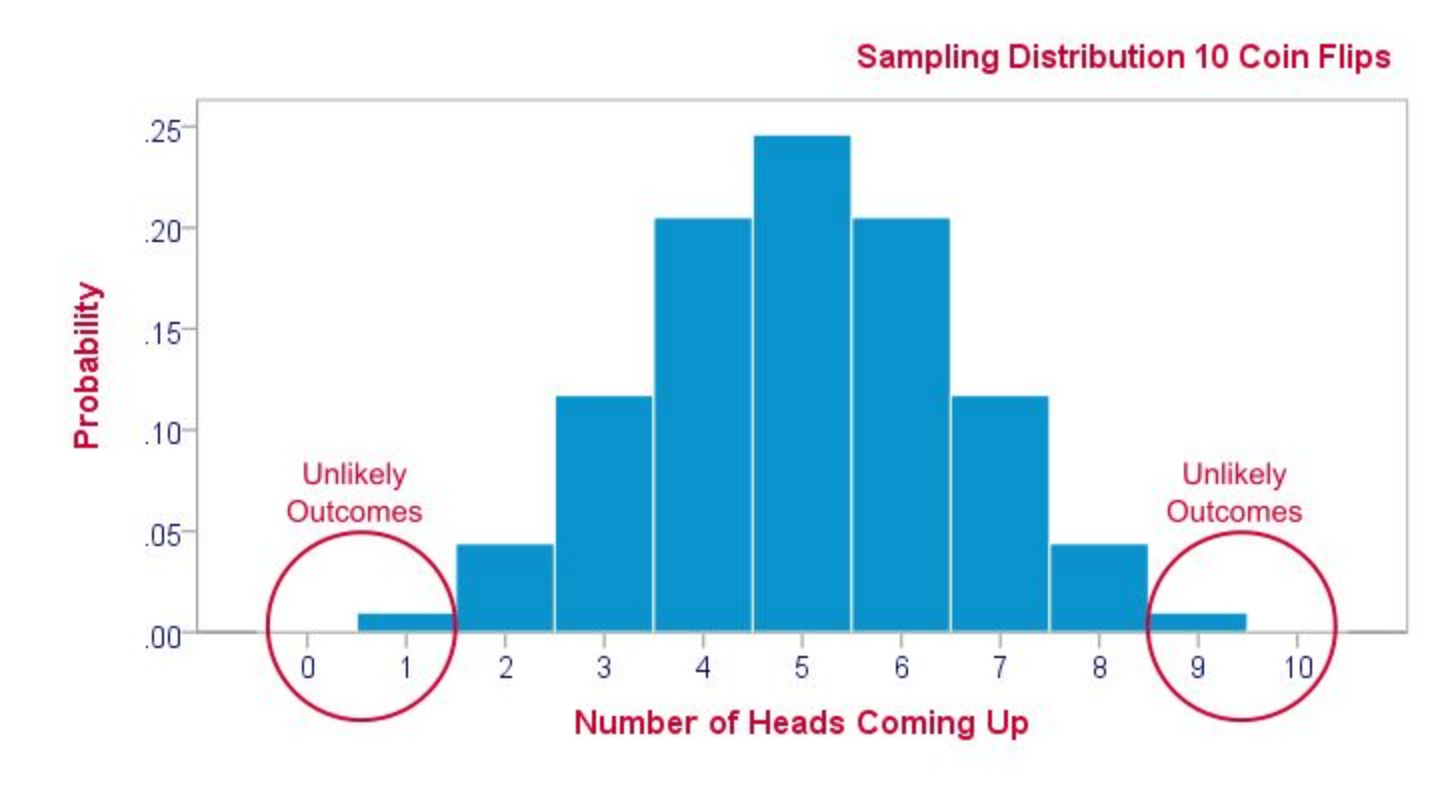
Review: signal or noise?

After opening a new school, a city wants to determine if the new school is performing well. The average school in the district has a graduation rate of 74%, with a standard deviation of 8%. The new school's graduation rate is 82%.

Review: signal or noise?

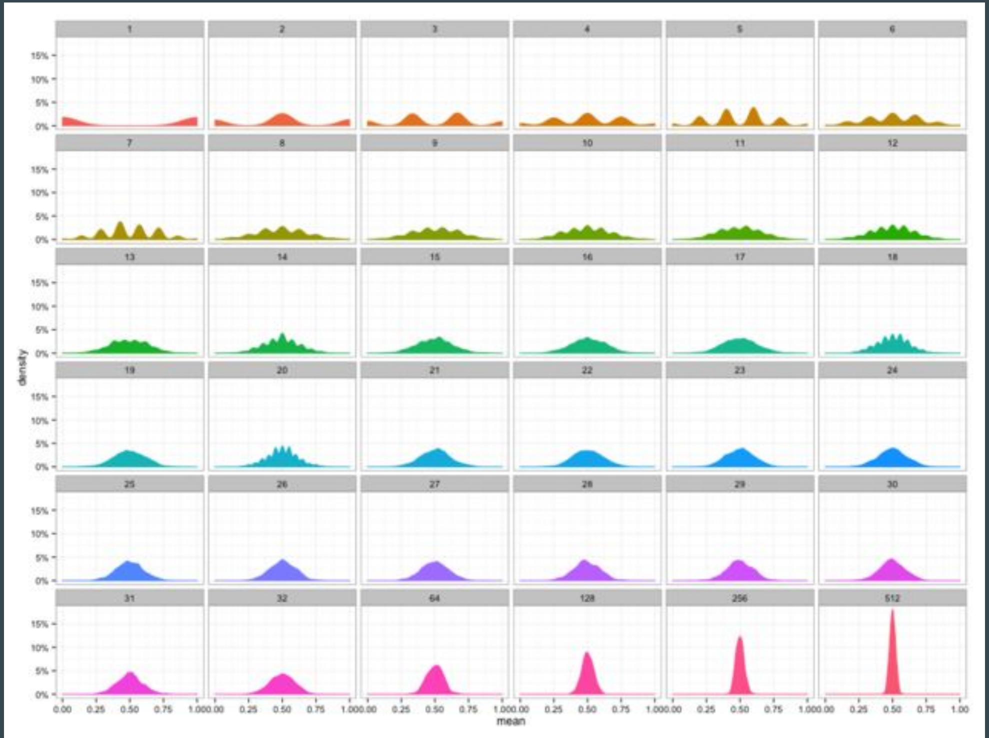
After opening a new school, a city wants to determine if the new school is performing well. The average school in the district has a graduation rate of 74%, with a standard deviation of 8%. The new school's graduation rate is 82%. Is this increase due to random chance, or has the new school done something different than the rest of the other schools to improve performance?

Review: sampling distributions



Review: CLT and normal approximation

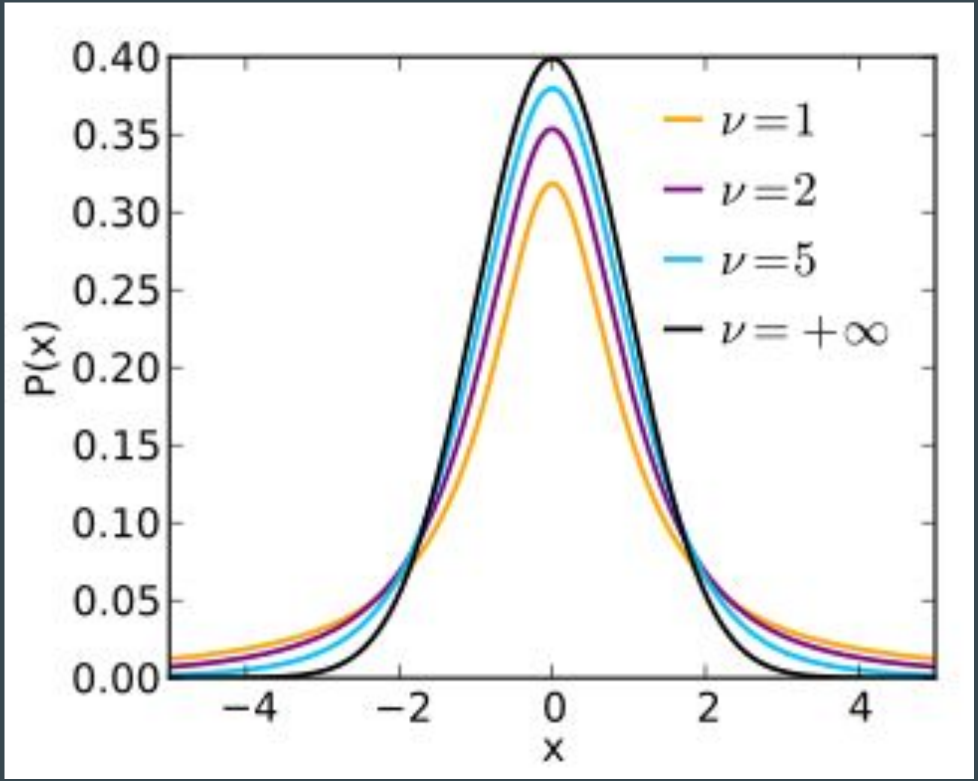
For all i.i.d. Variables, the means of repeated samples of sufficient size will be normally distributed.



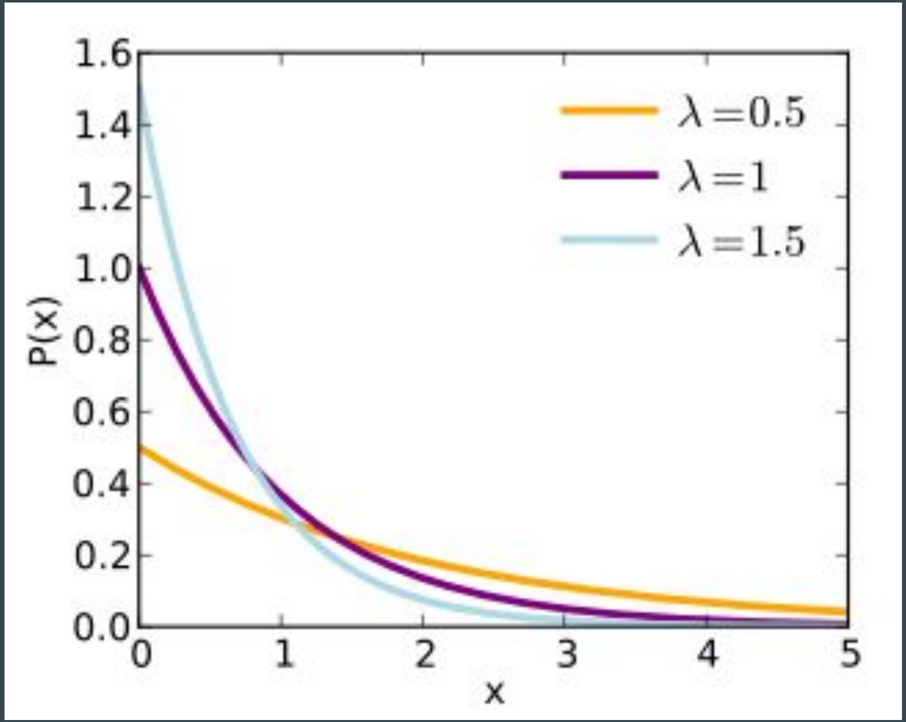
All distributions have parameters

- Binomial: p (probability)
- Poisson: λ (average count)
- Student's T: ν (normality parameter)
- Normal: μ, σ (mean, standard deviation)

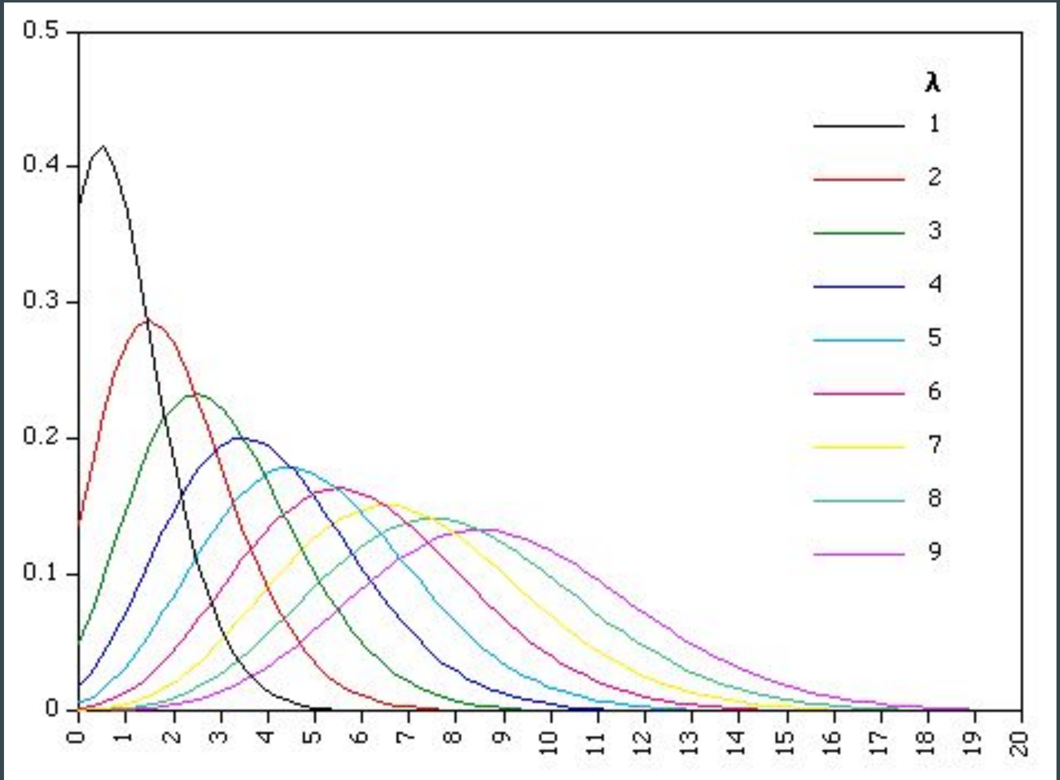
Parameters control the shape and location of a distribution



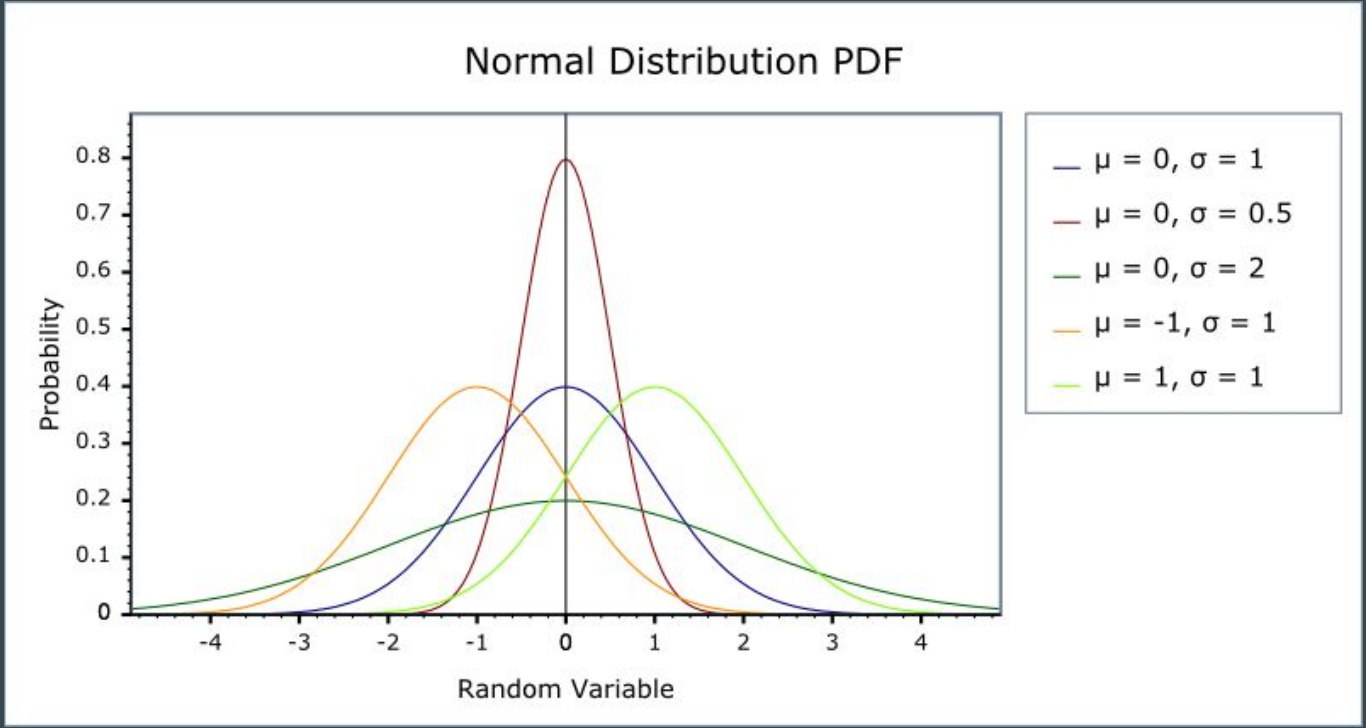
Parameters control the shape and location of a distribution



Parameters control the shape and location of a distribution



Parameters control the shape and location of a distribution



Parameters are probabilistic

- They are characteristics of infinite processes
- All observations in the universe
- We estimate the value of a parameter with a *point estimate*
- Population = parameter
- Sample = point estimate

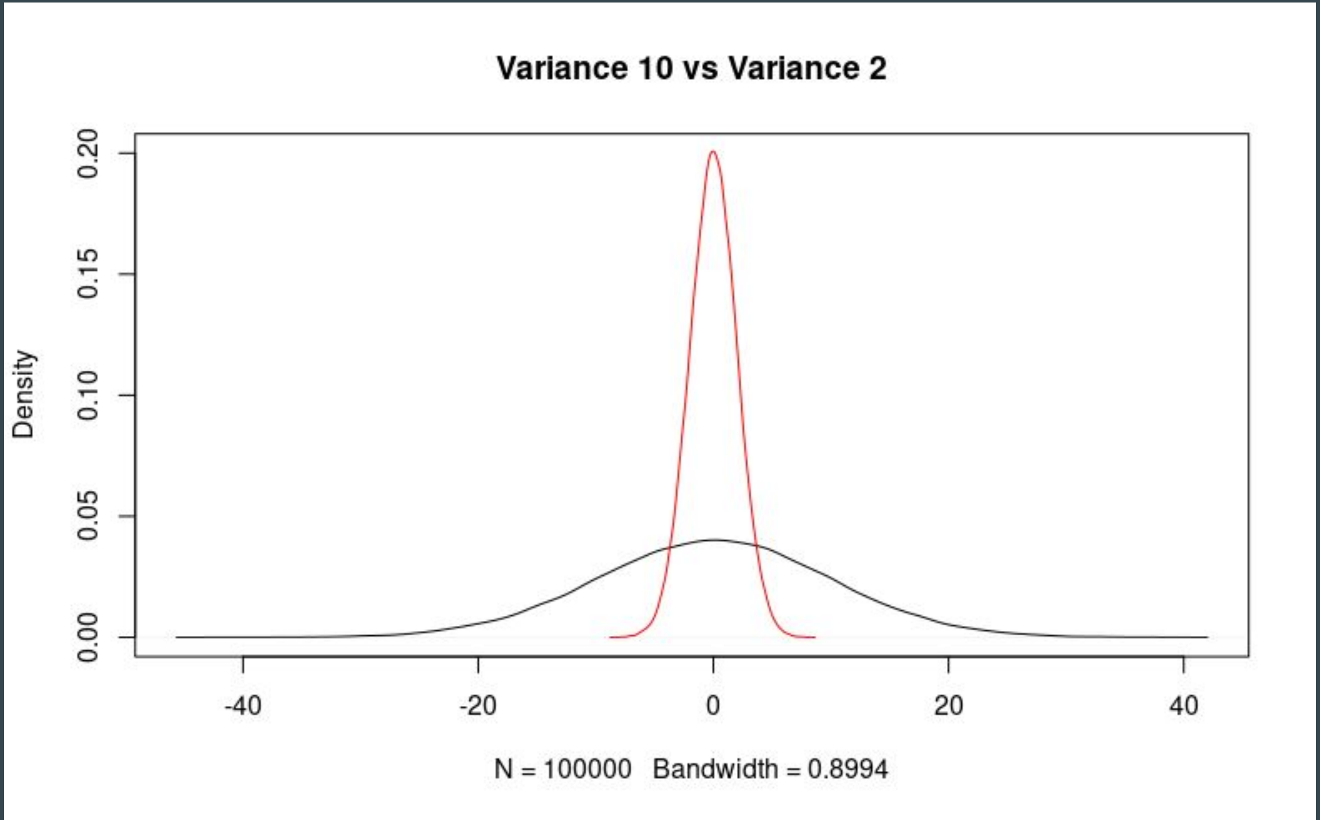
Some point estimates

- Binomial: p (probability)
  - $P = k / n$
- Poisson: λ (average count)
  - λ = mean(X)
- Normal: μ, σ (mean, standard deviation)
  - μ: mean(x)
  - σ: sqrt(var(x))

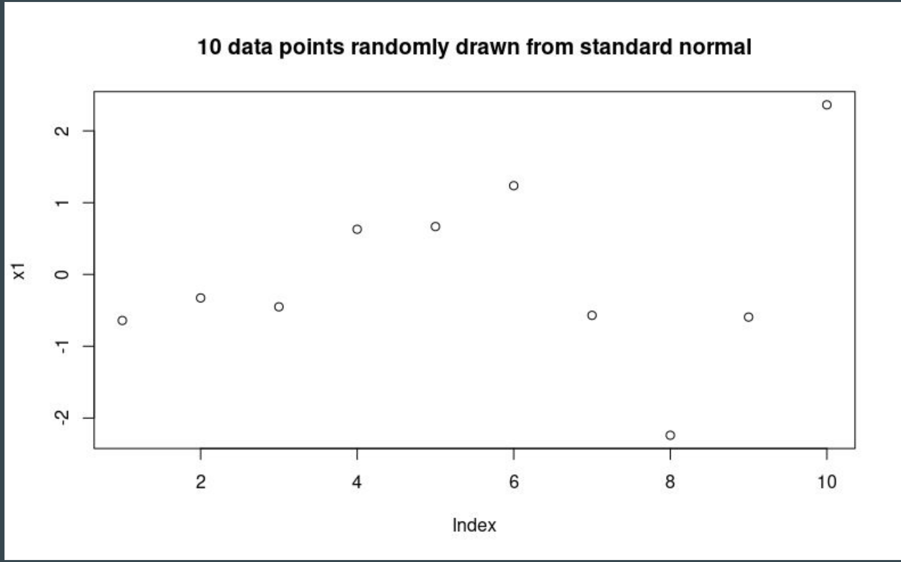
Point estimate examples

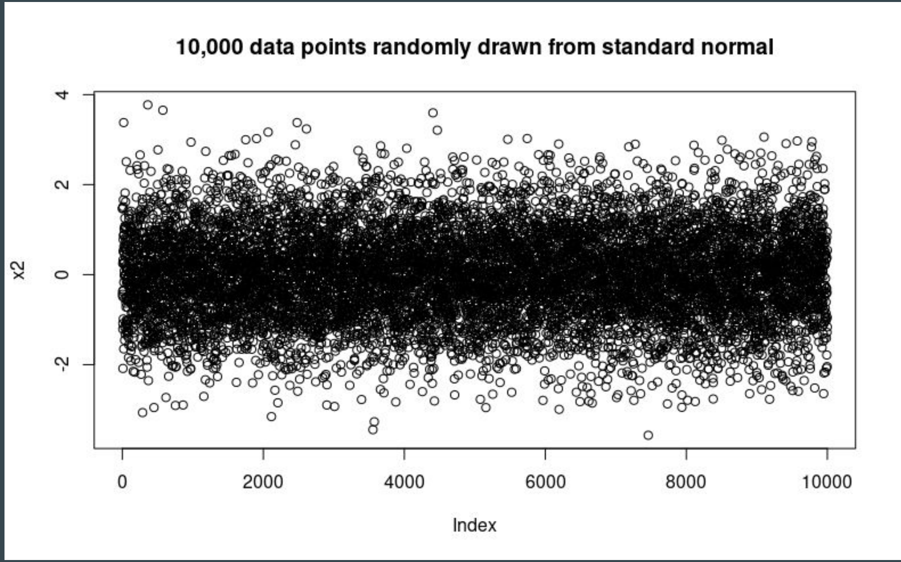
- Survey of 1000 likely voters finds 41% approval for President Trump
    - Point estimate: mean(approval) for sample
    - Parameter: average support amongst entire population
- The variance of test scores in a particular school is 8.72
    - Point estimate: variance for that school
    - Parameter: overall variance of all schools in the district

Uncertainty in point estimates



Uncertainty in point estimates





Uncertainty in point estimates

Uncertainty comes from two places:

- Variance
- Sample size

Standard Error

- Standard error captures the *uncertainty* of point estimates

$$SE = \frac{\sigma}{\sqrt{n}}$$

Standard Error

A sample of 1000 likely voters finds 62% support legalizing marijuana, with a variance of 4%.

SE = sqrt(4%) / sqrt(1000) = 6.3%

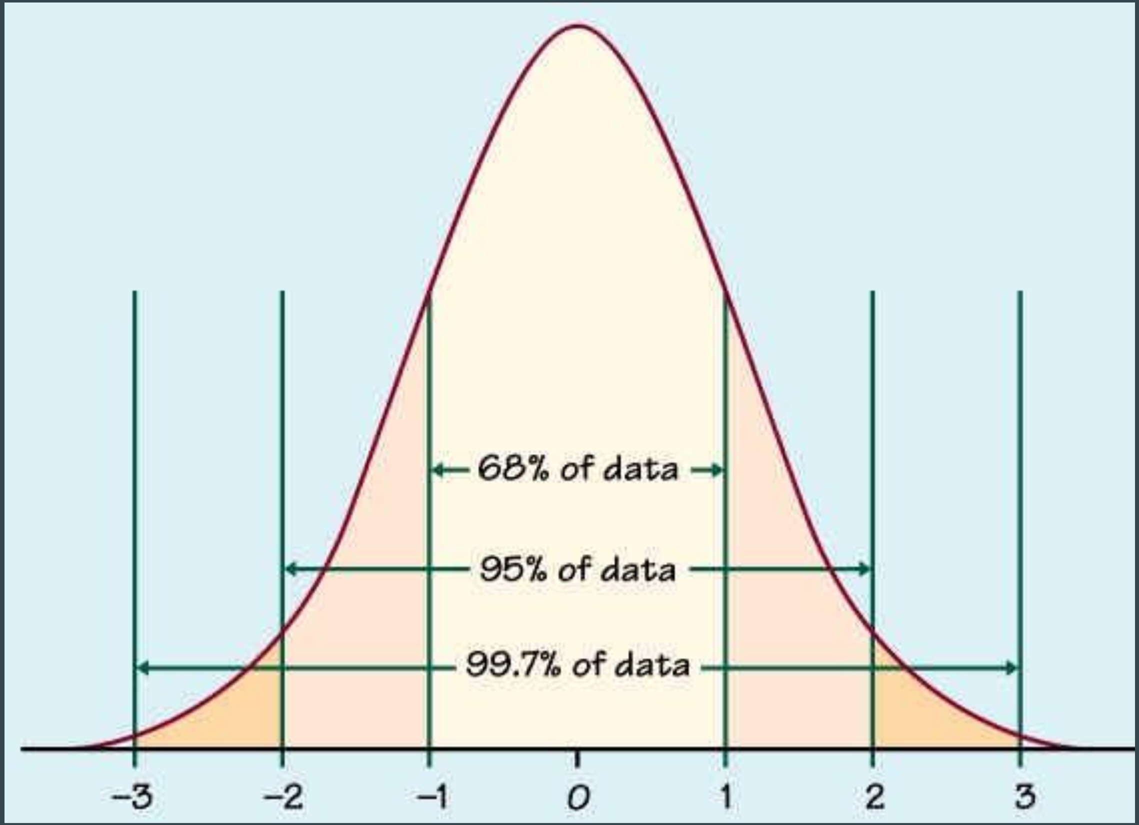
What do we do with standard errors?

- Standard errors provide a measure of uncertainty for our point estimates
- We can compare the quality of point estimates on the same scale...
- ...but not really helpful if we’re only looking at a single point

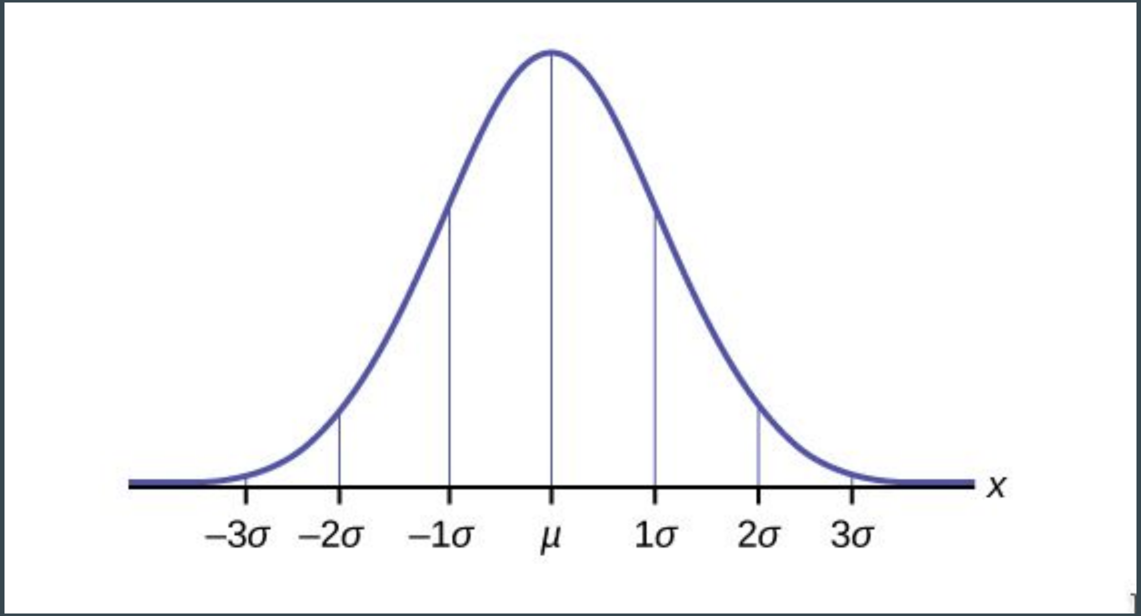
Confidence intervals

- Confidence intervals also measure uncertainty, like standard errors
- They translate standard errors from being in units of X to units of Y (i.e. probability)
- We want to know the *probability* that our point estimate is within a certain range of values

Point estimates follow a normal distribution



Point estimates follow a normal distribution



Confidence interval formula

- Select z-score for probability level you're interested in
  - Z = 1; 68% confident
  - Z = 2; 95% confident
  - Z = 3; 99.7% confident
- CI = point estimate $\mp$ z * SE

Confidence Interval example

A sample of 1000 likely voters finds 62% support legalizing marijuana, with a variance of 4%.

SE = sqrt(4%) / sqrt(1000) = 6.3%

95% CI = 62% ∓ 2 * 6.3%

95% CI = 49.4% - 74.6%

Confidence interval interpretation

A confidence interval means that there is 95% chance the true population parameter lies within the range of the C.I.

- In other words, 95 out of 100 perfectly random samples will have C.I. that overlap with the true population mean
- All based on CLT, so all CLT assumptions apply

Null Hypothesis Significance Testing: overview

The null hypothesis states that nothing changes.

H_0: Nothing changes

- Any deviation from E(X) due to random chance

H_A: Something changes

- Deviation from E(X) due to systematic change

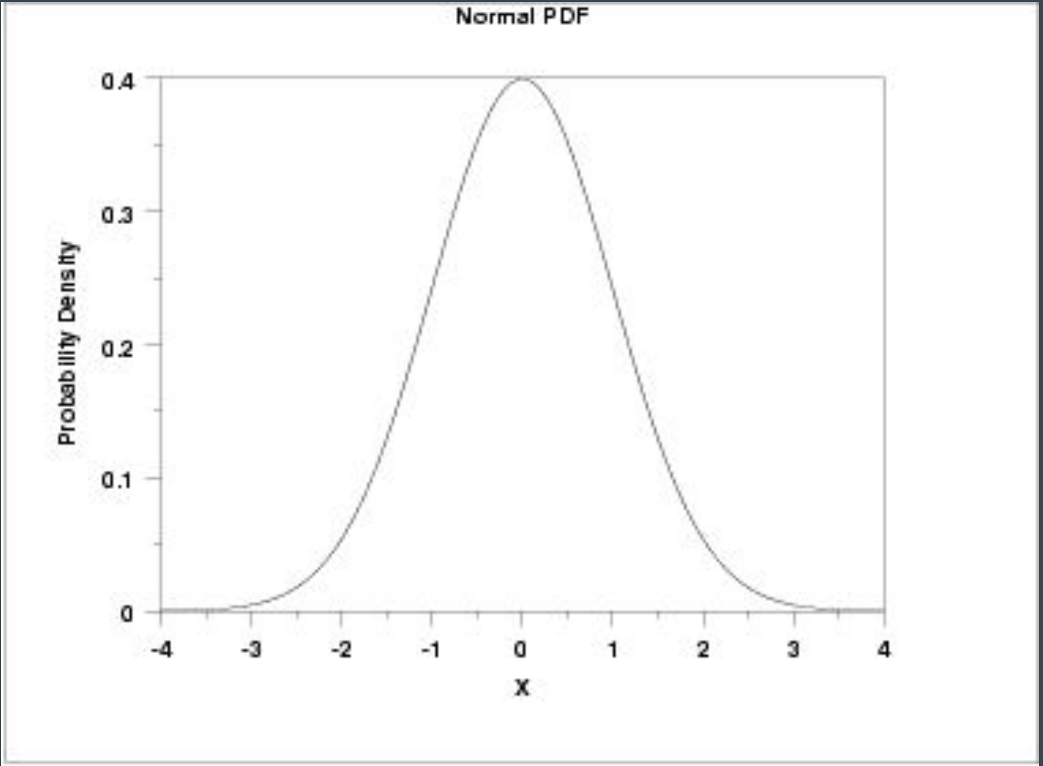
Null Hypothesis Significance Testing: overview

The null hypothesis states that $p(x_i = \text{point estimate})$ □ null distribution

$$H_0: p(x_i = \text{point estimate}) \;\square\; \text{null distribution}$$

$$H_A: p(x_i = \text{point estimate}) \;\square\; \text{alternative distribution}$$

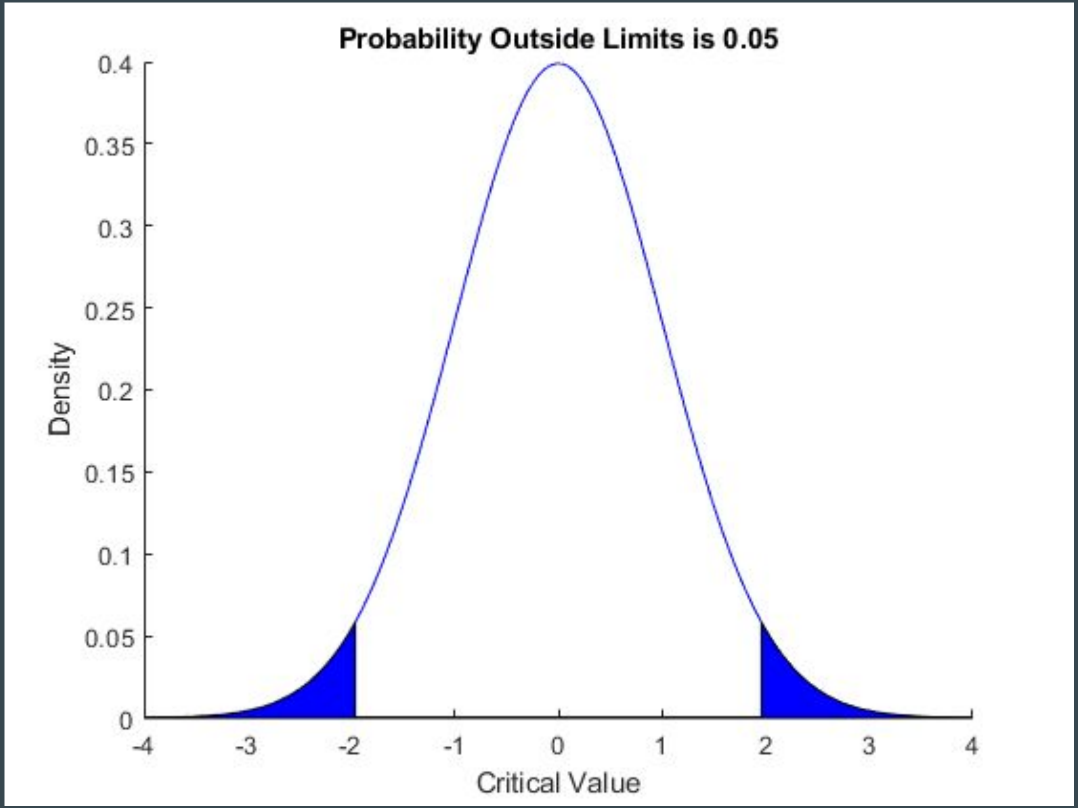
Null distribution: nothing changes



α: critical value

- This is the level at which you deem an observation “extreme” or “out of the ordinary”
- In other words, it’s so extreme, that we actually think it is drawn from a different distribution than the rest of our data
- α is often set to the fifth percentile of extreme values

α: critical value



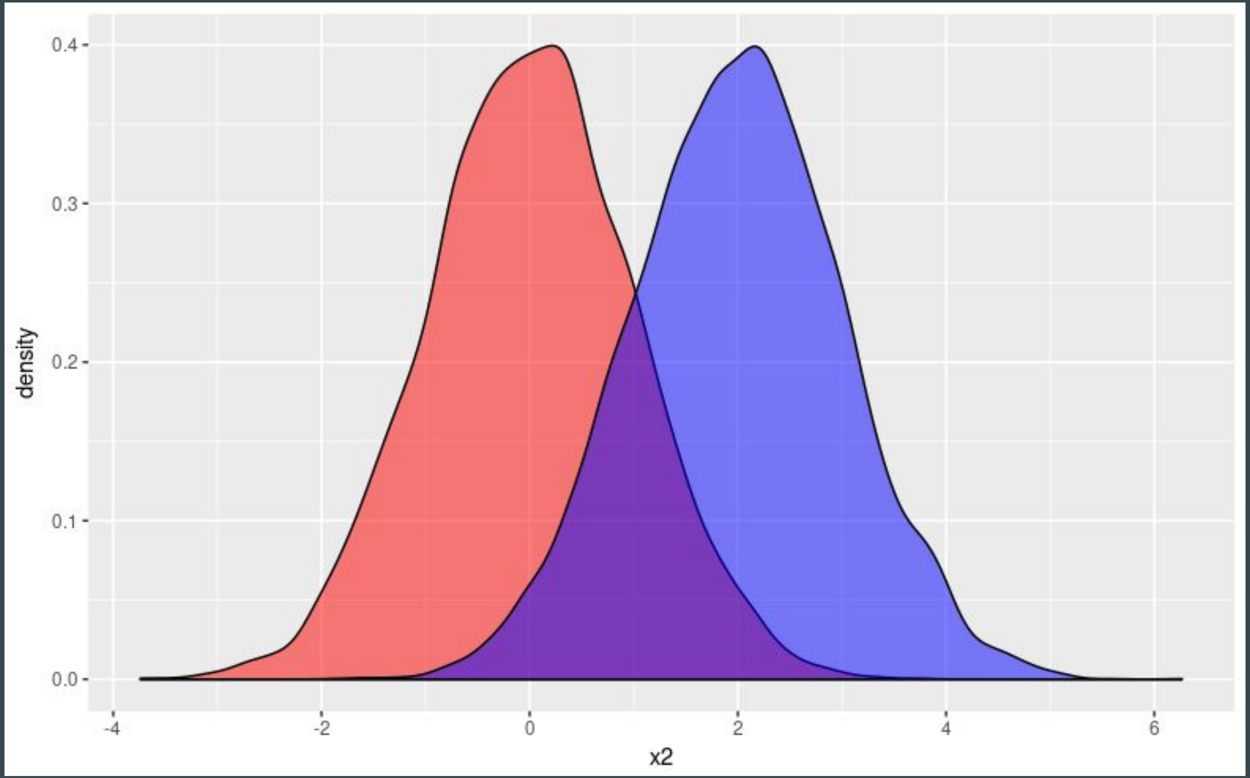
NHST steps

1. Calculate z-score for point estimate you're interested in
   - Use standard error instead of standard deviation
   - $P(x_i) = P(z_i)$
2. Evaluate z-score with respect to the null distribution
   - Normal(μ=0, σ=1)
3. If $p(z_i < \alpha)$ is true
   - Reject H_0 / Accept H_A
4. If $p(z_i > \alpha)$ is true
   - Accept H_0 / Reject H_A

$p(z_i)$ = P-value

- If the null distribution is true, we have an α chance to get the point estimate that we got
- If the p-value is below α we say that we accept the alternative hypothesis
    - Our point estimate is so rare in the null distribution, that we think it follows some alternative distribution

Null vs alternative distributions



NHST example

After a presidential debate, the candidate has an approval rating, drawn from a sample of 100 voters, of 53%. The average of past approval polls has been 48% with a 0.5% standard deviation. Did the debate make a difference in the candidate's approval?

1. Calculate z-score

After a presidential debate, the candidate has an approval rating, drawn from a sample of 100 voters, of 53%. The average of past approval polls has been 48% with a 0.5% standard deviation. Did the debate make a difference in the candidate's approval?

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

$where,$

$\overline{X} = the\ sample\ mean$

$\mu = the\ population\ mean$

$\sigma = the\ population\ standard\ deviation$

$N = the\ sample\ size$

1. Calculate z-score

After a presidential debate, the candidate has an approval rating, drawn from a sample of 100 voters, of 53%. The average of past approval polls has been 48% with a 5% standard deviation. Did the debate make a difference in the candidate's approval?

$$Z = 53 - 48 / (5/\text{sqrt}(100)) = 2.5$$

2. Evaluate with respect to the null distribution

After a presidential debate, the candidate has an approval rating, drawn from a sample of 100 voters, of 53%. The average of past approval polls has been 48% with a 5% standard deviation. Did the debate make a difference in the candidate's approval?

$$Z_i = 2.5$$

$$P(z_i = 2.5 \mid \text{null distribution}) = .006$$

3. If $p(z_i < \alpha)$, reject H_0 and accept H_A

After a presidential debate, the candidate has an approval rating, drawn from a sample of 100 voters, of 53%. The average of past approval polls has been 48% with a 5% standard deviation. Did the debate make a difference in the candidate's approval?

We would expect to see a 53% approval rating, assuming nothing changed, only 0.6% of the time. Therefore, we argue that the debate changed the underlying distribution of voter approval, and our 53% approval rating was drawn from this alternative distribution.

We can also do NHST through confidence intervals

After a presidential debate, the candidate has an approval rating, drawn from a sample of 100 voters, of 53%. The average of past approval polls has been 48% with a 5% standard deviation. Did the debate make a difference in the candidate's approval?

Point estimate ± z * SE

53 ± 2 * (5 / sqrt(100)) = (52%, 54%)

(52, 54) does not overlap with 48%

What p-values do not do

1. P-values are not statements about whether the alternative hypothesis is true, or whether the null hypothesis is false
2. P-values are not the probability that a given event is the product of random chance
3. Smaller p-values do not necessarily mean larger effects

P-values are not a "truth percentage"

- A p-value is NOT the % chance that the alternative distribution is true!
- The only reason we know the null distribution is because of CLT
  - Main assumption of CLT is i.i.d.
- We have no idea what the alternative distribution is
- P-values can only be interpreted with respect to the null distribution

P-values are not likelihood something due to random chance

- Say we calculate a .04 p-value
- That does not mean there is a 4% chance this is due to random chance
- NHST *assumes null distribution is TRUE*
  - *Assumes that all variation is purely due to random chance*
- Calculates the likelihood of a certain point estimate under that assumption
- This tells us nothing about whether that assumption is true!

P-values are not about size of effect

After a presidential debate, the candidate has an approval rating, drawn from a sample of 100 voters, of 53%. The average of past approval polls has been 48% with a 5% standard deviation. Did the debate make a difference in the candidate's approval?

P-values are not about size of effect

After a presidential debate, the candidate has an approval rating, drawn from a sample of 1000 voters, of 53%. The average of past approval polls has been 48% with a 5% standard deviation. Did the debate make a difference in the candidate's approval?

Z = 53 - 48 / (5/sqrt(1000)) = 31

P-value = .000000000003

Review

- We want to estimate parameters of distributions, like the mean
  - We call these “point estimates”
- We use standard errors to quantify our uncertainty about those point estimates
- Using the CLT, we can construct confidence intervals to provide a range of plausible values
- We can also use the CLT to do Null Hypothesis Significance Testing
  - Assume the null distribution
  - Provide a probability (p-value) of the likelihood of a given point estimate under the null
  - Reject or accept a value as too extreme to suggest a different distribution