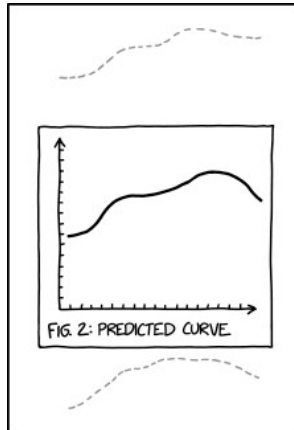


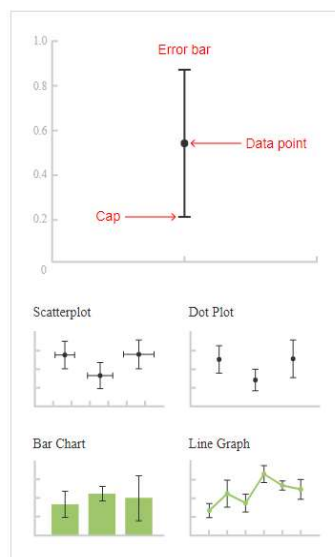
Graphic Representations of Confidence Intervals



SCIENCE TIP: IF YOUR MODEL IS BAD ENOUGH, THE CONFIDENCE INTERVALS WILL FALL OUTSIDE THE PRINTABLE AREA.



Error Bars



Types of Error Bars

- Types of error bars:
 - Confidence interval (90%, 95%, 99%)
 - Standard error (or standard error times a multiplier, typically a z-score of 2)
 - Standard deviation (or standard deviation times a multiplier)
- Recommended: confidence interval error bars
- When examining graphs with error bars, check the notes for the type of error bars used

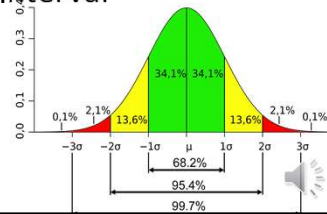


Comparison for the Same Data

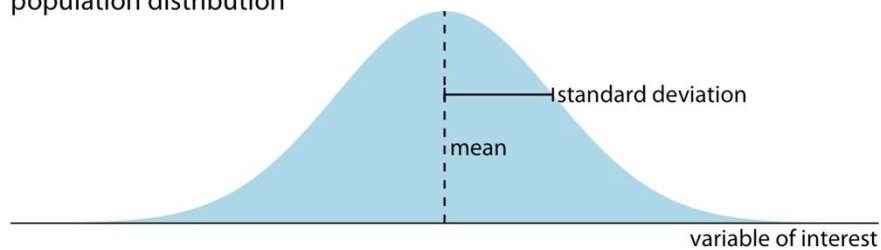


Clarifying Terminology

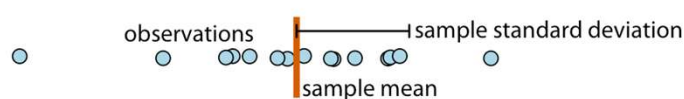
- Error bar – a way to graphically display either a confidence interval or a standard error (look the same)
- Standard error – measure of variation of the sampling distribution; for the mean, SEM: $\sigma_{\bar{X}} = s/\sqrt{n}$
- Standard deviation (s) – measure of variation for the variable itself, always larger than standard error
- Confidence interval – an interval estimate of the mean.
95% confidence interval = 4 SE wide ($\bar{X} \pm 1.96 \cdot \sigma_{\bar{X}}$)
- If an error bar displays the standard error, it can be viewed as a 68% confidence interval = 2 SE wide
- Margin of error – $\frac{1}{2}$ of a confidence interval
- For 95% CI, margin of error $\pm 2SE$



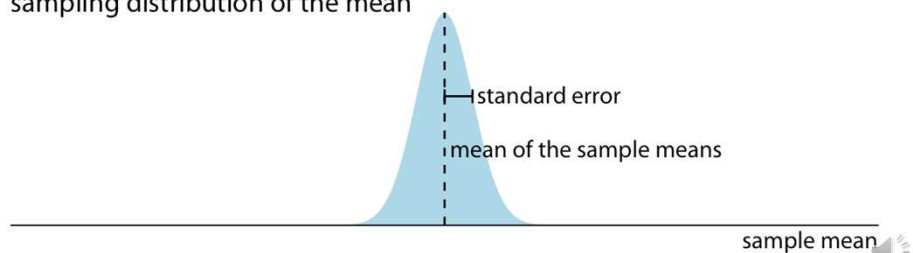
population distribution

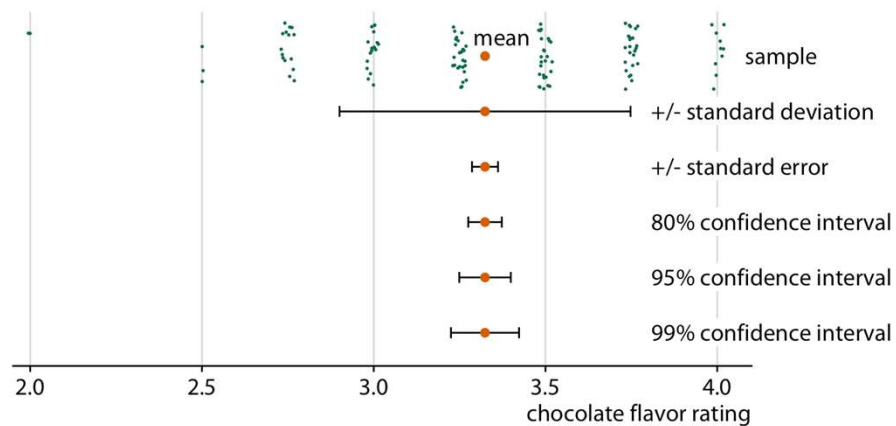
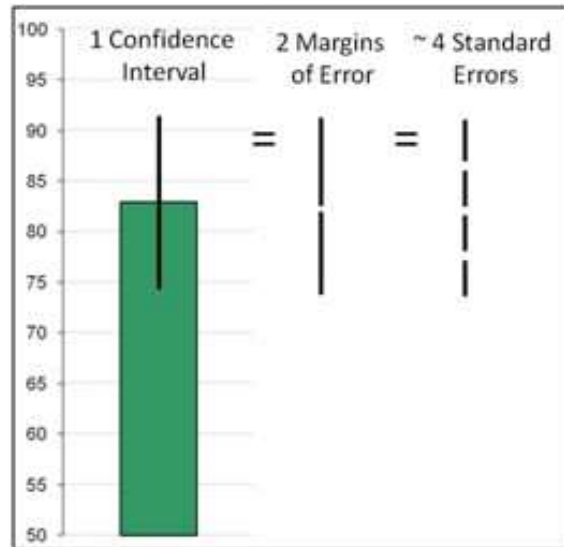


sample



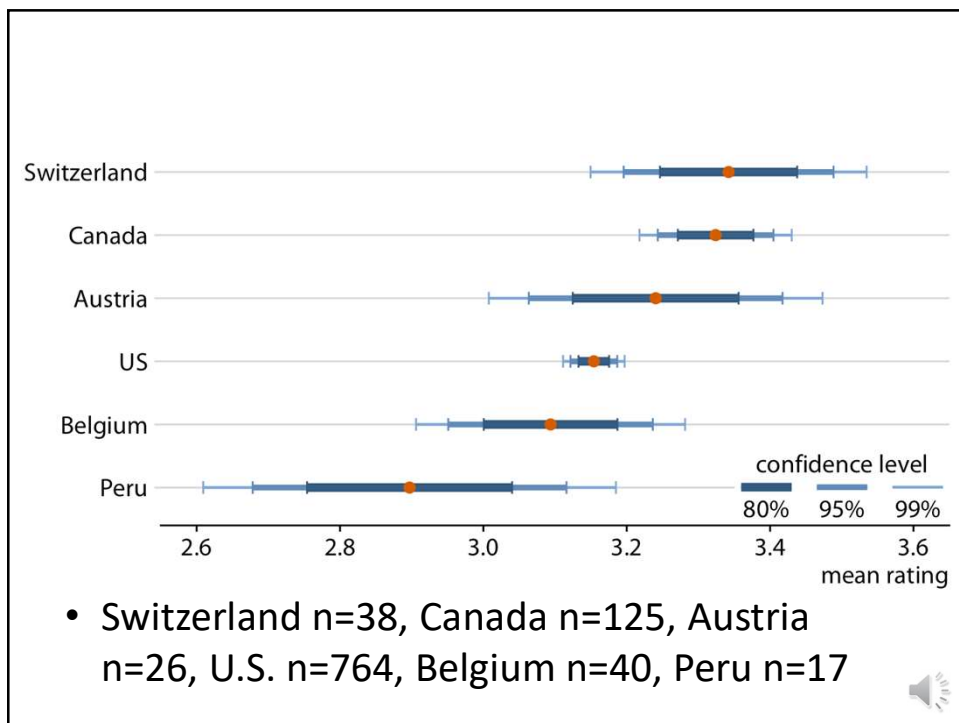
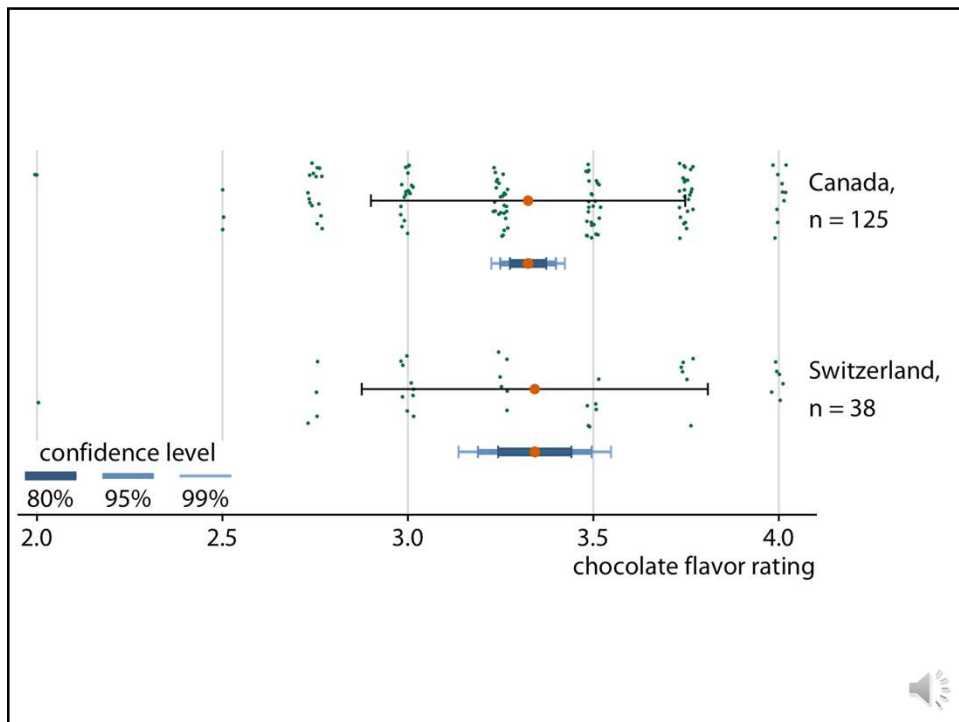
sampling distribution of the mean





The observations (jittered green dots) represent expert ratings of 125 chocolate bars made in Canada, on a scale from 1 (unpleasant) to 5 (elite). Data source: Brady Brelinski, Manhattan Chocolate Society. From: <https://clauswilke.com/dataviz/visualizing-uncertainty.html>





What Makes Error Bars Longer? (Same Things That Make Confidence Intervals Wider)

- Standard deviation > 99% CI > 95% CI > 90% CI > standard error
- Smaller sample size → longer error bars
- Less homogenous population → longer error bars

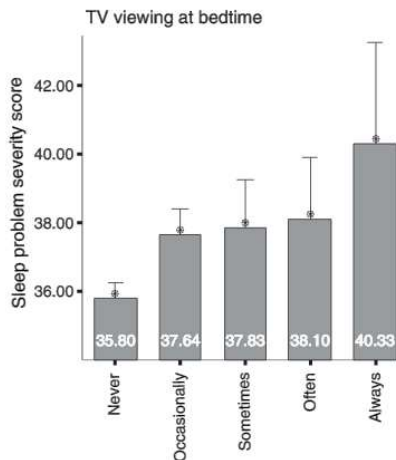
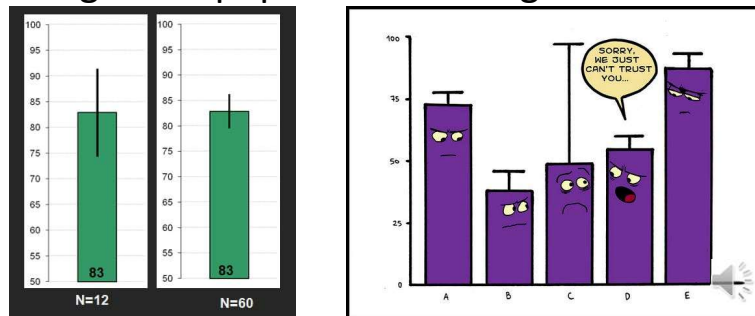
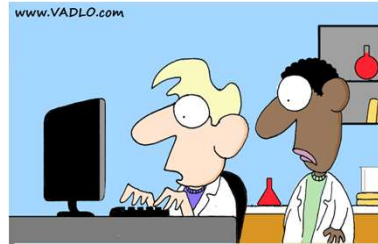
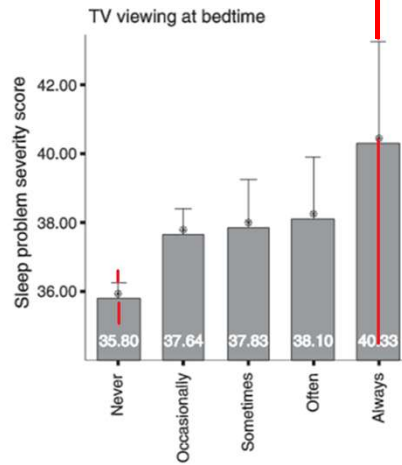


Figure 2. Mean sleep problem severity scores in children with different frequency of TV viewing at bedtime. The error bars represent the standard error of mean.

- Source: E. Juulia Paavonen, Marjo Penonen, Mira Roine, Satu Valkonen, & Anja Riitta Lahikainenj. "TV Exposure Associated With Sleep Disturbances in 5- To 6-Year Old Children." 2006. Sleep Research 15, 154–161

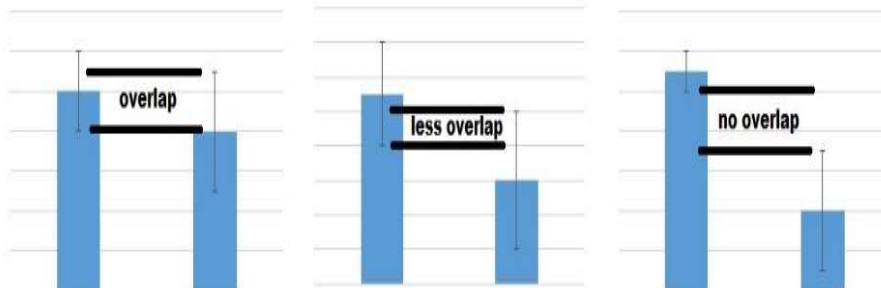


"I can see your error bars using google earth"

Figure 2. Mean sleep problem severity scores in children with different frequency of TV viewing at bedtime. The error bars represent the standard error of mean.

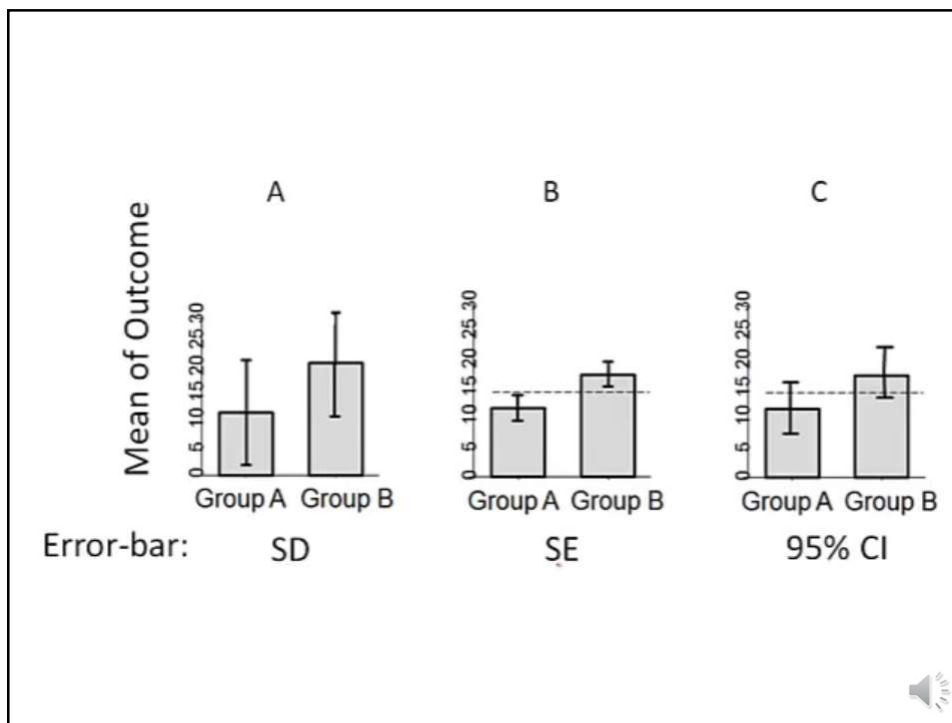
Using Error Bars to Compare Groups: Look for Overlap

- You should assess both confidence intervals and standard errors for overlap

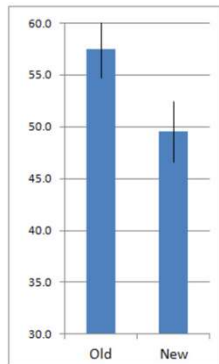


Decision about Group Difference

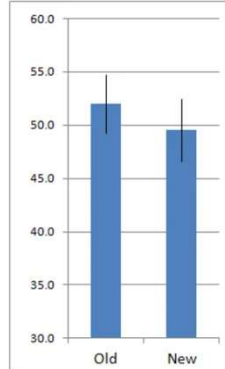
- Both confidence intervals and standard errors overlap → two groups are not different
- Neither confidence intervals nor standard errors overlap → two groups are different
- Confidence intervals overlap but standard errors do not → further testing is needed
- Also note:
 - If SE overlap → CI will always overlap
 - If CI do not overlap → SE will also not overlap



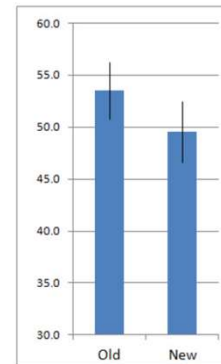
Bar Graphs with 95% Confidence Intervals



No overlap,
different
(95% sure)



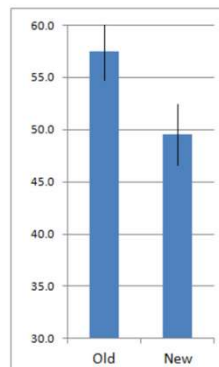
Lots of overlap:
Check SE



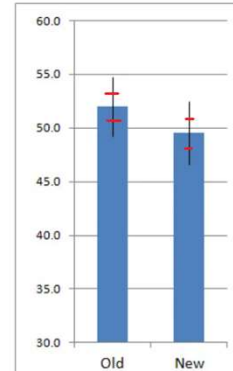
Some overlap:
Check SE



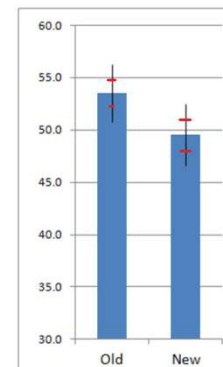
Bar Graphs with 95% Confidence Intervals



No overlap,
different
(95% sure)



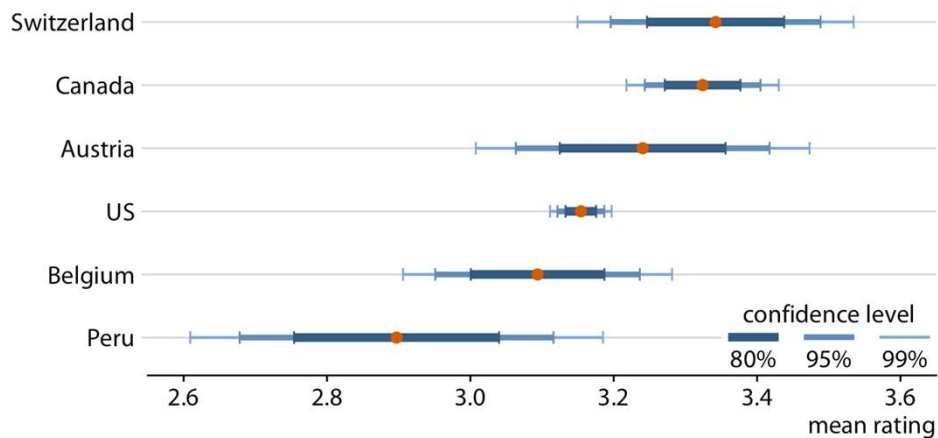
Lots of overlap:
Check SE →
SE also overlap,
no difference



Some overlap:
Check SE →
SE don't overlap,
need more testing



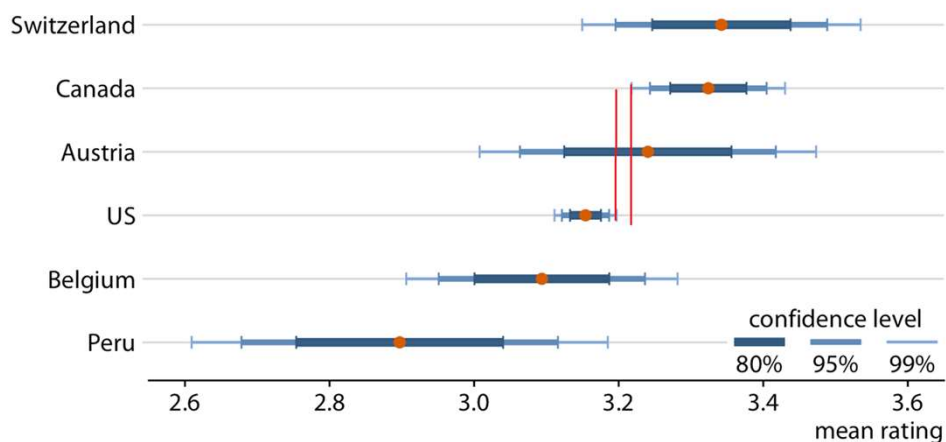
Back to Chocolate Bars



- Switzerland n=38, Canada n=125, Austria n=26, U.S. n=764, Belgium n=40, Peru n=17



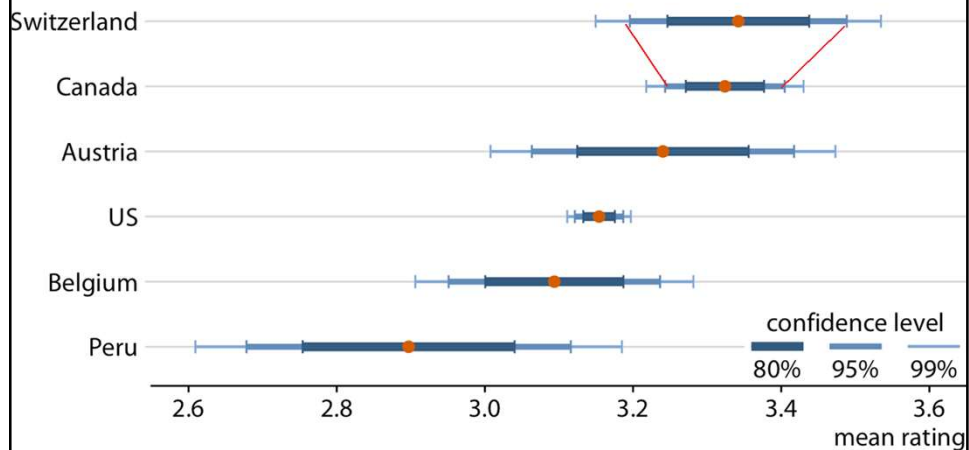
Canada vs US: No Overlap in 99% CI



- Conclusion: 99% sure that different



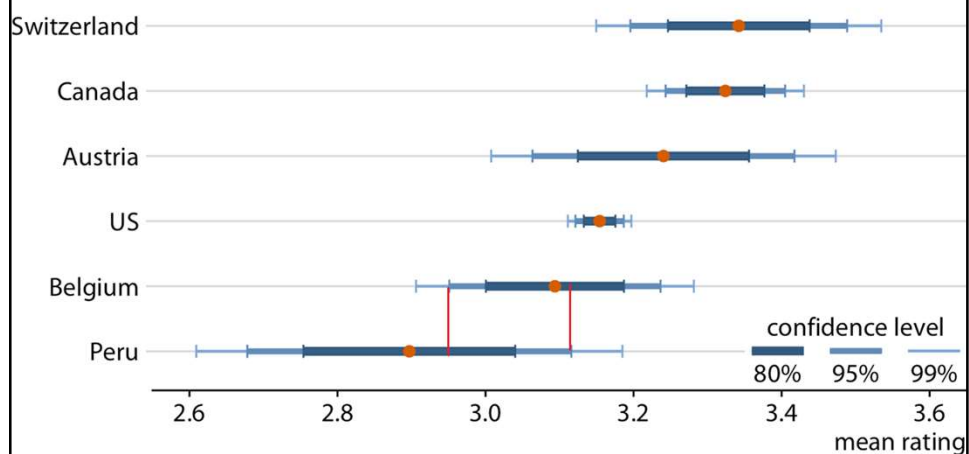
Canada vs Switzerland: Full Overlap



- Conclusion: SE also overlap → not different



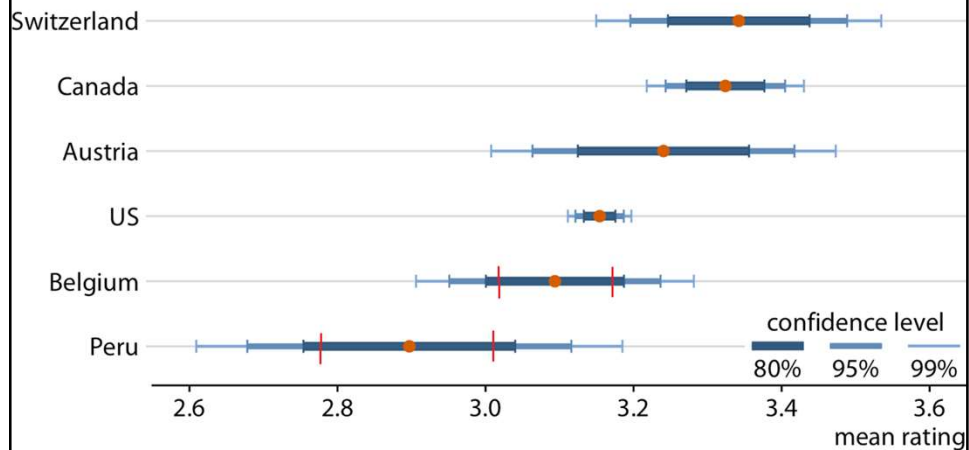
Belgium vs Peru: Partial Overlap



- Need to check if SE overlap



Belgium vs Peru: Standard Errors



- 95% CI overlap but SE do not → need further testing



Example: Confidence Interval Error Bars

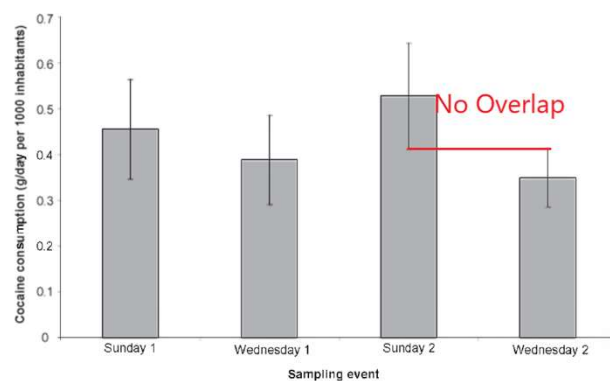
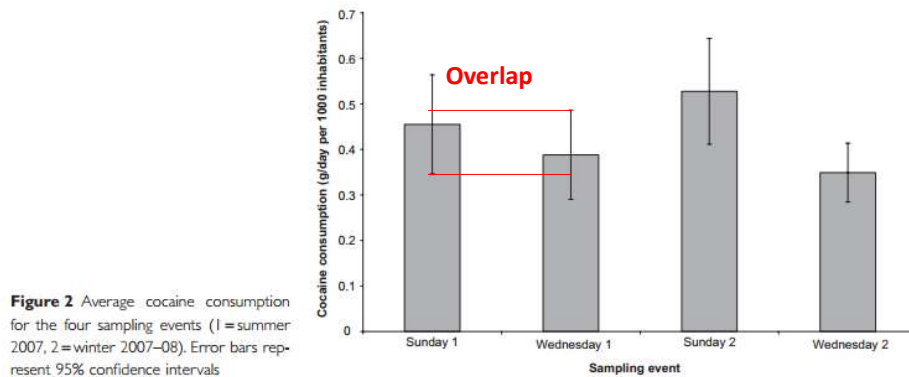


Figure 2 Average cocaine consumption for the four sampling events (1 = summer 2007, 2 = winter 2007–08). Error bars represent 95% confidence intervals

- Alexander L. N. van Nuijs et al. 2009. "Can cocaine use be evaluated through analysis of wastewater? A nation-wide approach conducted in Belgium." *Addiction*, 104, 734–741.
- Sunday2 vs Wednesday2: CI don't overlap → different



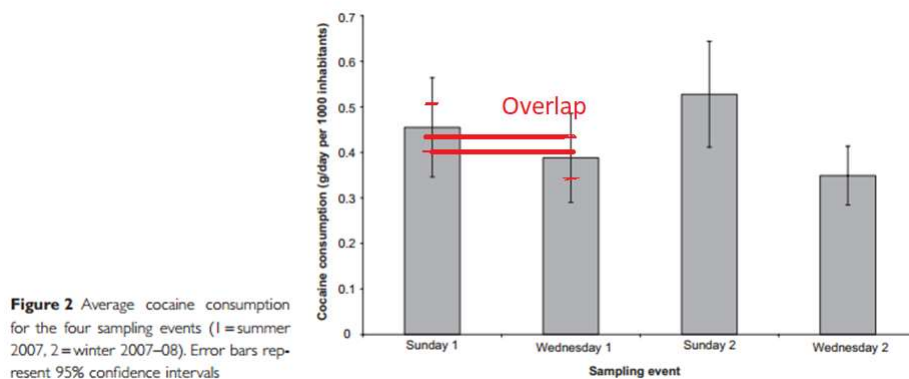
Example: Confidence Interval Error Bars



- Sunday1 vs Wednesday1: CIs overlap, check SE



Example: Confidence Interval Error Bars



- Sunday1 vs Wednesday1: SE also overlap → no difference



Example: Standard Error Bars

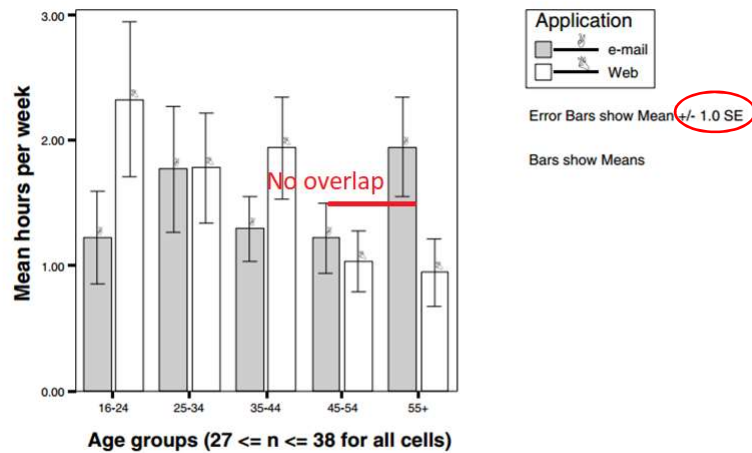


Figure 3: Mean Hours per Week Spent on E-Mail or Web Browsing by Diary Respondents at Wave 2 in 2000

- Ben Anderson & Karina Tracey. 2001. "Digital Living: The Impact (or Otherwise) of the Internet on Everyday Life." American Behavioral Scientist 45(3):456-475.

Example: Standard Error Bars

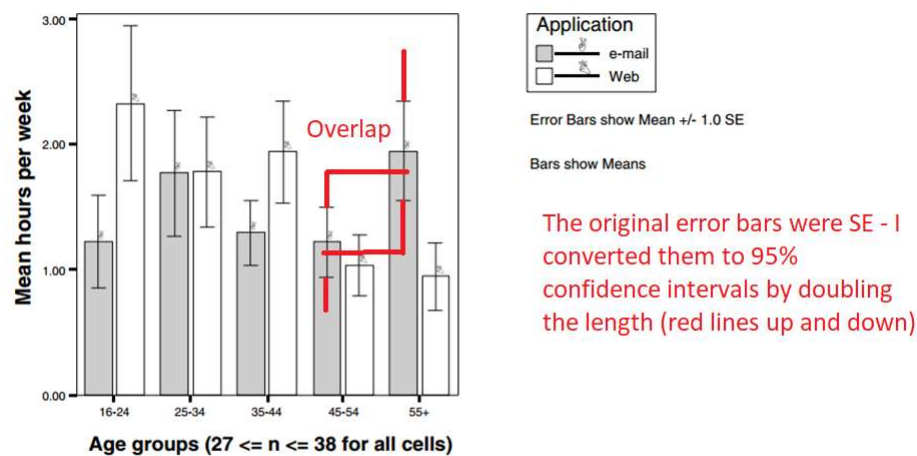


Figure 3: Mean Hours per Week Spent on E-Mail or Web Browsing by Diary Respondents at Wave 2 in 2000

- SE didn't overlap but CIs do $\rightarrow$ need more testing

Error Bars are Misleading for Pre-test and Post-test

- Don't use if two bars are based on data from the same people → because retests are highly correlated

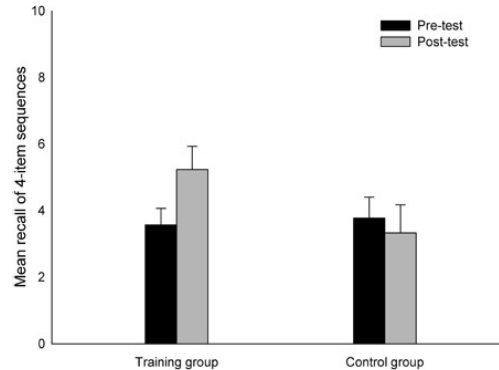
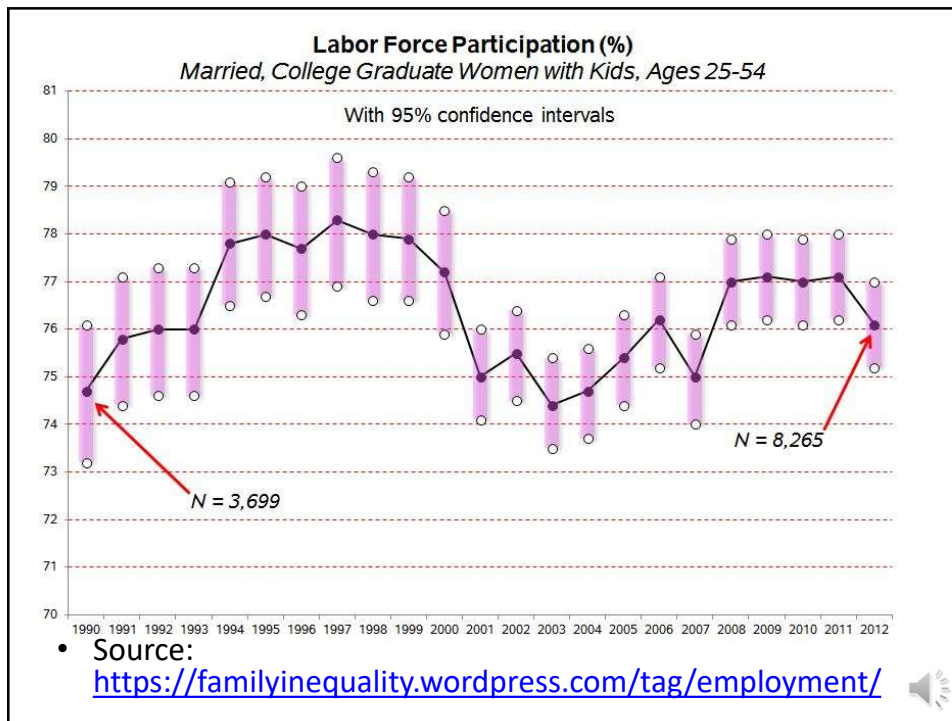
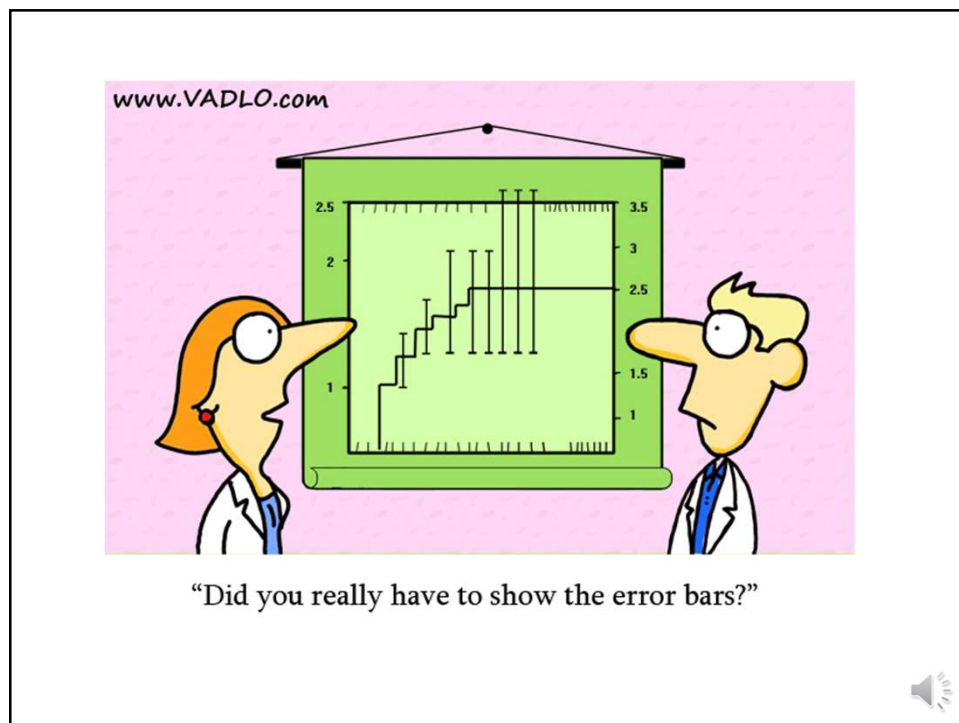
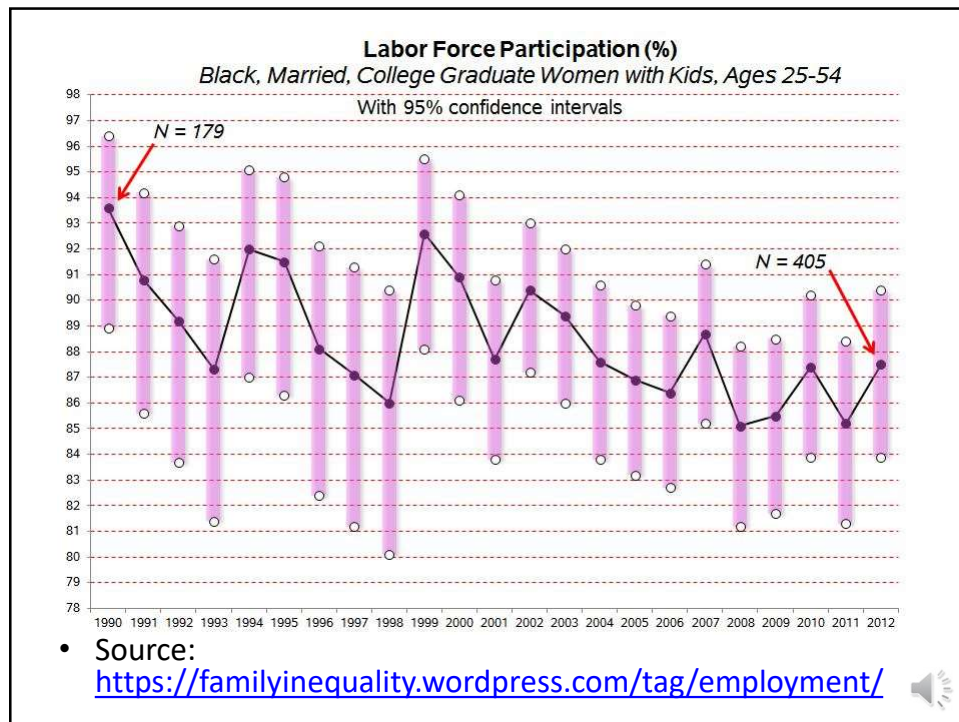


Figure 3. The number of correctly reported four-item sequences in the VS updating task. Performance for both groups is illustrated separately for pre-test and post-test. Error bars indicate standard errors of the mean.

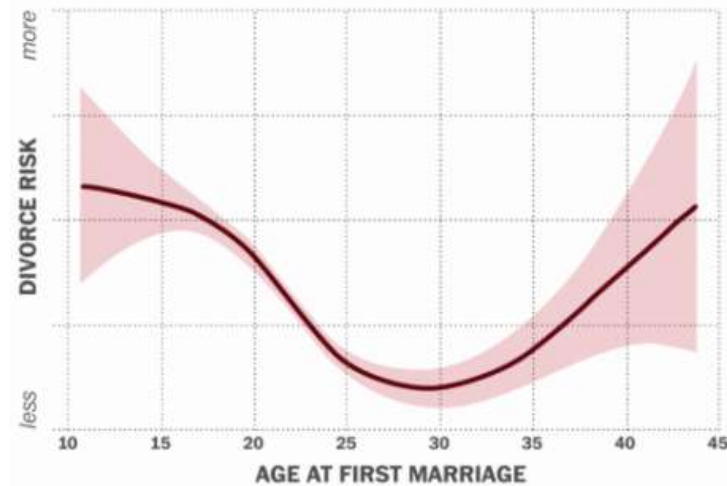
- Source: Salminen T., Strobach T. & Schubert T. 2012. "On the impacts of working memory training on executive functioning." *Frontiers in Human Neuroscience* 6:166.





When (not) to get married.

Age at first marriage and divorce risk as of 2006–2010. Shaded areas represent confidence intervals.

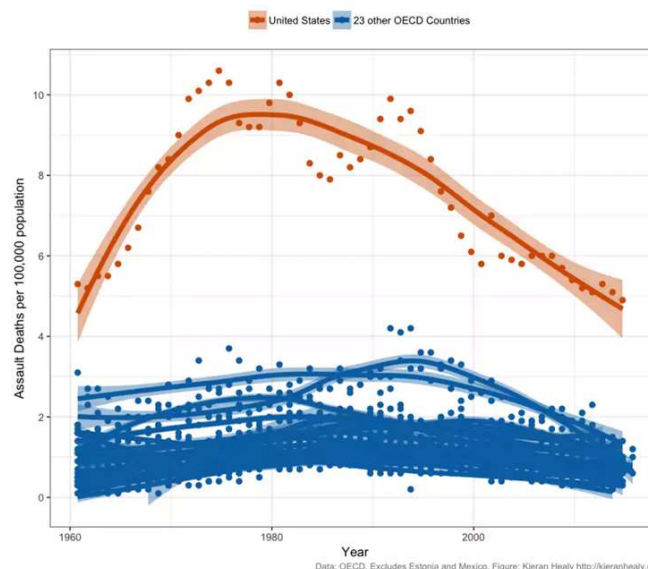


WAP0.ST/WONKBLOG

Source: Nicholas H. Wolfinger analysis of National Survey of Family Growth Data

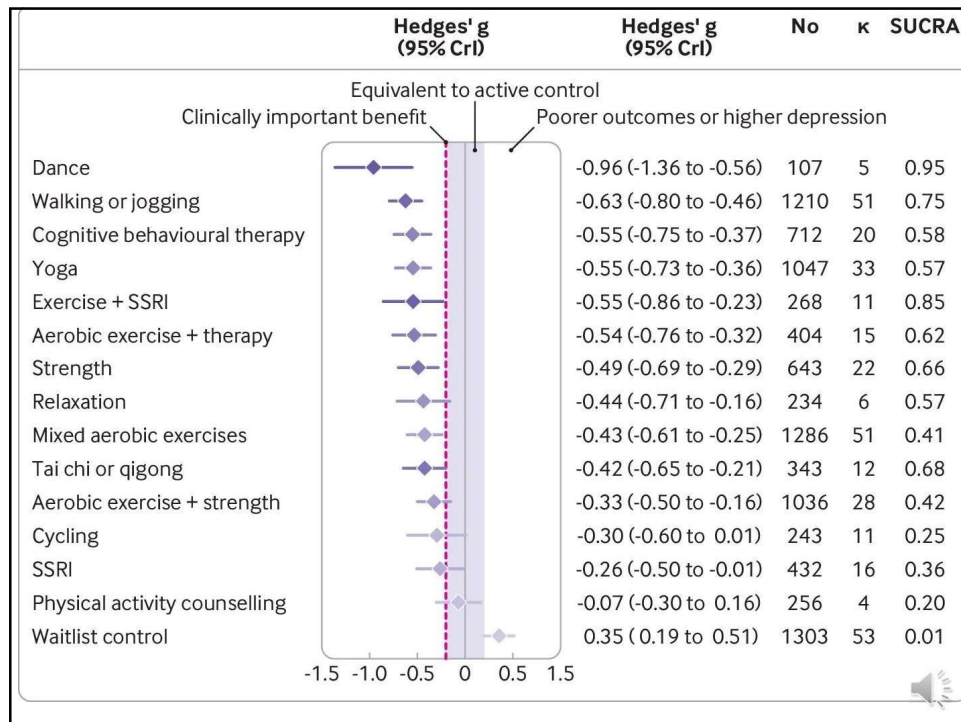


Assault Death Rates in the OECD, 1960-2015



Data: OECD. Excludes Estonia and Mexico. Figure: Kieran Healy <http://kieranhealy.org>





Hurricane Tracking



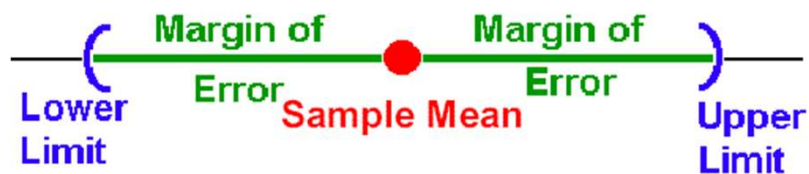
- Forecast cones:
 - Designed to capture the hurricane **center** two-thirds of the time (but impact can extend well away from where the center tracks)
 - That is ~68% confidence interval, which is $\pm 1SE$
 - So the center can end up outside the cone about a third of the time
 - Error used = average error over the prior five years (same cone width for all storms in a given year)
 - In 2025, they are slightly smaller than before (up to 6% in the Atlantic)
- Over the past 20 years, 1-3 day forecast errors have been reduced by 75%, and 4-5 day forecast errors -- by 60%
- The improvement in forecast accuracy has now slowed down – might be at the limit (chaotic nature of the atmosphere)

Public Opinion Polls & Margin of Error



Confidence Interval vs Margin of Error

- Margin of error = the "radius" (or half the width) of a confidence interval





"I'm going to need a Margin of Error or
I can't publish your prediction of
six more weeks of winter."



Confidence Intervals for Proportions

- Same as for mean but do not need a separate estimate of standard deviation

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

Example. For a sample size $n = 900$ and an estimated sample proportion = 0.2:

$$95\% \text{ CI} = 0.2 \pm 1.96 * \text{sqrt}(0.2 * 0.8 / 900) = \\ = 0.2 \pm 0.026$$

$$\text{Probability}(.174 \leq p \leq 0.226) = .95$$

$$\text{Margin of error} = 0.026 * 100 = 2.6\%$$



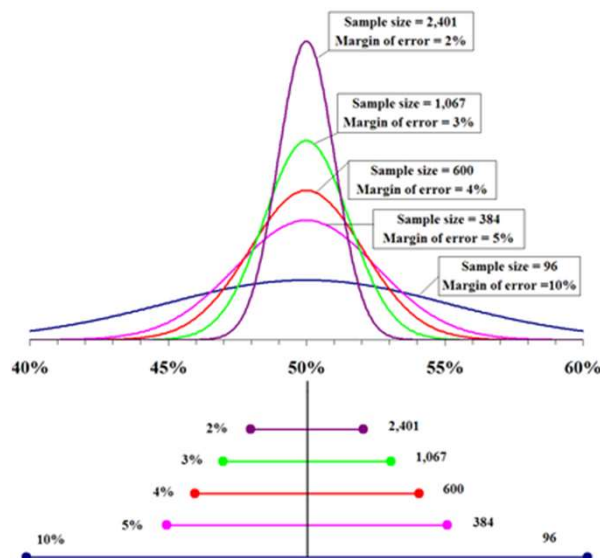
Margin of Error in Polls

- Media usually report the maximum margin of error for any percentage from that poll
- The margin of error for a particular percentage
→ usually smaller
- Maximum margin of error: for percentage=50% (which is proportion =.5); based on sample size:

$$\begin{aligned} CI(p) &= \pm 1.96 \times \sqrt{\frac{.5 \times .5}{n-1}} \\ &\approx \pm 2 \times \frac{.5}{\sqrt{n}} \\ &\approx \pm \frac{1}{\sqrt{n}} \end{aligned}$$



Margin of error and 95% confidence intervals for an estimate of 50%



Example

- Sample size $n = 400$
- Maximum margin of error = $1/\sqrt{n} = 1/\sqrt{400} = 1/20 = 0.05$
- Answer: margin of error = $\pm 5\%$
- But: margin of error applies only when we consider the whole sample, not particular subgroups (e.g., women, Republicans, etc.)

