PennState

Interactive Log files of students' interaction with software applications: Replicative Bayesian Networks on multiple years data

Yiyun Gong*, Dennis Pearl, Matt Beckman, Neil Hatfield, Shunqi Zhang*

*** Undergraduate Researchers**

Project SMILES

- 28 interactive songs
- Students answer prompts on statistical content
- Student responses are inserted into the song using a synthetic voice
- Project website at www.CAUSEweb.org/smiles
- Examples:
  - My Family's Mean
  - Super Bowl Poll

Interactive Log Files

A LEARNING LOCKER TO GATHER LOGFILE DATA ON STUDENTS' INTERACTION WITH SOFTWARE APPLICATIONS

EVALUATIONS ON STUDENTS' UNDERSTANDING OF EACH CONCEPT

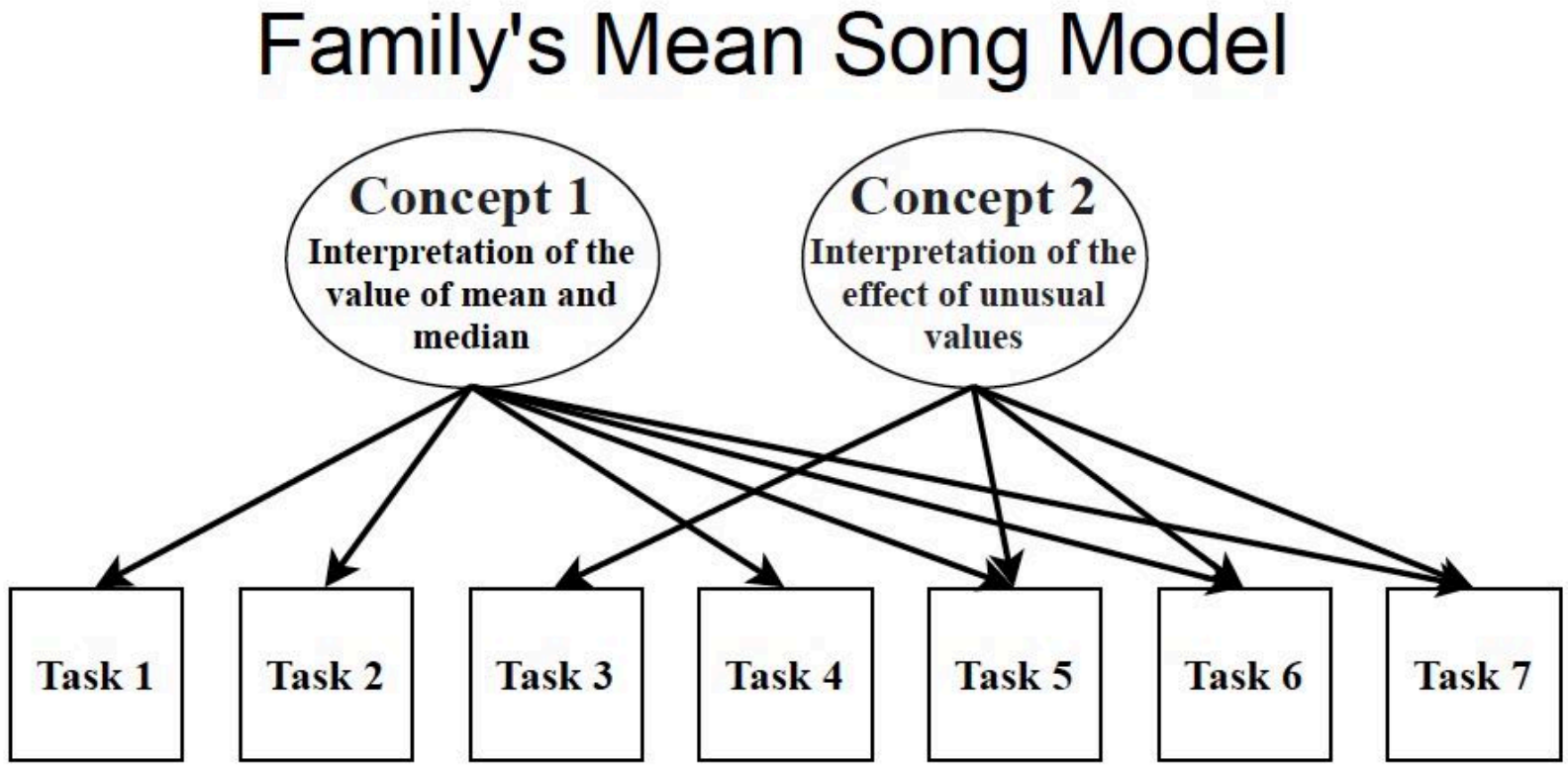
IMPROVE THE SOFTWARE (HINTS, FEEDBACK, INSTRUCTIONS ETC.)

Multiple Years Data

- Fall semesters of the past three years (2018, 2019, 2020)
- 2018: Pilot version
- 2019: Experiences on Large classrooms
- 2020: Various instructional modes due to Covid-19

Bayesian Network Analysis
My Family's Mean

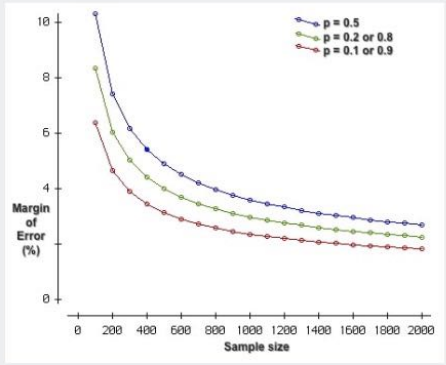
Tasks	Objectives
1	Median as "50th percentile"
2	Mean as numerical average
3	Mean as sensitive to outliers
4	Median as "typical" value
5	Median as robust to outliers
6	Median as robust to outliers
7	Median < mean for skewed right histogram



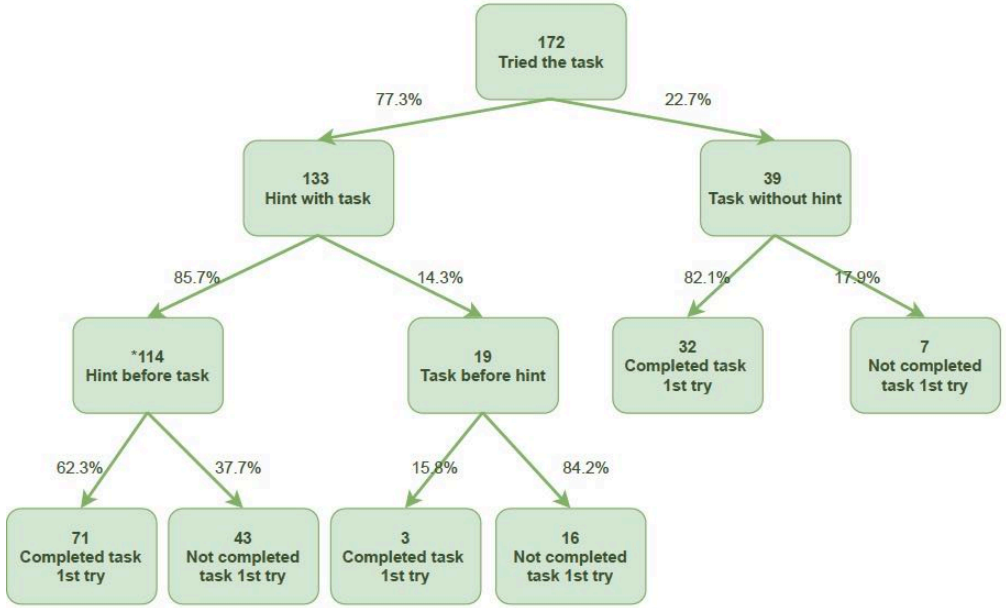
Tree Diagram on Hint Use - Super Bowl Poll

2. The margin of error for a **sample proportion** for a survey of 1000 people would be about ____ %. Hint

Here is a graph for how the margin of error changes with sample size and the population proportion p:

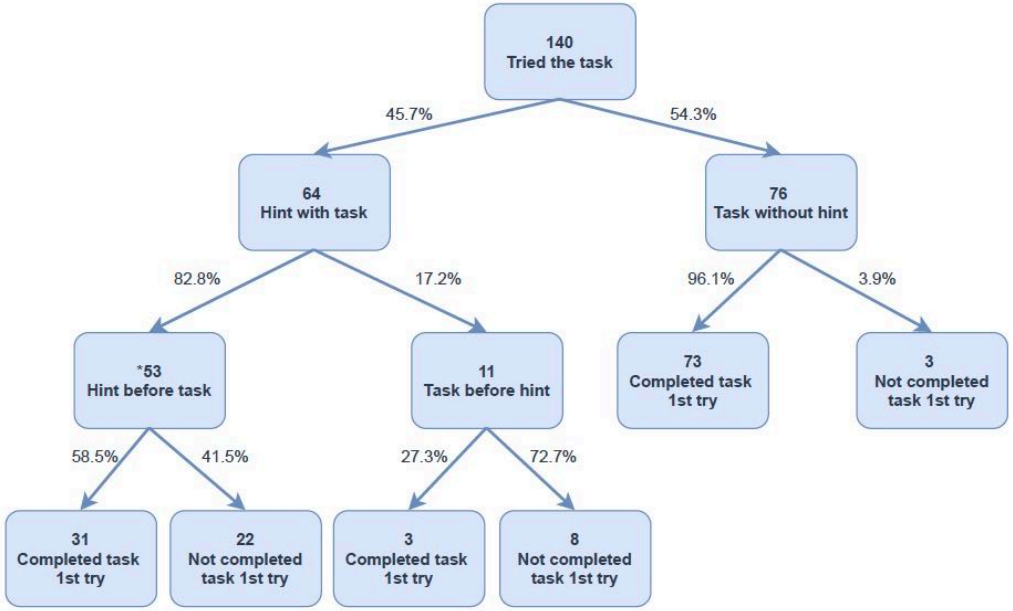


2019



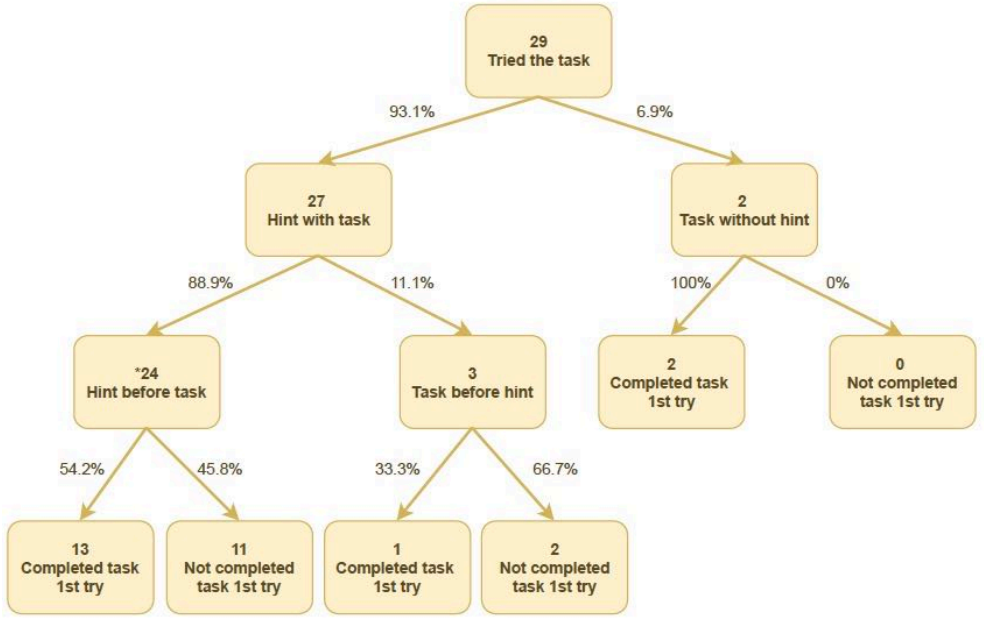
*: Samples that was used to train the Bayesian Network Model

2020



*: Samples that was used to train the Bayesian Network Model

2018



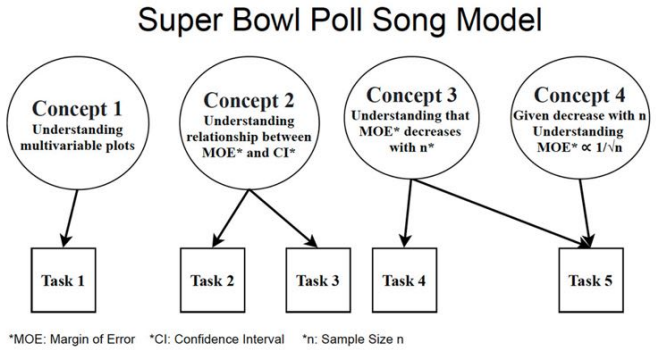
*: Samples that was used to train the Bayesian Network Model

Bayesian Network Analysis
Super Bowl Poll

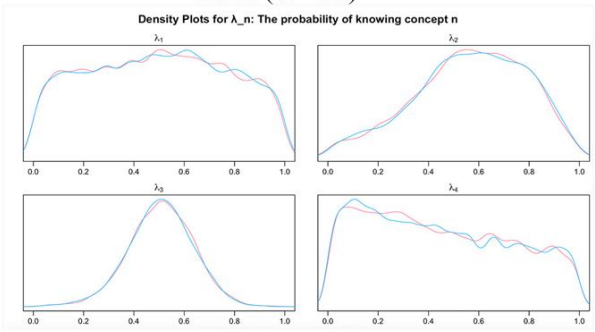
Super Bowl Poll

1. The NFL is a professional American football league of 32 teams. Pick your favorite team — or just any team you think might have a chance to win the Super Bowl (championship game): Arizona Cardinals · This question is unused in the model
2. The margin of error for a **sample proportion** for a survey of 1000 people would be about 3 %. Hint
3. If 17% is the **sample percentage**, then the **margin of error** you entered in the above item gives an interval estimate as low as 14 % and as high as 20 %.
4. If you multiplied the **sample size** by a factor of nine, that would decrease the margin of error.
5. The **margin of error** would then decrease by a factor of 3 .

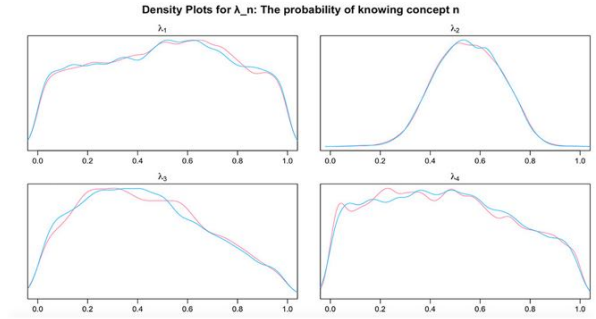
Continue



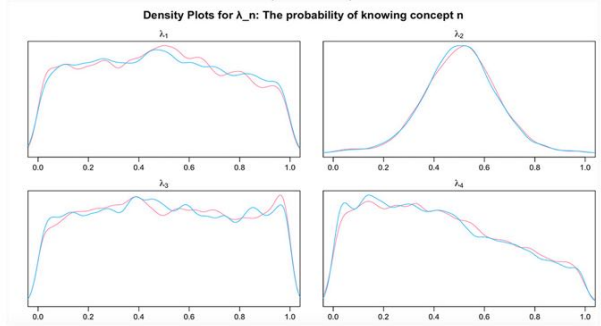
2020 (n = 53)



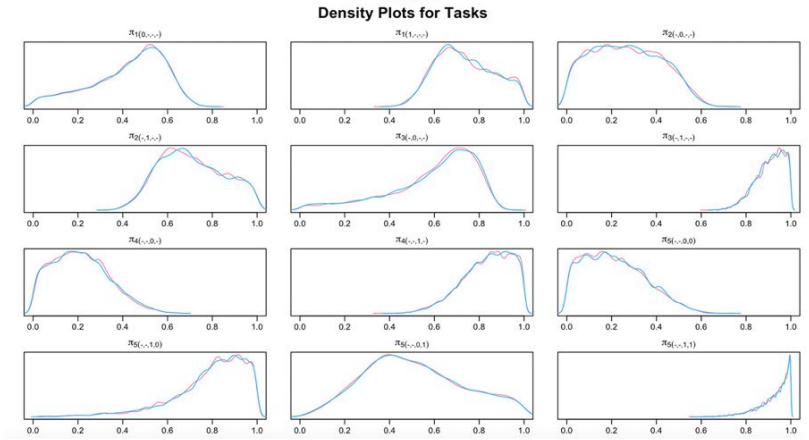
2019 (n = 114)



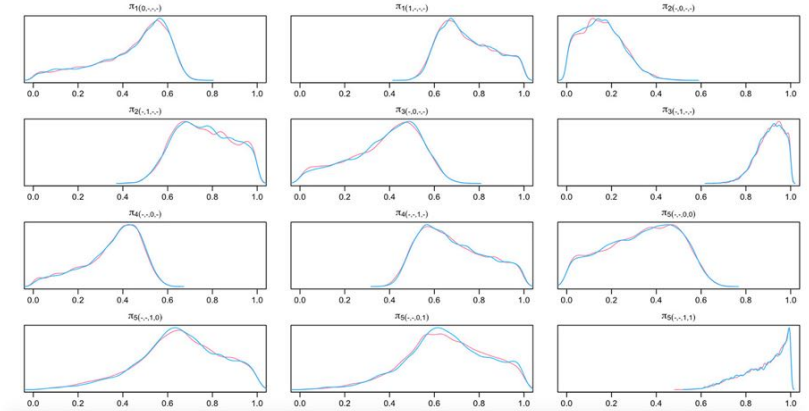
2018 (n = 24)



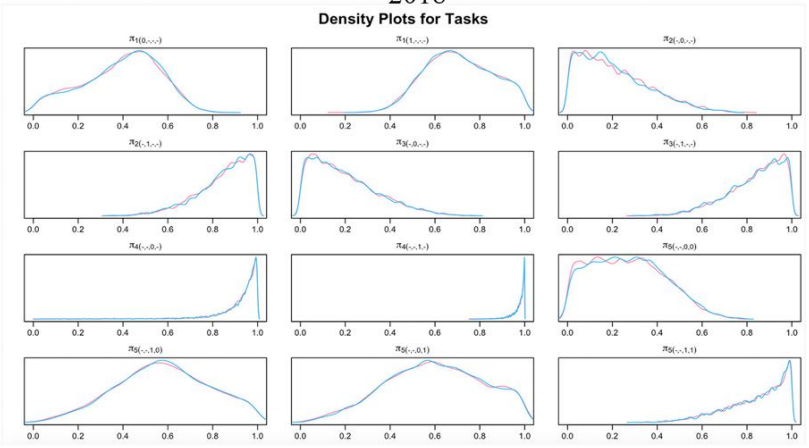
2020



2019



2018



Conclusions & Discussions

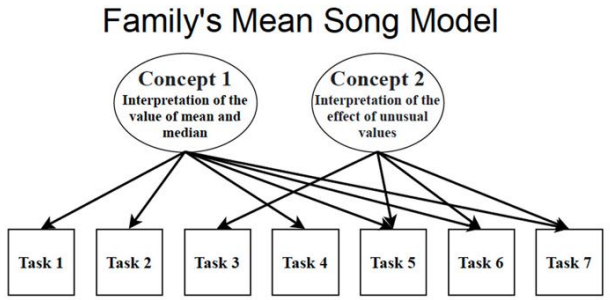
- Advantages of Log Files
  - Student flow through the app (For example, use of hint)
  - Studies of conceptual knowledge versus the ability to do tasks (For example, Bayesian Network)
  - Potential interface upgrade base on where students get stuck
  - Anonymous data with little/no human subject concerns
- Multi-Year Cross-Sectional Comparisons
- Limitation: No longitudinal follow-up without log in

Q&A Appendix

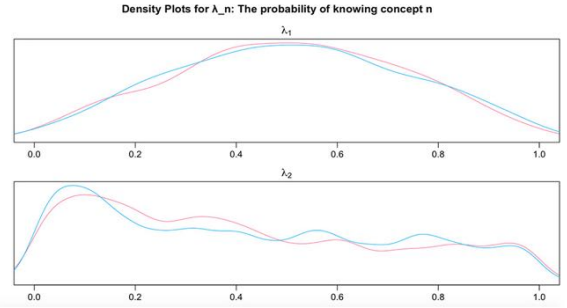
PennState

Bayesian Network Analysis
My Family's Mean

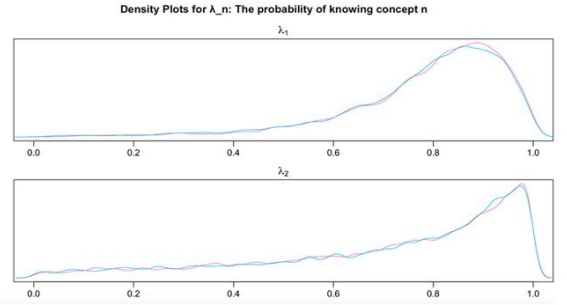
Tasks	Objectives
1	Median as "50th percentile"
2	Mean as numerical average
3	Mean as sensitive to outliers
4	Median as "typical" value
5	Median as robust to outliers
6	Median as robust to outliers
7	Median < mean for skewed right histogram



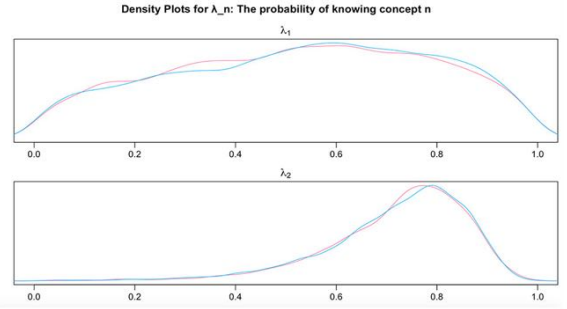
2020 (n = 4)



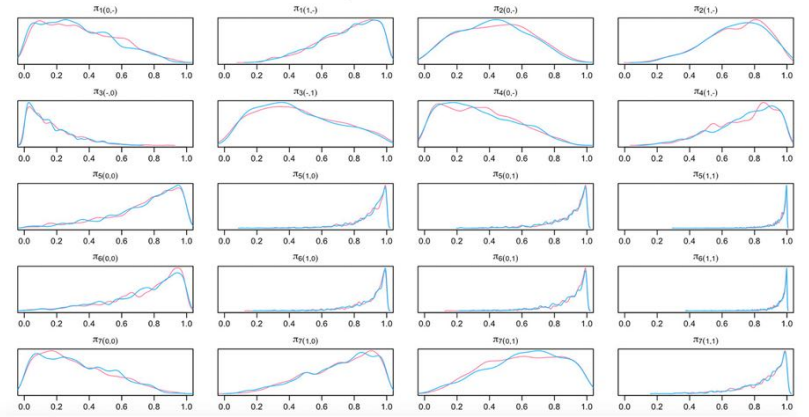
2019 (n = 37)



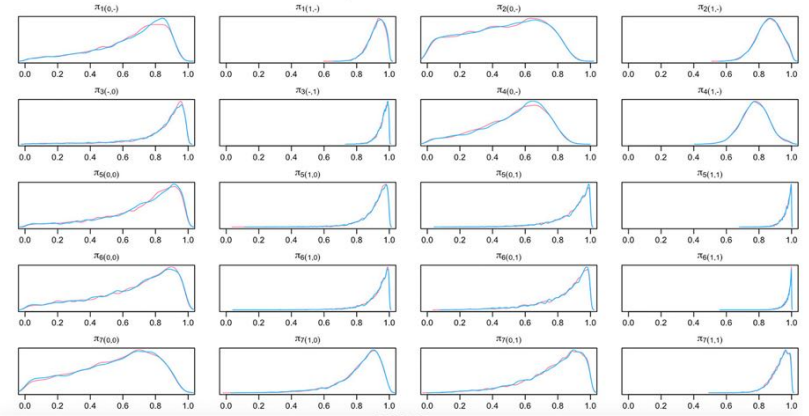
2018 (n = 74)



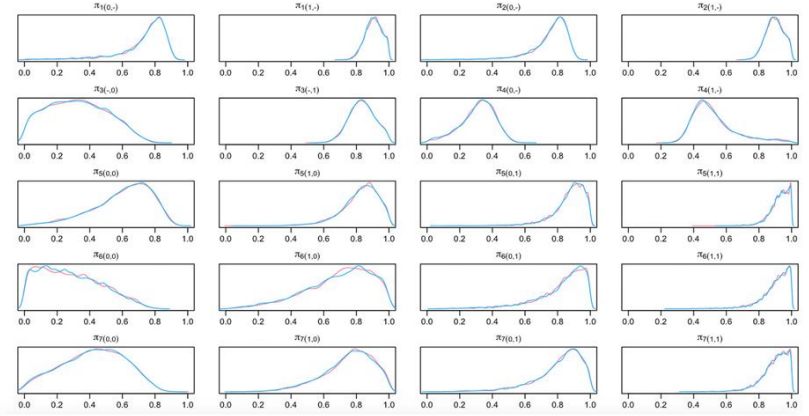
2020



2019



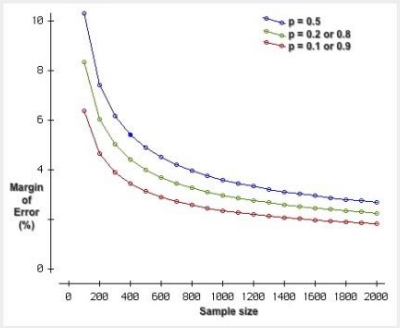
2018



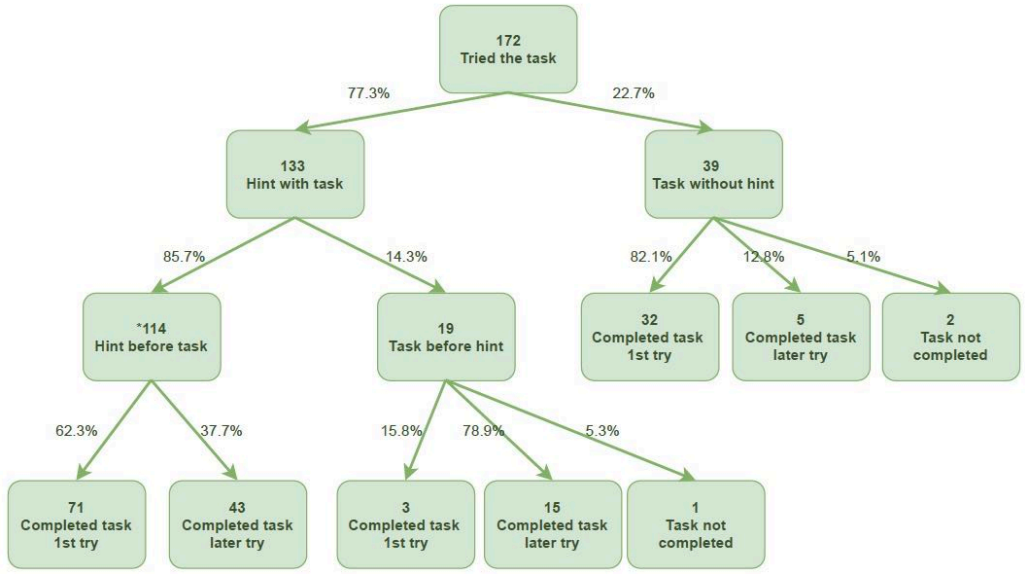
Tree Diagram on Hint Use - Super Bowl Poll

2. The margin of error for a **sample proportion** for a survey of 1000 people would be about ( ) %. Hint

Here is a graph for how the margin of error changes with sample size and the population proportion p:

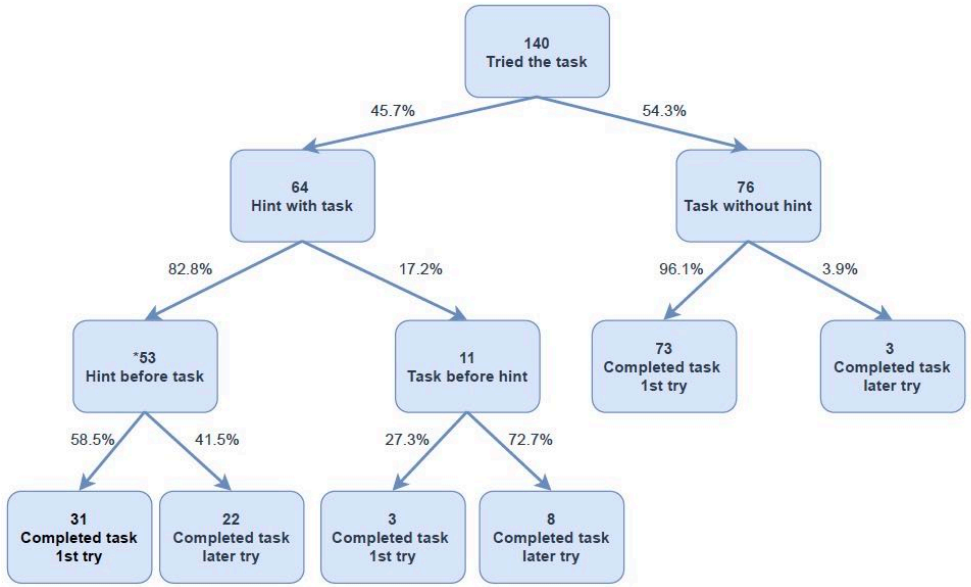


2019



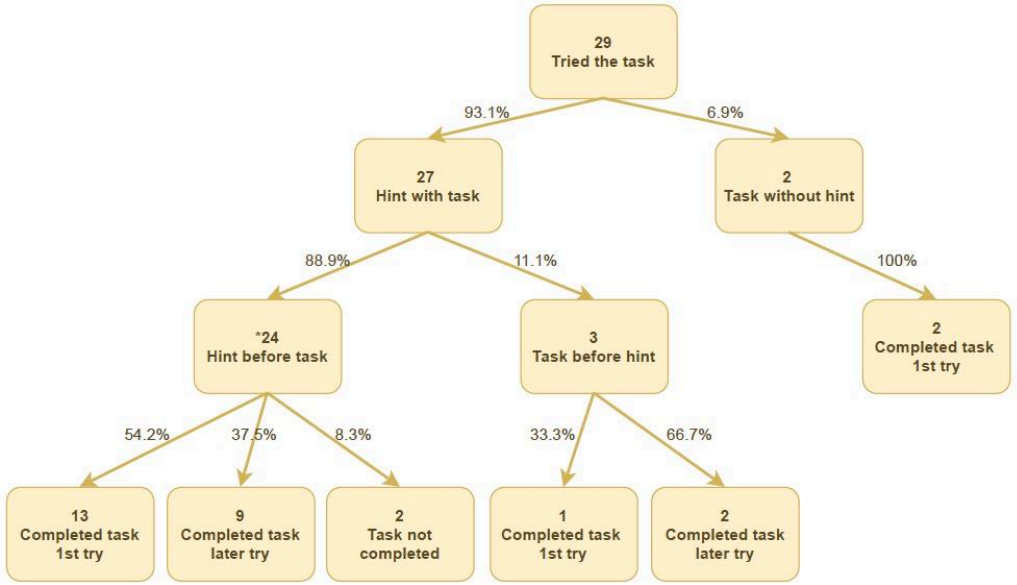
*: Samples that was used to train the Bayesian Network Model

2020



*: Samples that was used to train the Bayesian Network Model

2018



*: Samples that was used to train the Bayesian Network Model