Class 10

DATA1220-55, Fall 2024

Sarah E. Grabinski

2024-09-20

Homework 2

- Instructions (`homework2_instructions.pdf`), a Quarto markdown template (`homework2_template.qmd`), and an example HTML output (`homework2_example.html`) are available for download under Chapter 2 on the Modules page in Canvas.
- Video walk-through of Homework 2 under Tutorials on the Modules page in Canvas. Make sure you're caught up on the video walk-through of homework 1.
- Upload ***TWO*** (2) documents to Homework 2 on the Assignments page in Canvas by Friday 9/20/2024 by 6:00pm: `homework2_yourlastname.qmd` and `homework2_yourlastname.html`

Homework Hints

- *Read the instructions!* Some of the issues you're having are because you did not follow them correctly.
- *Please answer in complete sentences where possible!* I want you to practice effectively communicating data, and life is not a multiple choice question. I will be more clear about indicating this on future homework.
- Real world distributions are harder to describe than idealized theoretical distributions. *Combining visual and numeric summaries is more powerful than using either alone.*

Campuswire Hints

- *Turn on notifications.* Your question may have already been asked and answered. Campuswire can email you when there are new posts, so you can keep up with the discussion.
- *Be specific!* A detailed question is more likely to get a (useful) answer than a general question.
- *Include code & error messages.* It is much easier to troubleshoot "My document won't render. I've copy-pasted the error message and the lines of code where it breaks." than "My document won't render." *Click here for more info on how to ask good debugging questions.*

How can I get help with homework?

- **Read the textbook.** Many of you are asking for additional examples. Luckily, there are tons we didn't go over in the textbook.
- **Ask a question on our Campuswire class feed.** I'm only one person, and I may not be able to give you a prompt answer. However, the 20+ other people in the class might be able to.

I will try to keep an eye on Campuswire posts between 4-6pm before the homework is due, but I have other things going on and might miss something.

Last time... defining probability

- ***Probability:*** The proportion of times that a particular outcome would occur if we observed a random process an infinite number of times ($\mathrm{P}(\mathrm{Event} = \mathrm{A})$.
  - Ranges from 0 to 1 or 0% to 100%
  - $0 \leq \mathrm{probability} \leq 1$
  - Probability = Proportion
- ***Random process:*** you know which outcomes are possible (i.e. the **sample space**) but you don't know which outcome comes next

Last time... representing probability

- ***Sample space:*** all possible outcomes of a random process (S)
- ***Disjoint events:*** events that CANNOT occur at the same time (***mutually exclusive***)
- ***Complement:*** the complement of any event A which exists in sample space S is any outcome also in sample space S which is NOT A (A^C or A')
  - Complements are always disjoint
  - The probability of event A occuring OR the complement of event A occuring is always 1
- ***Non-disjoint events:*** events that CAN occur at the same time

Last time... calculating probabilities

Remember...

- $\mathrm{P}(S) = 1$
- $\mathrm{P}(S) = \mathrm{P}(A) + \mathrm{P}(A')$
- $\mathrm{P}(A) + \mathrm{P}(A') = 1$
- $\mathrm{P}(A') = 1 - \mathrm{P}(A)$

Last time... population probability

- ***Population Probability:*** the theoretical "true" probability of an outcome in the population of interest, the "ground truth" (p)

$$p = \frac{\text{count}(\text{events} = \text{A})}{\text{count}(\text{alleventsinsamplespace})}$$

Last time... sample probability

- ***Sample Probability:*** the probability of an outcome observed in a sample of size n from a population with probability p, an estimate of the population probability ($\hat{p}_n$)

$$\hat{p}_n = \frac{\text{count}(\text{observation} = \text{A})}{\text{count}(\text{observationsinsample})}$$

Last time... Law of Large Numbers

- How well the sample proportion $\hat{p}_n$ represents the population proportion p depends on the size of the denominator.
- As more observations are collected, the sample proportion $\hat{p}_n$ of a particular outcome approaches the population proportion p of that outcome.
- $\lim_{n\to\infty} \hat{p}_n = p$ (As $n \to \infty$, $\hat{p}_n \to p$)