PHP 2510 Principles of Biostatistics & Data Analysis

Week 2: Intro to R & Probability

This Week’s Plan

1. R & Rstudio
   - Data manipulation
2. Probability Games
   - Births
   - Baseball
   - Let’s Make a Deal
3. Contingency Tables
   - Diagnostic Tests
   - Dice
4. Q&A

OUTCOMES

After this week's classes, along with the required readings (SPEEGLE Chapter 1; CHIHARA Appendix A; SPEEGLE Chapter 2), you should be able to:

- Execute basic R commands for arithmetic, variable assignment, and installing packages
- Calculate the probability of complex events by applying the fundamental rules of probability
- Explain the concepts of marginal, joint, and conditional probability
- Solve for conditional probabilities like sensitivity, specificity, and positive predictive value in the context of diagnostic testing

Part 1

R – basic commands

Working with a neighbor, use R to:

- Pretend this data was collected last week and input it
- Merge the data with with the true ages: 14, 58, 37, 26, 80

Then:

- Calculate which picture had the most accurate guesses (on average)
- Calculate which picture had the guesses with the largest variance
- Does it seem like bias is associated with standard deviation? What does this mean?
- Calculate any other statistic you'd like. What does it tell you?

Discussion topic:

- What is a question you could ask that would require statistical *inference*?

pic	guess
1	10
1	15
1	18
2	55
2	70
2	49
2	65
3	33
3	55
4	40
4	50
4	57
5	70
5	77
5	79
5	80
5	81

R – basic commands cont.

Review example A.2 on pg 495 of CHIHARA

Now inspect the var() or sd() help pages in R

What do you notice?

Should you change your calculation from the previous slide?

Probability

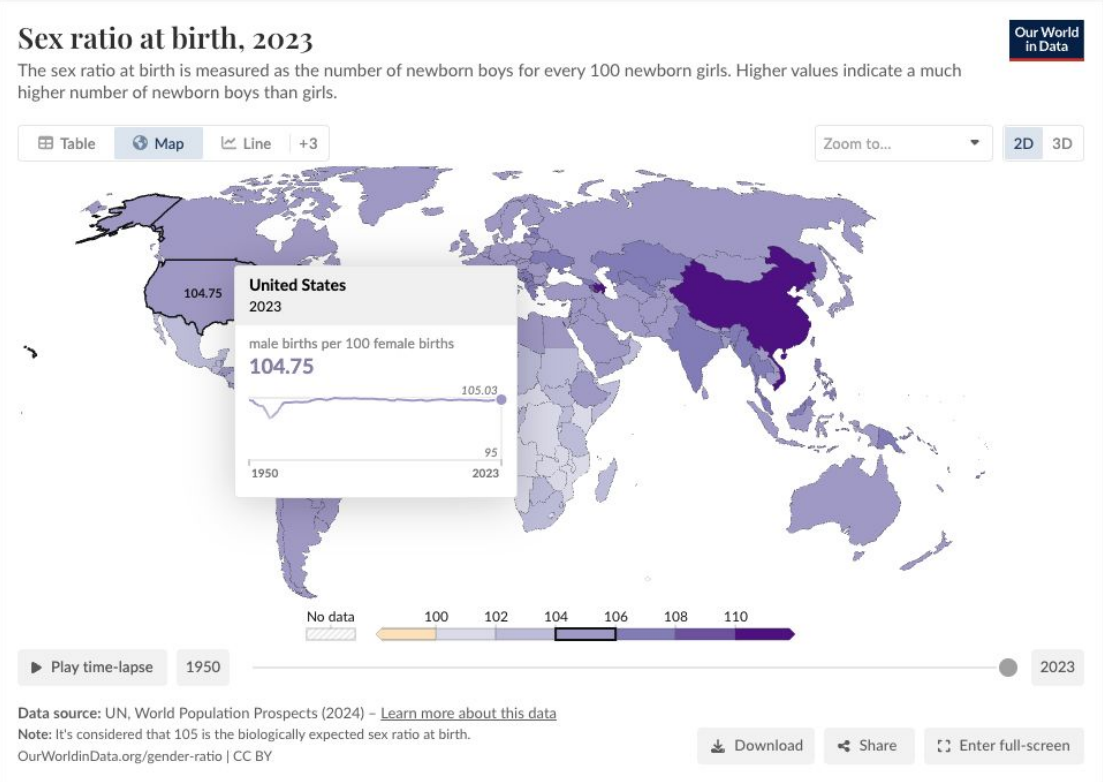
Part 2

Probability Games: Births

- Given a family has 4 children, what sequence is most likely: *bbbb, gggg, bgbg?* Calculate the probability of each, assuming equal probabilities of each sex for a single birth.

Part 2

Probability Games: Births



https://ourworldindata.org/grapher/sex-ratio-at-birth?mapSelect=~USA

https://www.npr.org/sections/health-shots/2015/03/30/396384911/why-are-more-baby-boys-born-than-girls/

Re-calculate the probability

#1

Part 2

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Degrees & Programs Admissions Faculty & Research

Home / News / Biological sex at birth isn't random, study finds

Biological sex at birth isn't random, study finds

By Karen Feldscher July 22, 2025

Share Post

The research described in this article was made possible in part by federal funding awarded to Harvard Chan School scientists in the interest of protecting and promoting health for all. The future of research like this is now in question due to the government's actions to terminate large numbers of grants and contracts and freeze funding for scientific inquiry and innovation across Harvard University.

A child's biological sex at birth may not be a 50-50 toss-up, according to a new study from Harvard T.H. Chan School of Public Health.

The study, published July 18 in the journal Science Advances, found that birth sex appears to be associated with maternal age, certain genes, and the sexes of older siblings.

"If you've had two girls or three girls and you're trying for a boy, you should know your odds are not 50-50. You're more likely than not to have another girl," said Jorge Chavarro, professor of nutrition and epidemiology and senior author of the study, in a July 18

The Washington Post

Your baby's sex isn't random. A study shows what could influence it.

A new study finds that the odds of having a boy or girl aren't 50-50. Instead, birth sex is associated with maternal age, specific genes and the sexes of older siblings.

July 18 at 2:00 PM 2025

Before a Brazilian soccer match, fans hold a gender reveal party in the stands for a girl who was to be named Luiza. (Gilson Lobo/AP)

By Sabrina Malhi

A baby's sex may not be up to mere chance.

A study published Friday in the journal

Today News+ Sports Audio Following

https://hsph.harvard.edu/news/biological-sex-at-birth-isnt-random-study-finds/

Part 2

Probability Games: Births cont.

The likelihood of having identical twins is 1/300; fraternal twins 1/125

- What is the probability of having two boys in a single birth event?
- Given a twin birth event, what is the probability that both are boys?

Probability Games: Baseball

The world series is a best-of-7 competition. What's the likelihood of requiring exactly n = {4, 5, 6, 7} games to declare a champion?

start with the assumption of a 50–50 chance between teams and independence between games

Part 2

Probability Games: Baseball

Number of Games	Theoretical	Actual
4	12.5%	19/116 = 16.4%
5	25%	30/116 = 26.9%
6	31.25%	27/116 = 23.3%
7	31.25%	40/116 = 34.5%

How would we redo the calculation if we instead assumed a 55/45 split in winning likelihood between teams?

Probability Games: Notecards

We will be working with sets of 3 notecards:

- Write “A” on both sides of a notecard
- Write “B” on both sides of a notecard
- Write “A” on one side, and “B” on the other side, of a note card

We will randomly grab a card and show just one side of it.

We see the letter “A”. What is the likelihood the other side also has an “A”?

Part 2

Contingency Tables

- Tabulates multiple categorical variables
- Common application: 2×2 tables for diagnostic tests

- Let's go over some key terms:
  a. True Positive (TP)
  b. True Negative (TN)
  c. False Positive (FP)
  d. False Negative (FN)
  e. Prevalence
  f. Sensitivity (also called TPR or recall)
  g. Specificity (also called TNR)
  h. False Positive Rate (FPR)
  i. False Negative Rate (FNR)
  j. Positive Predictive Value (PPV; also called precision)

		DISEASE STATUS	
		+	-
TEST RESULT	+	9	90
	-	1	4900

- How much evidence do we get from a positive test?
- What happens as the disease gets more rare?

Part 3

Contingency Tables

- Come up with your own (realistic!) 2×2 table
  a. Think about your disease prevalence
  b. Think about your test performance
- Swap with your neighbor and ask them to calculate the following
  a. Sensitivity
  b. Specificity
  c. PPV
- Check their work

		DISEASE STATUS	
		+	-
TEST RESULT	+	<>	<>
	-	<>	<>