# 1010 Consumer Statistics and Probability

Here are portions of questions from class to help you with your notes or later practice. The wording and ordering may change and we may not have time to cover all of them. Here we actively practice concepts, computational strategies, critical & creative thinking, and communication. Making mistakes is integral to the learning process and enriches our understanding as we extend content and clear up misconceptions.

- **Think** about a possible answer(s) on your own.
- **Pair up**: discuss your thoughts in a group. We may reorganize groups at times.
- Prepare to **share** from your group's discussion.
- Use any lag to **review** related concepts and examples and **add** to your notes.

Appalachian's General Education Program prepares students to employ various modes of communication. Successful communicators interact effectively with people of both similar and different experiences and values and in this class you will practice oral and written communication during class.

## probability

- What is the probability of NOT rolling a 6 on a dice?

- If the probability of a person being left-handed is $\frac{1}{10}$, and the probability of being blue-eyed is $\frac{1}{3}$, then what is the probability of being left-handed and blue-eyed (assuming these are independent of each other)?

- What is the probability of two people having the same birthday?

- What are some factors that may invalidate independency/confounding variables for the birthdays?

- Based on the decision matrix, would you have selected Friend or Foe? Explain your reasoning.

- What is the expected value of grades in a class where engagement is 20%, exams are 50%, and projects are 30% if the current averages are 85, 90, and 70, respectively?

- "The WORLD FACTBOOK," the CIA lists Eswatini (a small landlocked country in southern Africa formerly known as Swaziland) as having the highest "percentage of adults (aged 15–49) living with HIV/AIDS" (`https://www.cia.gov/the-world-factbook/field/hiv-aids-adult-prevalence-rate/country-comparison`, last updated Oct 2021). The total population is 1,113,276 and the percentage of people living with HIV is 26.8% Calculate how many people are HIV negative (round to the nearest whole person)?

- Compute the probability of getting exactly 3 heads if you toss a fair coin 4 times Assume independence.

- Appalachian has approximately 20,641 students (Sep 2021 email from Chancellor). The following table lists theoretical numbers of students registered by the number of courses they are taking. First fill in the table by computing the probabilities—number of registered / 20,641. Round to 2 decimals:

| number of courses | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| number registered | 206 | 619 | 2683 | 5160 | 8050 | 3509 | 413 |
| probability | .01 | | | | | | |

  Next, compute the expected value of the number of courses by taking a weighted average of the number of courses (in row 1) and the probabilities (in row 3), such as $.01 \cdot 1 + ...$

- Even though these are theoretical registration numbers, do you think this reasonably approximates reality at Appalachian? Explain your reasoning.

- Compute the approximate probability to find a shared birthday in a group of 4 people? Assume independence and exclude February 29th.

- Based on the readings from *The Heart of Mathematics* is it likely that two people in our class have the same birthday?

**data analysis**

- What does the application of Benford's Law to the country populations data set show us?

- Share something you found interesting or surprising from the country populations.

- Is it appropriate for scientists and mathematicians today to use data obtained under what we agree were morally reprehensible conditions, like the syphilis experiment in Tuskegee, Alabama or many experiments in Nazi Germany? **Answer all of the following**:
  a) First write down what you think is the strongest argument from the "yes" side.
  b) Next write down what you think is the strongest argument from the "no" side.
  c) Lastly, what do you think? [select from yes, only in certain circumstances, or no]

- Kat is making measurements in lab and is confident that they have set it up properly. When Kat tries to do the required calculations to verify the formulas in the book, the data seems wrong. So Kat does the mathematical calculations to determine what a correct set of data would be and simply changes the data to match the calculations. Is this problematic?

- A research group wants to study the effectiveness of a quercetin supplement, and has contacts at the Watauga County Detention Center, Rikers Island in New York, and Silivri Penitentiaries Campus in Turkey. Quercetin is a flavonoid found in fruits and vegetables that is a strong antioxidant. Do you think Appalachian State University's Institutional Research Board (IRB) will approve such research?

- A realtor wants to advertise how inexpensive it is to live in an area. Should they use the mean or median. Why?

- Here is a data set that measures population growth rates in the US from 1910–1919:

| year | 1910 | 1911 | 1912 | 1913 | 1914 | 1915 | 1916 | 1917 | 1918 | 1919 | 1920 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| population growth | 2.1% | 1.56% | 1.56% | 1.96% | 1.92% | 1.44% | 1.4 % | 1.27 % | −.06% | 1.26 % | 1.85% |

Use a scale balancing idea to analyze how the rate in 1918 would impact the mean/average:

- What happened in real-life in 1918 that led to this rate?

- Here on the left is Nielsen ratings (roughly represents the percentage of households tuned in). Use the boxplots to award "best network" with ABC (top boxplot), CBS (middle), NBC (bottom)



Reaction time with cell phone usage
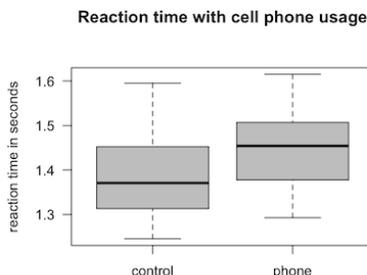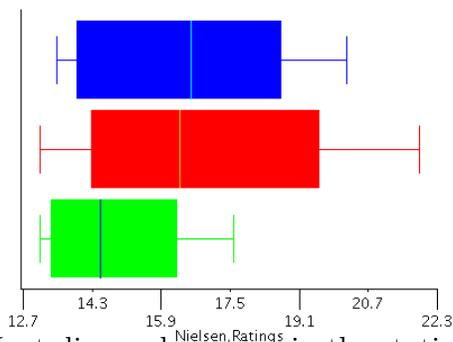
Picture: http://digitheadslabnotebook.blogspot.com/2010_06_01_archive.html

- Next discuss how to spin the statistics positively (but still truthfully) for each of the other networks you didn't select: Here's good news, we are the best network because... In one of them, it may be challenging to say something positive but truthful, but think creatively!

- Which is true for CBS, the middle boxplot when comparing the mean and median?

- With regard to the 1936 Landon and Roosevelt election and the *Literary Digest* poll that predicted the winner as Landon:
  a) the sample size was not large enough
  b) the sample was not diverse enough
  c) Landon would have won—but Roosevelt's win was due to a last minute change in sentiment
  d) other

- Which of the following in *The Heart of Mathematics* readings did you find most compelling?
  a) inconsistencies in scaling can lead to false interpretations
  b) the average American or average salary of Lakeside school can be very misleading. Half of the people are not necessarily below average.
  c) sampling pitfalls such as convenience sampling, voluntary responses, and asking unclear or misleading questions
  d) other

## inferences and case studies

- Do hours without sleep the night before an exam predict midterm errors versus midterm errors causing hours without sleep the night before an exam?

- Why is it plausible that there is a pattern to spirals in nature? Where else do we find the same pattern?

- In the above right picture we see two side-by-side boxplots on reaction times. The first boxplot is a control group and the second are people using their cell phones. What can you say about the data from the median to Q3, the third quartile, of the reaction times?

- Would you have been drafted for Vietnam in the 1969? Is there anyone in the class with the same birthday?

- You worked with data from a study comparing the number of years of experience of New York City police to the number of tickets they give per week. What does the $y$-intercept mean when $x$=#years experience and $y$=#tickets police give out per week and the best fit line is $y = -2.932x + 55.038$?

- Using the same data as the last question, if $r = -.86$ so that $r^2 = 73.96\%$, this tells us that

- The $r^2$ value is strong but the line $y = -2.932(25) + 55.038$ predicts that the police *receives* tickets after 25 years, rather than give them out. Resolve the apparent conflict.

- If the $r^2$ value was 100% for a stock in the stock market, would we be assured to make money by using the best fit line to predict the future performance? Why or why not?

- As a researcher, was it ethical to remove the armspan/height points that I eliminated from the class data and keep the remaining points?

- In the egg bungee experiment the similarity of the rubber bands led to an almost constant slope of $\frac{\Delta y}{\Delta x} = \frac{\Delta \text{distance dropped}}{\Delta \text{rubber bands}}$

- How many rubber bands did you use?

- Even without $r^2$ in front of us, we can visually inspect and categorize relationships. From Gapminder did you think that recent Income is at least a statistically moderate predictor of Life expectancy (i.e. moderate/strong versus a no/weak correlation)?

- What did you find interesting or surprising, or that you had a question on, in the Gapminder exploration?

- On April 4, 2017, Gallup published poll results under the headline, "Affordable Care Act Gains Majority Approval for First Time." For 2012, Gallup gives a 95% confident margin of error of 3%, which had 48% of the sample "approved." For 2017, 55% of the sample "approved" and the margin of error was plus or minus 4% for the 95% confidence level.

  If this was a simple random sample of the 1023 adults in 2017, what would the conservative 95% confidence interval margin of error be?

  Is the headline "Affordable Care Act Gains Majority Approval for First Time" statistically valid? Why or why not?

- Assume little to no bias and truly a random sample. If a polling company conducted 100 such polls with a 95% confidence interval, then about how many of them are likely to include the true population percentage?

- Is there any way to know which intervals from the 100 polls contain the true percentage and which ones don't?

- Gallup typically targets both landline and cellphone users in its polls. Are there any voices that are left out?

- Search for the percentage of people in the US that have a phone? List your response and source.

- Search for how many US adults do NOT use the internet? List your response and your source.

- How should we interpret the margin of error if the sample is biased?

- For a simple random sample at the 95% confidence level, what sample size would be required to achieve a 1% margin of error (use the conservative computation)?

- In which of the following examples will the margin of error be the smallest? Assume each refers to a random sample that is not biased for a 95% confidence interval.
    a) a sample of $n = 400$ from a population of 50,000
    b) a sample of $n = 1000$ from a population of 10 million
    c) a sample of $n = 2500$ from a population of 200 million

- What was the main point of Fisher's experiment on the Lady Tasting Tea?

- In *The Heart of Mathematics*, on pp. 663–664 you chose one to complete: amazing stats #15 or internet askew #18. Share as below:

If you selected #15: What is the question you asked? What is the sample and the sample size? What is the dubious conclusion? How could one reduce bias in this instance?

If you selected #18: What is the title or topic? Summarize the graph. Is there distortion or bias? Explain. How could one reduce bias in this instance?

- Which did you find most compelling about the "price of life" readings on deciding personal and private policy?

- What is the problem with testing the entire US population for HIV?

- A retail version of OraQuick costs $30 and gives results in about 20 to 40 minutes. About half of US states test every inmate for HIV on admission or during incarceration. Voluntary testing programs are often ineffective because prisoners do not want to admit to high-risk behaviors. Given this and your statistical analyses from the case studies worksheet, consider whether we should support mandatory HIV testing of newly admitted prison inmates, as you **respond to all of these**:

a) Summarize what you believe is the strongest argument from the "yes" side
b) Summarize what you believe is the strongest argument from the "no" side
c) What do you think—yes or no?

- What can we do with multiple regressions? Use both pieces of information on the slides to come up with a prediction of the person's height.

- Discuss whether you or someone you know have ever experienced something similar to stereotype threat as part of some kind of group (for example, gender, race, sibling, hair color, athlete, southern accent...) where external expectations from someone else (teacher, society, parents, friends...) affected your performance.

- Mental Rotations Test (MRT)–Is Time on your Side?

- How can we mediate the impact of stereotype threat?

- What your Gallup Poll which you analyzed in the lab (and has a Survey Methods section that includes a margin of sampling error %).

- Name at least one item from your Gallup Poll that you found interesting or surprising, or that you had a question on.

- What are some recent examples of confidence intervals and statistically invalid Gallup statements (where they are possibly invalidated by taking into account the margin of error)

- General Electric (GE) funded the National Research Council (NRC) of the National Academy of Sciences to study how worker productivity is tied to lighting during a 1924 illumination study. What do you think the NRC concluded?

- What are possible underlying variables/confounding issues that could explain why birds fly south in the winter?

- Can you think of experiments to understand why birds fly south in the winter? What makes good experiments on this issue difficult?

- Studies have found that college students who used marijuana regularly had impaired skills related to attention, memory, and learning 24 hours after the last use. What are some confounding issues?

**review**

- List an instance from the statistics and probability segment where the theme of local to global played a role. What was local? global?

- Reflect on this segment to discuss when are we reasonably convinced of truth and/or causation?

- Review and write down a few people and their impacts from our statistics and probability segment.

- Review and write down a few ideas relating to what mathematics is and has to offer from our statistics and probability segment.

- Write down as many items, scenarios, equations, visualizations, and more that we covered and are related to each of the following (i.e. use these as a way to review what we covered):
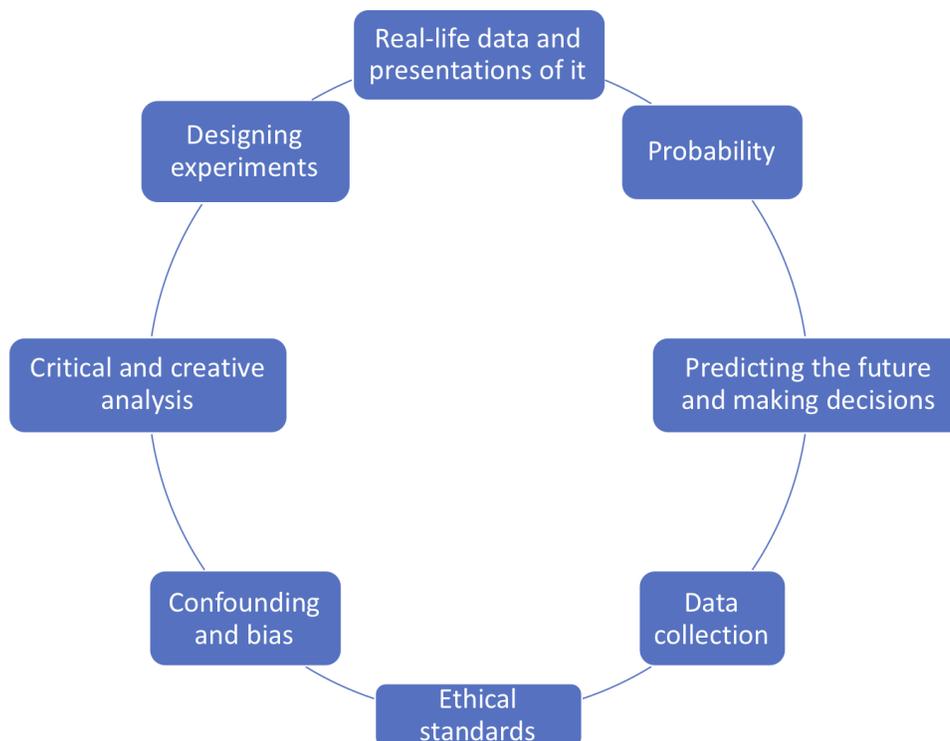
Problem solving at the heart of mathematics: visualization, generalization, making decisions