Dear Students,

Here's a list of some things I have in mind as I prepare the upcoming exam. Keywords and concepts and so on.

I hope it helps!

Best,
Felipe A. Ramírez

Contents

1 Exam 1 1

2 Exam 2 3

1 Exam 1

- Definition of *statistics*, the field of study
- What is a population? A variable? Data? Sample?
- Different kinds of variables. Quantivative vs. categorical.
- How do we communicate descriptions of data? Averages, dot plots, percentiles, etc.
- Parameters and statistics. What are these, and how are they related?
- Location, spread, shape. What do these terms refer to?
- Commonly used notations: $x_1, x_2, \ldots, x_n$ for a data set from a sample of size n, $\bar{x}$ for mean. Can you write the mathematical formula for $\bar{x}$ using these notations?
- What about formula for range?
- What are quartiles? Five-number summary?
- Sample variance, sample standard deviation. What are they? Are you comfortable computing them by hand for small data sets?
- Population variance (σ^2) and population standard deviation (σ). What's the difference between these parameters and the statistics?
- How to study the relationship between two quantitative variables.
  - What kind of data does one gather?

  - How does one depict those data?
  - How does one describe what one sees?

- Regression analysis, regression equation, regression line. What are these?
- Least squares method. What's going on here?
- The correlation coefficient, r, and its square, r^2. What do they tell us? How do you compute them? Recall the interpretation of r^2 as the proportion of variation in y that is explained by x. What does that mean?
- General advice: This list has already included a few things that, in actual practice, one would compute with the help of technology. (Mean, standard deviation, least squares, correlation coefficient, etc.) On this exam, I am more interested in whether you understand what these things actually are—their mathematical formulas, and what they tell us about data. I might ask you to compute them for small examples.
- How to study the relationship between two categorical variables.
  - What kind of data does one gather?
  - How does one depict that data?
  - How does one describe what one sees?
- Sampling. What is meant by "representative sample"? What is a simple random sample? Stratified sample? Clustered?
- Are you comfortable with the basics of probability theory? How do you interpret the statement "an event has probability p of occurring"?
- Know terminology: sample space, simple event, compount event, event, complementary event, mutually exclusive (or disjoint) events, independent events, conditional probability, etc. I may ask you to define some of these terms, but even if I don't, I will certainly *use* them with the expectation that you know what they mean.
- Are you comfortable with notational conventions? For example, the use of letters, like A, to denote events? Can you translate a mathematical expression into a spoken sentence? For example, how would you translate

  $$\mathrm{P}(\mathrm{A}) = 0.2$$

  into English? ("The probability of the event A is 0.2.") Can you translate

  $$\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = 1/3, \qquad \mathrm{P}(\mathrm{A} \cup \mathrm{B}) = 1/7, \qquad \mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B}),$$

  and other similar notational expressions into English?
- Can you translate natural English sentences into mathematical expressions such as the ones above?

- Know basic rules of probability:
  - $0 \leqslant P(A) \leqslant 1$ for any event A.
  - Suppose A and B are disjoint events. What is $P(A \cap B)$ in terms of the probabilities of A and B? What is $P(A \cup B)$?
  - Suppose A and B are independent events. What is $P(A \cap B)$? What is $P(A \mid B)$?
  - Suppose A and B are events. What is $P(A \cup B)$ and how does this relate to the answer when A and B are mutually exclusive?
  - Bayes' Rule. What is it? What's it for?
  - ...and so on.

  There are several of these. Are you comfortable with them?
- I suggest you practice "tree diagrams" and "hypothetical tables." These devices make it easier to understand certain unintuitive results from probability theory.

2 Exam 2

- Are you comfortable with the material from the previous exam? Though it is not specifically tested, it is now assumed as background knowledge.
- Terminology: random variable, discrete/continuous random variable, probability distribution function, probability density function, cumulative distribution function, expected value of a random variable, variance of a random variable, standard deviation of a random variable.
- What is a random variable? Can you give examples?
- What is the difference between discrete random variables and a continuous random variables? Can you give examples of each?
- Probability distribution function vs probability density function?
- What's the difference between pdf and cdf? How are they related? Can you give the pdf/cdf of basic random variables?
- Can you define expected value and give a general formula? $(E(X) = ?)$ What about standard deviation? What do these numbers tell us about a random variable?
- Two very simple random variables that come up all the time: 1) X is the outcome of a die roll. 2) X is the outcome of a coin flip $(\text{heads} = 1, \text{tails} = 0)$. Can you analyze these? Their pdfs, cdfs, expected values, etc.?
- Generalizing that last one, make sure you're comfortable with binomial random variables. What they are, how they're distributed, what they're good for, etc.

- What is the connection between binomial random variables and normal distributions?
- What is a normal distribution? What is the standard normal distribution?
- What is the empirical rule?
- Inference: What are the roles of (population) parameters, samples, sample estimates, sampling distributions, confidence intervals, standard errors, etc, etc.?
- "The Big Five Parameters." What are they? Can you come up with questions exemplifying each?
- Are you comfortable with the analysis of the ones we have discussed?
- What is a sampling distribution?
- Of what use is standard error? How is it different from standard deviation? When/why do we use it instead of standard deviation?
- What is a confidence interval, and how do we interpret "confidence"?
- Are you comfortable finding confidence intervals in the estimation of proportions?
- Are you comfortable with the methods for estimating the difference between two population proportions? What changes?