Transition into Probability

2019-02-21

What have you learned?

> Manipulating, visualizing, generating data
> Contingency Tables summarizing
> Sampling theory

* Prior to the midterm,

> continuous
> Discrete

} To be clear, depending on the abstract nature of a variable, there is a natural answer.

DATA TYPES.

Review

BMI Categories — ordinal, categorical

BMI Values/#'s — continuous, Quantitative

* You also learned about STUDY DESIGN and SAMPLING METHODS. We design studies and sample in certain ways all to properly capture information about the group we are interested in.

Population → sample

We can take many many samples to build what is called a SAMPLING DISTRIBUTION.

(Talked about in Lab 3)

Estimates about POPULATION PARAMETERS will be more precise with "large enough" n and number of samples taken total.

Really quick, analogy: Population : Parameter
sample : statistic

μ is population mean.
$\bar{x}$ is the sample mean.

And definition:
X (capital) is a RANDOM VARIABLE, abstract placeholder

x (lowercase) is what you observe (realization)

As motivation, we are aiming to go down the yellow brick road of statistics.

Build up language of chance (Probability) → Use the language to formalize probability theory (understand how samples can behave as $n \to \infty$) → Use tools based off of the theory developed (statistical Testing)

Recall that we know how to explore and visualize (summarize) data by now. This is called EXPLORATORY DATA ANALYSIS. After knowing you have quality data, EDA is the step you take to motivate some sort of statistical testing.

P P D A C

YOU ARE HERE.

(*) Now, we are pivoting into the world of probability. Welcome. We have rules.

RULE $0 \leq P(A) \leq 1$ where (A) is any event

Imagine that A = The event you ate today.
A^c = The event you have not

These two events belong to a SAMPLE SPACE.

$$S = \{A, A^c\}$$

They are also DISJOINT/MUTUALLY EXCLUSIVE ! Because you cannot have done A and A^c.

$$P(A \cap B) = 0$$
$$\begin{cases} P(A|B) = 0 \\ P(B|0) = 0 \end{cases}$$

They are also DEPENDENT.
$P(A) \neq P(A|B)$

RULE $P(A) + P(A^c) = 1$

But wait... what is independence? In math:

INDEPENDENCE
$$P(A) = P(A|B)$$

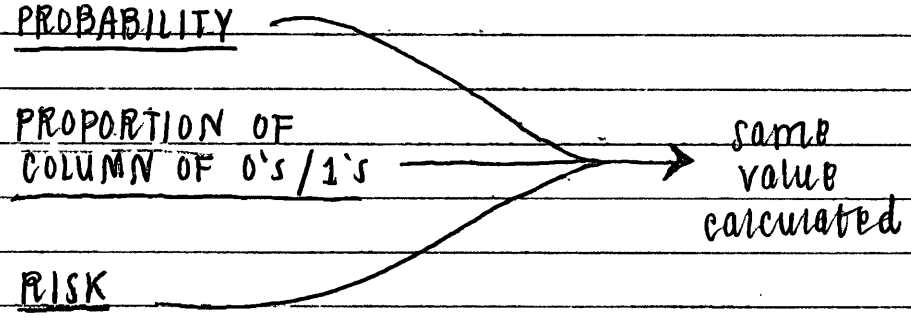
(*) (This may be a bit of a reach...)
In words? Imagine:
A = The event you are self-actualized, happy, and stable
B = The event you are in love

WHETHER OR NOT YOU ARE IN LOVE OR NOT, the chance you are (A) is the same!

⊛ These notes skip over density curves. We will return to those when we start seeing limit theorems.

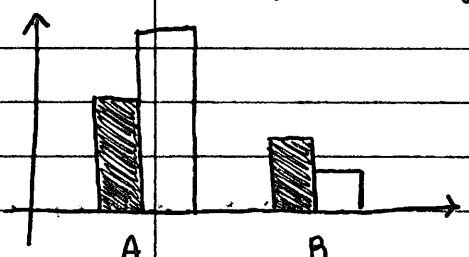
Probability between 2 events can be BEAUTIFULLY visualized with venn diagrams.
> See Sarah Johnson's slides! Wow!

⊛ When you read about probability,



⊛ CATEGORICAL DATA are represented in contingency tables. They show counts GIVEN certain categorical characteristics.

Recall dodged histograms can be displayed like:



	A	B	Total
X			
Y			
Total			

From these tables, we can use counts to calculate CONDITIONAL PROBABILITIES.

(Lab 4)