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	BMI values/#'s
ordinal	continuous
categorical	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.

Population : Parameter

(*) Really quick, analogy: sample : statistic

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

And definition:

X (capital) is a <u>RANDOM VARIABLE</u>, 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) $\Rightarrow$ use the language to formalize probability theory (understand how samples can behave as $n \to \infty$) $\Rightarrow$ use tools based off of the theory developed (statistical Testing)

Recall that we know how to explore (summarize) and visualize data by now. This is called <u>EXPLORATORY DATA ANALYSIS</u>. 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.

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

$$\left\{ \begin{array}{l} P(A \cap B) = 0 \\ P(A|B) = 0 \\ P(B|A) = 0 \end{array} \right.$$

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

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

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

This is going to be a stretch. :P

Let (A) be a PERSON who is super strong.
Let (B) be a person she's interested in.

When (B) is not there, i.e. when (A) is just P(A) she has value.
When (B) is there, i.e. when we have P(AlB) she has the same value.

REGARDLESS of whether (B) is there, (A) is the same!


DESTINY'S CHILD
INDEPENDENT WOMEN
© 2000 Sony Music Entertainment Inc./44K 79493

* 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,

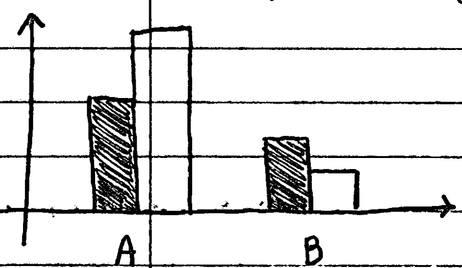
PROBABILITY

PROPORTION OF COLUMN OF 0's / 1's → same value calculated

RISK

* 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)