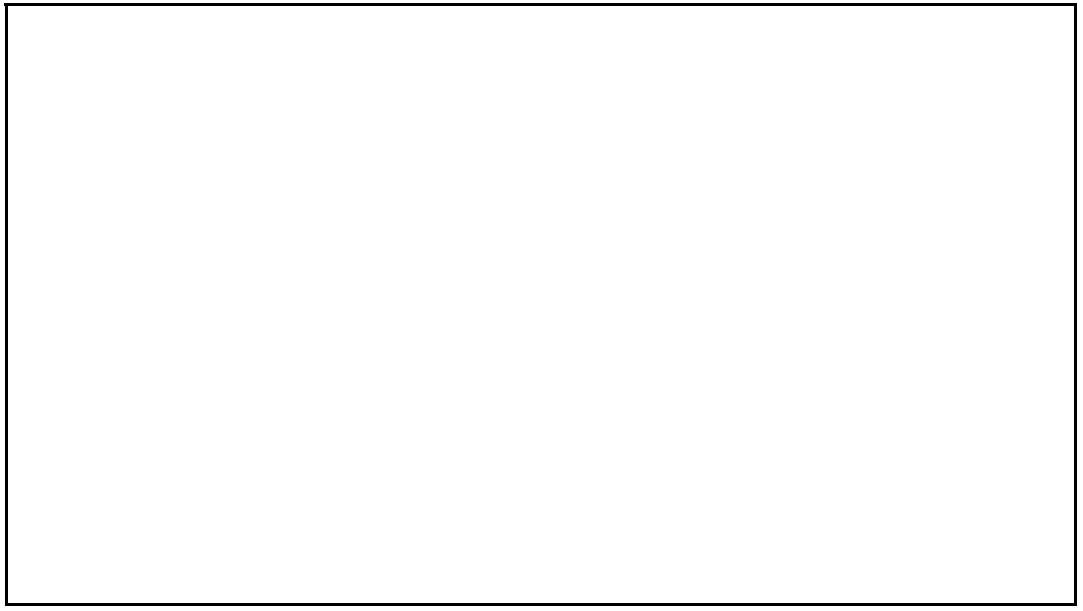
INDEPENDENCE & CONDITIONAL PROBABILITY



	6:19 / 6:19	1.0x				

SPEAKER: MICHAEL J. MAHOMETA, Ph.D.

We know that using a contingency table is

the first step in determining if there is

a relationship between two categorical variables.

And then, that the Row or Column percentages

- the conditional distributions - will then

help us to tell the story of the contingency table.

It's our job as data scientists or examiners

to determine which of the conditional distributions to use,

which story we want to tell from our contingency table.

Help

We also want to examine and answer our original question of interest

to determine if there is in fact a relationship between two

categorical variables.

And to do that, we need to look at both our marginal distributions

and our conditional distributions together.

To help remind us, here is the contingency table

looking at the relationship between Survivorship and Class on the Titanic.

And here are the two possible conditional distributions we could use.

One for the Row percentages - the conditional distribution

of Class based on Survivorship.

And the other for the Column percentages -

the conditional distribution of Survivorship based on Class.

So how do we know if there is a relationship

between Class membership and Survivorship?

Well, it comes down to examining our conditional distribution

to our marginal distribution.

How do we do that?

We start with asking - from our contingency table -

what the variable is that represents the outcome.

Much like in correlation when we determined what the outcome

variable was (or the dependent variable),

so we also need to come up with an outcome variable

here for our contingency table.

So which variable is it?

Survivorship or Class membership?

I think it's pretty obvious that the final variable - the only variable that

makes sense as an outcome variable - is in fact Survivorship.

Now determining this variable - this outcome variable - may not be obvious

all the time for every contingency table.

It's up to us to determine which outcome variable

or which variable makes the most sense for the story that we want to tell.

Here, survivorship is the variable of interest

for me.

Now that we have determined what are variable

of interest is from our contingency table,

we need to look at how this outcome variable behaves

in its normal state of affairs.

What do I mean by this?

I mean, how does our outcome variable look

without any consideration for our other categorical variable?

How does Survivorship look in terms of its distribution of outcomes

without any consideration of Class membership?

To answer this fundamental question, we turn to our marginal distributions.

Our marginal distribution out here for Survivorship

shows us what we should expect the normal state of affairs

to be of surviving or not surviving the Titanic catastrophe

if we didn't consider at all class membership.

To discover if there's a relationship between our two categorical variables,

we want to see whether or not this marginal distribution holds up

when we compare it to the conditional distribution of Survivorship

- to the conditional distribution of our outcome variable

or our variable of interest.

When we compare our marginal distribution of Survivorship

to our conditional distribution of Survivorship,

we can see that things don't actually match.

Just looking within the first class level,

we see that our conditional probability of surviving

doesn't match our marginal probability of survival.

It doesn't even come close.

If we didn't take into account Class, the rate of survival from the Titanic

would be about 32%.

But, if we take into account Class level, the first class

level in particular, we can see that the likelihood

of surviving if you were in the first class of passengers is about 62%.

The same disparity of the conditional and

marginal probabilities

hold true for the second class and the third class and even the crew.

Our conditional probability of surviving within any of our classes

doesn't actually match our marginal probability -

the normal state of affairs if we didn't take class into consideration.

And that's how we determine if there's a relationship between our two

categorical variables.

We determine our variable of interest - our outcome variable in our contingency

table.

We determine our marginal distributions and our conditional distributions

for that variable of interest.

And then we compare those two distribution probabilities.

If they don't match one another - or come close to matching one another

- then we have a relationship between our two categorical variables.

If we fail this mathematical statement that the probability of A given B

is equal to the probability of A, then we actually

have a relationship between the two categorical variables.

So is there a relationship between Class level and Survivorship on the Titanic?

Absolutely.

Your survival on the Titanic was is some way related to your Class membership.

You might say that the probability of surviving

was driven by Class membership.

Some classes of passengers had a higher probability

of surviving than the expected 32%.

But just saying that and showing the disparity of probabilities

may not really hit home the answer to our question of interest.

To do that, we'll really need to "see" what's

going on in our contingency table with a good visualization.

Help

Comprehension Check

Below is a contingency table showing data from a University of Texas Southwestern Medical Center study on Hepatitis C.

	Tattoo in Commercial Parlor	Tattoo Done Elsewhere	No Tattoo	Total
Has Hepatitis C	17	8	18	43
Does Not Have Hepatitis C	35	53	495	583
Total	52	61	513	626

(6/6 points)

1) How many simple events (outcomes) were possible for participants in this study?

- Three
- Six ✔
- Nine
- Twelve

2) What was the total number of participants in this study?

- 513
- 583
- 626 ✔
- 1113

Help

3) What was the marginal distribution for Hepatitis status in this study?

- 25 had tattoos, and 18 did not.
- 626 participants had Hepatitis.
- Of the 513 participants with no tattoo, most did not have hepatitis.
- 43 had Hepatitis; 583 did not have Hepatitis. ✔

4) Overall, what percentage of participants had a tattoo? (*Round to 1 decimal place and do not include % sign.*)

18.1

18.1

Answer: 18.1

5) What percentage of those participants with Hepatitis C had a tattoo done in a commercial parlor? (*Round to 1 decimal and do not include % sign*.)

39.5

39.5

Answer: 39.5

Help

6) What percentage of those who had a tattoo done in a commercial parlor have Hepatitis C? (*Round to 1 decimal and do not include % sign*.)

32.7

32.7

Answer: 32.7

Check Hide Answer

Calculate the probability that a randomly selected participant from the study would have Hepatitis C:

$$P\left(Hepatitis\right) = \frac{outcomes\ with\ Hepatitis}{total\ \ outcomes\ in\ sample\ space} = \frac{A}{B} = C$$

(4/4 points)

7) What is the value of **A**?

43

43

Answer: 43

8) What is the value of **B**?

626

626

Answer: 626

9) What is the value of **C**, the probability of randomly selecting a participant with Hepatitis? (*Round to 3 decimal places*.)

0.069

0.069

Answer: 0.069

10) In general, what must be true of P(A)?

- (•) It must be between the values of 0 and 1, inclusive. ✔
- ( ) It must be calculated from real data; it cannot be determined theoretically.
- ( ) It must include all possible outcomes of an experiment.
- ( ) It must not be greater than P(B).

Help

Check Hide Answer

Help

EdX offers interactive online classes and MOOCs from the world's best universities. Online courses from MITx, HarvardX, BerkeleyX, UTx and many other universities. Topics include biology, business, chemistry, computer science, economics, finance, electronics, engineering, food and nutrition, history, humanities, law, literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

Terms of Service and Honor Code

Privacy Policy (Revised 4/16/2014)

About edX

About

News

Contact

FAQ

edX Blog

Donate to edX

Jobs at edX

Follow Us

Twitter

Facebook

Meetup

LinkedIn

Google+