UTAustinX: UT.7.01x Foundations of Data Analysis

KarenWest ▾

Courseware Course Info Discussion Syllabus Download R and RStudio R Tutorials Readings Contact Us

Progress Office Hours Community

GRAPHING THE CONTINGENCY TABLE

6:54 / 6:54 1.0x

SPEAKER: MICHAEL J. MAHOMETA, Ph.D.

So now that we know that there's a relationship between our two

categorical variables by showing the disparity

between the marginal and conditional probabilities of an event occurring,

we need to tell the rest of our story.

After all, a good plot only takes us so far.

We need to really show the rea

Help

actually happening.

Luckily for us, the visualization of our contingency table

is pretty straightforward.

But we still need to make sure that our story is

being told from the correct perspective.

Remember, like most stories based on data,

we'll also want to use a picture to help our reader along -

a picture based on the conditional distribution

that we chose to use previously.

Now, we answered our question of interest

about our relationship between the two categorical variables

by determining the variable of interest -

the outcome variable that made the most sense.

And that's mostly true for our visualization idea as well.

As it turns out we can visualize both conditional probabilities.

The one we choose to use depends on our perspective and the role

we wish to play in talking about our data.

Imagine you are one of the crew of the Carpathia

Help

- the ship that heard Titanic's distress call.

You arrive at the location, to find the Titanic lost below the surface

- and either people in lifeboats, floating in the icy water

holding on for dear life, or, unfortunately, dead.

In this situation, I think the Row percentages

- the conditional distribution of Class based on Survivorship

would make the most sense to use for a visualization.

You come upon a survivor; you might want to know what Class they are in.

That's what this conditional distribution tells us.

Given a Survivorship status, what's the likelihood

that a person will be in a specific class.

For example, you come upon a non-survivor.

What's the likelihood that they will be a first class member?

It's about 8%.

You come upon another non-survivor - what's the likelihood that they will be

a crew member?

Help

It's about 45%.

If we want to graph this idea, it would look like this: a bar plot.

Remember that since we have counts (or percentages in this case)

of a categorical variable, we want use a bar plot for our visualization.

Now, at first, this doesn't seem that strange.

It actually follows the marginal distribution of Class membership

pretty closely.

Unfortunately things start seeming a little bit off

if we look at the survivors graph.

Again, imagine you come upon a survivor, what's

the likelihood they will be a first class member?

Well it's about 29%.

The likelihood they will be a crew member is about 30%.

Here's the graph of Survivorship.

These proportions don't really match the marginal distribution.

But you might be saying: "but they're about equal

though, each class on the Titanic had an almost equal chance of surviving."

Help

And you, unfortunately, would be wrong.

You would be right with this particular graph

if you said "among survivors of the Titanic,

there's an almost equal chance that the survivor will

be among one of the four classes."

So why was the first statement wrong - why

was "each class on the Titanic had an almost equal chance of surviving"

statement wrong?

Because that statement is from the perspective of the Class

member, not the perspective of the Survivor.

It's the perspective of our outcome variable.

Remember our variable of interest was Survivorship -

the likelihood of surviving the Titanic.

Let's look at our Column percentages - the conditional distribution

of Survivorship based on Class - to make this clear.

Here's the table with our conditional distribution

of Survivorship based on Class level.

Let's look at this from the perspective of a first class passenger.

The Titanic has just struck the iceberg.

OK, fast forward about an hour into the movie.

You realize that you need to evacuate the ship.

Help

As a first class passenger, what's the likelihood that you will survive?

Well, it's about 62%.

Sweet.

What's the likelihood that you'll survive if you're a third class

passenger - one of Leonardo's friends?

Well, it's about 25%.

Ouch.

Here's the graph of Survivorship based on Class.

This graph is actually redundant, as the sum of both bars (Survived and not

Survived) for each Class level will add up to 100%.

So to make things clearer, we might use this graph -

the likelihood of surviving for each Class level.

In the end, we want to use a graph that MATCHES our conditional probabilities

conclusion.

The graph that takes advantage of our

Help

variable of interest.

So to finally answer our question, our full question -

"Is there a relationship between the class of a passenger

and if they survived the Titanic disaster?"

and "What does that relationship between Survivorship and Class

on the Titanic look like?" - we do several things:

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

table.

We determine both our marginal and our conditional distributions

for that variable of interest, and then we compare those probabilities.

If those values don't come close to matching,

then we have a relationship between our two categorical variables.

Finally, we graph the relationship.

Although we can pick one of several different graphs

based on our perspective, it's always best

to use the graph that coincides with our conditional probabilities

from our variable of interest.

In our case - yes: Survivorship is in fact related to Class membership.

The likelihood of surviving decreases as your Class level gets lower and lower.

Help

Comprehension Check

Is there an association between car color and marital status? Below are data collected to address this question.

	Red	Black/White/Silver	Other
Married	40	22	19
Unmarried	45	10	12

(6/6 points)

Using the contingency table above, solve for each of the following probabilities. (*Report as proportions rounded to 3 decimal places*.)

Help

1) **P (black/white/silver)**

0.216

0.216

2) **P (married)**

0.547

0.547

3) **P (unmarried and black/white/silver car)**

0.068

0.068

4) **P (red car | married)**

0.494

0.494

Help

5) **P (married | red car)**

0.471

0.471

6) What ratio would you use to solve for P(red car| unmarried) using the contingency table?

- 40/148
- 40/81
- 45/85
- 45/67 ✔

Check Show Answer

Now solve for P (red car | unmarried) using the formula below. (*Probabilities should be reported as proportions rounded to 3 decimal places.*)

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$P(A \cap B)$ = **X**

$P(B)$ = **Y**

$P(A \mid B)$ = **Z**

(8/8 points)

7) **X**

0.304

0.304

Answer: .304

8) **Y**

0.453

0.453

Answer: .453

9) **Z**

Help

0.672

0.672

Answer: .672

Help

Solve for these three values. (*Report each as a proportion rounded to three decimal places*.)

10) **P (red)**

0.574

0.574

Answer: .574

11) **P (red | married)**

0.494

0.494

Answer: .494

12) **P (red | unmarried)**

0.672

Help

0.672

Answer: .672

13) If car color and marital status are independent, what should be true?

- P(red | married) = P(married | red)
- P(red) = P(married)
- P(red) = P(red | married)
- P(red) = P(married | red)

14) Do car color and marital status appear to be independent?

- Yes, because the probability of having a red car is identical for everyone.
- Yes, because there are more married people and more red cars in the sample.
- No, because more people drive red cars than other colored cars. Therefore, color is not independent of marital status.
- No, because the probability of having a red car is different for married and unmarried people. The difference is small, however

Check

Hide Answer