

MODULE 15

2-27-19

2 CATEGORICAL VARIABLES

- ALWAYS TO LOOK @ 2 CATEGORICAL VAR'S TOGETHER SLIDE 1

$$\% \text{ --- } = \text{ --- } X \text{ OUT OF } Y = \frac{X}{Y} = \%$$

- CONDITIONAL PROBABILITIES

- THE %'S ARE THE CONDITIONAL PROBABILITY

SLIDE 2 WHAT % SURVIVORS WERE IN 1ST CLASS

$$\underline{203} \text{ OUT OF } \underline{711} = \frac{203}{711} = 0.28551 \approx 28.6\%$$

SLIDE 3 WHAT % DEAD WERE IN 1ST CLASS?

$$\underline{122} \text{ OUT OF } \underline{1490} = \frac{122}{1490} \approx 8.2\%$$

* NOMENCLATURE (PROBABILITY NOTATION):

$P(\text{---}) \Rightarrow$ "PROBABILITY OF"

\Rightarrow EXAMPLE: $P(\text{SURVIVAL}) = \frac{711}{2201} = 32\%$

\hookrightarrow WHOLE BOAT TOGETHER

\Rightarrow EXAMPLE: $P(\text{FIRST CLASS} | \text{SURVIVED}) = \frac{203}{711} \approx 28.6\%$

THIS MEANS
A CONDITIONAL
PROBABILITY

$P(A|B) =$ PROBABILITY OF "A" GIVEN "B"

\hookrightarrow DETERMINES THE DENOMINATOR

$\hookrightarrow P(A|B) = \frac{\# \text{ OF } A \text{ AND } B}{\# \text{ OF } B}$

SLIDE 2

$$P(1^{\text{ST}} / \text{SURVIVE}) = \frac{\textcircled{203}}{\textcircled{711}}$$

\swarrow 1ST CLASS THAT SURVIVED
 \searrow # OF TOTAL SURVIVORS

2-25-19

* EXAMPLE:

IS THERE AN ASSOCIATION BETWEEN SURVIVAL & CLASS?

⇒ YES, THERE IS. THE % OF SURVIVORS WHO WERE IN 1ST CLASS IS 29.6% WHERE AS IF THE SURVIVORS WERE TRULY RANDOM, WE WOULD EXPECT ABOUT 15% IN 1ST CLASS; ALMOST TWICE AS MANY AS EXPECTED SURVIVED.

- STATISTICS: FINDING #'S THAT DESCRIBE A SAMPLE
↳ STATISTIC ⇒ SAMPLE
- PARAMETERS: FINDING #'S THAT DESCRIBE A POPULATION
↳ PARAMETERS ⇒ POPULATION

2-25-19

* EXAMPLE.

$$P(\text{FIRST CLASS} | \text{DEAD}) = \frac{122}{1490} \approx 8.2\%$$

=> IN SENTENCE FORM:

8.2% OF DEAD PEOPLE WERE IN FIRST CLASS OF THE TITANIC.

• FIND $P(3^{\text{RD}} \text{ CLASS} | \text{DEAD}) = \frac{528}{1490} = 0.3544 = 35.4\%$

SLIDE 3

35.4% OF THOSE WHO DIED WERE IN THIRD CLASS.

SEGMENTED BAR CHARTS SLIDE 4

• IF THERE IS NO ASSOCIATION BETWEEN TWO VARIABLES (IN THIS CASE CLASS & SURVIVAL), THE PROBABILITIES SHOULD MATCH NUMERICALLY & VISUALLY THEIR "RANDOM CHOSEN" PROBABILITIES. ($\pm 5\%$)

	<u>A LIVE</u>	<u>DEAD</u>	<u>RANDOM</u>
1 ST	28.6%	8.2%	$\frac{327}{220} \approx 15\%$
2 ND	16.6%	11.2%	$\frac{285}{2201} \approx 13\%$
3 RD	25%	35.4%	$\frac{706}{2201} \approx 32\%$
CREW	29.4%	45.2%	$\frac{885}{2201} \approx 40\%$

* NO ASSOCIATION MEANS THE VARIABLES ARE

INDEPENDENT OF EACH OTHER

$$P(A|B) = P(A) \text{ SAME!}$$