Probability with Tables

Dr. Tom Ilvento
FREC 408

Rules so far

Probability of A Union $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

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

Probability of an Intersection $P(A \cap B) = P(B)P(A \mid B)$

Cross-Tab of Treatment Type versus Still Smoking After 8 Weeks

We have a cross-tabulation of Treatment by Still Smoking after 8 weeks

	Yes	No	Row Total
Nicotine Patch	64	56	120
Placebo	96	24	120
Column Total	160	80	240

Handout of Class Data

- Let Event A = Received a Nicotine Patch.
- What is the probability of Event A?

SMOKING TREATMENT EXPERIMENT

Treatment	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Nicotine	120	50.0	120	50.0
Placebo	120	50.0	240	100.0

P(A) = 120/240 = .50

Still Smoking after 8 weeks

- Let Event B = No Longer Smoking.
- What is the probability of Event B?

STILL SMOKING AFTER 8 WEEKS

SMOKING	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	160	66.7	160	66.7
No	80	33.3	240	100.0

P(B) = 80/240 = .333

Cross-Tab of Treatment Type versus Still Smoking After 8 Weeks

We have a cross-tabulation of Treatment by Still Smoking after 8 weeks

	Yes	No	Row Total
Nicotine Patch	64	56	120
Placebo	96	24	120
Column Total	160	80	240

Union of Events A and B

- What is the union of Events A (Received Nicotine Patch) and B (No Longer Smoking)
- $(A \cup B)$ =
- ______ Everyone who received the patch
- + ______ Everyone who no longer smokes
- - ______ Everyone who is both

Union of Events A and B

120 Everyone who received the patch

80 Everyone no longer smoking

- 56 Everyone who is both – this is the INTERSECTION

144 $(A \cup B)$

Probability of the Union of Events A and B

$P(A \cup B)$ = **144/240 = .60**

Intersection of Receiving the Patch Versus No Longer Smoking

- What is the Intersection Receiving the Patch Versus No Longer Smoking?
- $(A \cap B)$

56 Everyone who both received the nicotine patch and is no longer smoking

$P(A \cap B)$ = **56/240 = .233**

Let's do a check using formulas

Probability of A Union

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

P(A) = .50

P(B) = .333

P(A) + P(B) = .833

$P(A \cap B)$ = .233

$P(A \cup B)$ = .833 - .233 = .60

Conditional Probability

- A Conditional Probability statement would be "The probability of No Longer Smoking given you received the Nicotine Patch" and is defined as

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

Conditional Probability

- The probability of No Longer Smoking given you received the Nicotine Patch

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

P(B|A) = **.233/.50 =** **.467**

Conditional Probability in a Table

- Note: I can solve for the P(A|C) directly, as long as I understand how to percentage my table
- There are 120 who received the Nicotine Patch in the study – see the number in the row margin
  - This is the given, as in *given you received the Nicotine Patch*

Conditional Probability in a Table

- And 56 of those that received the patch were not smoking after 8 weeks
- So, **56/120 = .467**
- In a cross-tab this is called the **row percentage**
- It is the conditional probability, conditioned on the row attribute

Conditional Probability in a Table

We have a cross-tabulation of Treatment by Still Smoking after 8 weeks

	Yes	**No**	Row Total
Nicotine Patch	64	**56**	**120**
Placebo	96	24	120
Column Total	160	80	240

56/120 = .467 which is the row proportion

The Complement of A

- The Complement of A would be "Received the Placebo"
  - Denoted as A^c
  - aka "Placebo"
- What is the $P(A^c)$ and $P(A^c \cap B)$?

$P(A^c)$ = **.50**

$P(A^c \cap B)$ = **24/240 = .10**

Conditional Probability for A^c

- The probability of No Longer Smoking given you received the Placebo

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

$P(B|A^c)$ = **.10/.50 =** **.20**

Also: 24/120 = .20

SAS Output Treatment by Still Smoking

```
TABLE OF TREATMENT BY STILL SMOKING

TREATMENT     STILL SMOKING
Frequency,
Percent  ,
Row Pct  ,
Col Pct  , YES    , NO     ,  Total
---------------------------
NICOTINE ,     64 ,     56 ,    120
         ,  26.67 ,  23.33 ,  50.00
         ,  53.33 ,  46.67 ,
         ,  40.00 ,  70.00 ,
---------------------------^
PLACEBO  ,     96 ,     24 ,    120
         ,  40.00 ,  10.00 ,  50.00
         ,  80.00 ,  20.00 ,
         ,  60.00 ,  30.00 ,
---------------------------^
Total         160       80      240
            66.67    33.33   100.00
```

Look at the First Cell – Nicotine Patch who Are Still Smoking

Percent	The cell value over the total	26.67 = 64/240 * 100
Row Pct	The cell value over the row margin on the right	53.33 = 64/120 * 100
Col Pct	The cell value over the column margin on the bottom	40.00 = 64/160 * 100

Look at the 2nd Cell – Those with a Nicotine Patch Who Are No Longer Smoking

$P(A \cap B) = .2333$	**is the percent**
$P(B\|A) = P(A \cap B)/P(A) = .467$	**is the row proportion for those No Longer Smoking who received the Nicotine Patch**
$P(A\|B) = P(A \cap B)/P(B) = .700$	**is the column proportion for those No Longer Smoking who received the Nicotine Patch**

Question 9 page 4

- **So what do you think the**
- **$P(A|B)$ is for our table?**
- This is the Probability of receiving a Nicotine Patch given you are No Longer Smoking
- **$P(A \cap B)/P(B) =$ .70**

- **Does this make any sense?**

How to Percentage a Table

- If you can specify a conditional probability
- Or if you can specify that one variable causes or influences a second variable
  - The first variable is called an **independent variable** (*this is the given*)
  - The second is the **dependent variable**

How to Percentage a Table

- **Percentage in the direction of the independent variable**
  - If the independent variable is at the top, use **column percentages**
  - If the independent variable is on the side, use **row percentages**

Let's go back to our table for the multiplicative rule

- We said the probability of the intersection between A and B is

$$P(A \cap B) = P(A)P(B \mid A)$$

Percent	P(A ∩ B) = **56/240 = .233**
Table Approach	P(B|A) = P(A ∩ B)/P(A) = **56/120** = **.4667**
Using formula	P(A ∩ B) = **.50 * .4667 = .233**

Independence

- Events A and B are **independent** events if the occurrence of B does not alter the probability that A has occurred.
  - P(A|B) = P(A)
  - P(B|A) = P(B)
- Events that are not independent are **dependent**

Independence

- Furthermore, if Events A and B are independent, the the probability of their intersection simplifies to:
  - P(A∩B) = P(A)P(B)
- Why???
  - P(A∩B) = P(A)P(B|A)
  - And if A and B are independent then
  - P(B|A) = P(B)

Be careful with in reporting data

- **The leading cause of death of children 1 to 4 years of age is accidents!**
- **36%** of all deaths for this age group comes from accidents
- One might conclude there is an accident waiting to happen for our children

Cause of death for children

- But the death rate for children 1 to 4 is 38.3 per 100,000 children
- This means the probability of a child aged 1 to 4 dying is only .000383
- And the overall probability of dying from an accident is .000138
- Anyone with toddlers knows that they are amazingly indestructible!
- But we also know how tragic it is when one dies from an accident

Here's How the data would look like

Cause of Death	Deaths	Nondeaths	
Accidents	2,147 .361	?	?
All Others	3,801 .639	?	?
Total	5,948	15,530,026	15,535,974

Odds and Odds Ratios – not in book

- Odds ratios have become popular ways of displaying scientific data – it was always popular in gambling
- Whenever you hear an expression such as, "One group is 3 times more likely to suffer from chronic disease", this is an odds ratio.
- It expresses the likelihood of one group experiencing a situation relative to another group.
- An ***Odds Ratio*** is the ratio of two odds

Odds

- The Odds of an event is the ratio of the probability of the event to probability of not in event
- For example, for the those who received a Nicotine Patch, the odds of Not Smoking versus Smoking
  - 56/64 = **.875**
  - Or 23.33/26.67 =

```
TABLE OF TREATMENT BY STILL SMOKING

TREATMENT     STILL SMOKING
Frequency,
Percent  ,
Row Pct  ,
Col Pct  , YES    , NO     ,  Total
---------------------------
NICOTINE ,     64 ,     56 ,    120
         ,  26.67 ,  23.33 ,  50.00
         ,  53.33 ,  46.67 ,
         ,  40.00 ,  70.00 ,
---------------------------^
PLACEBO  ,     96 ,     24 ,    120
         ,  40.00 ,  10.00 ,  50.00
         ,  80.00 ,  20.00 ,
         ,  60.00 ,  30.00 ,
---------------------------^
Total         160       80      240
            66.67    33.33   100.00
```

Odds

- **An odds of 1 means equal probabilities**
- Odds are bounded by zero on the bottom end, but unbounded on the upper end

Note: the pivot point for Odds and Odds Ratios is 1.0

Odds

- If you calculated it in terms of
- Still Smoking/Not Smoking
  - 64/56 = **1.143**
- It is the reciprocal of .875

Odds Ratio

- An **Odds Ratio** is the ratio of two odds
- It is a way to compare the odds for two groups
- We compare the two odds by taking a ratio of one to the other

An Example of Odds Ratios

- **Odds of Nicotine patch group, Not Smoking versus Smoking**
  - **56/64 = .875**
- **Odds of Placebo group, Not Smoking versus Smoking**

  24/96 = .25
- **Odds Ratio .875/.25 = 3.5**

Say it in words

Those who received a nicotine patch were 3.5 times more likely to not be smoking than those with a placebo

More on odds

- Odds and Odds Ratios are used often in research where the outcome is categorical
  - Health fields –
    - die/not die, cancer/no cancer
  - Marketing
    - Purchase/don't purchase

More on Odds

- Sometimes we take the log of the odds – called a **Logit**
  - The reference point for a Logit is zero, since the log of 1 is zero
- Odds can be very sensitive to extremes!
  - Odds of group 1 = 10/1000 = .01
  - Odds of group 2 = 1/1000 = .001
  - Odds Ratio = .01/.001 = 10

Recent study of Hormone Replacement Therapy (HRT)

- Study of 16,608 postmenopausal women aged 50-79 recruited in 1993-1998
- 8,506 received estrogen + progestin
- 8,102 received placebo
- They were tracked over time
- Let's look at the data for Cardiovascular disease

Cardiovascular disease

Cardiovascular Disease Present	**Treatment**	**Placebo**	**Row Margins**
Yes	164	122	286
No	8,342	7,980	16,322
Column Margins	8506	8102	16,608

Calculate

- What is the probability of having cardiovascular disease?
  - P(C) =
- What is the probability of having cardiovascular disease, given you received the treatment
  - P(C|T) =
- What is the probability of having cardiovascular disease, given you received the placebo?
  - P(C|P) =

Calculate

- What is the probability of having cardiovascular disease P(C) =?
  - **P(C) = 286/16,608 = .01722**
- What is the probability of having cardiovascular disease, *given* you received the treatment?
  - **P(C|T) = 164/8,506 = .01928**
- What is the probability of having cardiovascular disease, *given* you received the placebo?
  - **P(C|P) = 122/8,102 = .01506**

Calculate the Odds Ratio

- Odds of having cardiovascular disease versus not for those in the treatment group?
  - Odds =
- Odds of having cardiovascular disease versus not for those in the placebo group?
  - Odds =
- Odds Ratio =

Calculate the Odds Ratio

- Odds of having cardiovascular disease versus not for those in the treatment group?
  - **Odds = 164/8342 = .01966**
- Odds of having cardiovascular disease versus not for those in the placebo group?
  - **Odds = 122/7,980 = .01529**
- **Odds Ratio = 1.286 or 1.3**
- **Women who got the treatment were 1.3 times as likely to have cardiovascular disease (versus not) compared to the treatment group**