

BIOS 600: Principles of Statistical Inference

Probability

Fall 2012

- ▶ Pagano and Gauvreau, Chapter 6, Section 6.2

Additive Rule of Probability

Recall the additive rule of probability: if events A , B , C , \dots , are *mutually exclusive* – so that at most one of them may occur at any one time – then

$$\Pr(A \cup B \cup C \dots) = \Pr(A) + \Pr(B) + \Pr(C) + \dots$$

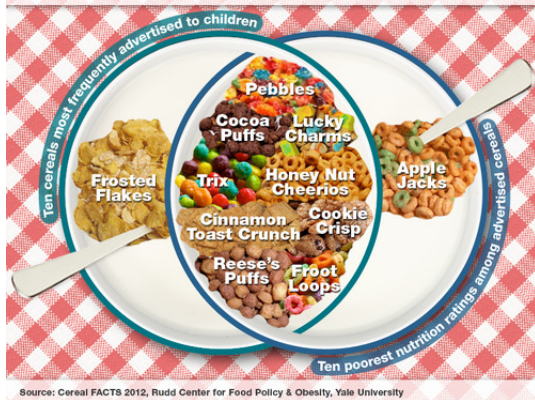
Interpretation: If events cannot happen together, the probability that any one of them will occur is the sum of their individual probabilities.

What if Events Can Overlap?

The Least Nutritious Cereals Are Those Most Frequently Advertised to Children

This graphic shows the overlap between the cereals most advertised to children and the cereals with the poorest nutrition ratings.

Read the latest Cereal FACTS report at www.cerealfacts.org



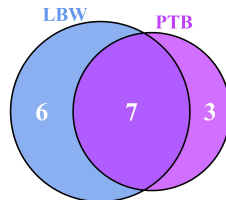
Non-Mutually Exclusive Events

When the events are not mutually exclusive, the additive rule does not apply. Let A be the event that a baby is preterm and B be the event that a baby is low birth weight. Because the two events often occur simultaneously, there is considerable overlap between A and B . If we just add the probability of individual events, the overlap area would be counted twice instead of just once. So we use the formula

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

Note that when the events are mutually exclusive, $\Pr(A \cap B) = 0$ and the formula reduces to the additive rule.

Figure 1: Percent low birth weight and preterm birth from sample of ≈ 5000 births in PIN Study at UNC



Asthma Patients

Use complementary therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

What is the probability that a patient responds that conventional medicines help (event A) or that the patient uses complementary therapy (event B)?

Find this probability directly from the table and then using the probability equations for $Pr(A \cup B)$.

Conditional Probability

Conditional probabilities can cause confusion, but in fact we deal with them all the time.

- ▶ Suppose we care about the probability a woman has breast cancer, $Pr(BCa)$

Conditional Probability

Conditional probabilities can cause confusion, but in fact we deal with them all the time.

- ▶ Suppose we care about the probability a woman has breast cancer, $Pr(BCa)$
- ▶ What if the woman has a suspicious mammogram?

Conditional Probability

Conditional probabilities can cause confusion, but in fact we deal with them all the time.

- ▶ Suppose we care about the probability a woman has breast cancer, $Pr(BCa)$
- ▶ What if the woman has a suspicious mammogram?
 - ▶ $Pr(BCa \mid M+)$ is the conditional probability that a woman has breast cancer if her mammogram is suspicious

Conditional Probability

Conditional probabilities can cause confusion, but in fact we deal with them all the time.

- ▶ Suppose we care about the probability a woman has breast cancer, $Pr(BCa)$
- ▶ What if the woman has a suspicious mammogram?
 - ▶ $Pr(BCa \mid M+)$ is the conditional probability that a woman has breast cancer if her mammogram is suspicious
- ▶ What if the woman has a clean mammogram?

Conditional Probability

Conditional probabilities can cause confusion, but in fact we deal with them all the time.

- ▶ Suppose we care about the probability a woman has breast cancer, $Pr(BCa)$
- ▶ What if the woman has a suspicious mammogram?
 - ▶ $Pr(BCa \mid M+)$ is the conditional probability that a woman has breast cancer if her mammogram is suspicious
- ▶ What if the woman has a clean mammogram?
 - ▶ $Pr(BCa \mid M-)$ is the conditional probability a woman has breast cancer if her mammogram is clean (hopefully this is lower than $Pr(BCa \mid M+)$!)

Conditional Probability

Conditional probabilities can cause confusion, but in fact we deal with them all the time.

- ▶ Suppose we care about the probability a woman has breast cancer, $Pr(BCa)$
- ▶ What if the woman has a suspicious mammogram?
 - ▶ $Pr(BCa | M+)$ is the conditional probability that a woman has breast cancer if her mammogram is suspicious
- ▶ What if the woman has a clean mammogram?
 - ▶ $Pr(BCa | M-)$ is the conditional probability a woman has breast cancer if her mammogram is clean (hopefully this is lower than $Pr(BCa | M+)$!)
- ▶ Knowledge of the mammogram result changes our best estimate of her probability of cancer

Conditional Probability

Conditional probabilities can cause confusion, but in fact we deal with them all the time.

- ▶ Suppose we care about the probability a woman has breast cancer, $Pr(BCa)$
- ▶ What if the woman has a suspicious mammogram?
 - ▶ $Pr(BCa | M+)$ is the conditional probability that a woman has breast cancer if her mammogram is suspicious
- ▶ What if the woman has a clean mammogram?
 - ▶ $Pr(BCa | M-)$ is the conditional probability a woman has breast cancer if her mammogram is clean (hopefully this is lower than $Pr(BCa | M+)$!)
- ▶ Knowledge of the mammogram result changes our best estimate of her probability of cancer
- ▶ The latter two probabilities are *conditional probabilities*. $Pr(BCa | M+)$ is the probability of breast cancer *within the group of women with suspicious mammograms*.

Multiplicative Rule of Probability

The *multiplicative rule of probability* is a useful probability rule that will help us find conditional probabilities. It states $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A) = \Pr(B)\Pr(A \mid B)$.

Multiplicative rule in English: the chance that two events will both happen equals the chance that the first will happen, multiplied by the chance that the second will happen given that the first happened.

Multiplicative Rule of Probability

- ▶ Let A =event someone uses complementary therapy and B =event that meds help

Use comp. therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

Multiplicative Rule of Probability

- ▶ Let A =event someone uses complementary therapy and B =event that meds help
- ▶ Find $Pr(A \cap B)$

Use comp. therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

Multiplicative Rule of Probability

- ▶ Let A =event someone uses complementary therapy and B =event that meds help
- ▶ Find $Pr(A \cap B)$
- ▶ From table, this is $\frac{122}{1000} = 0.122$

Use comp. therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

Multiplicative Rule of Probability

► Find

$Pr(\text{use complementary therapy and meds help})$

Use comp. therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

Multiplicative Rule of Probability

► Find

$Pr(\text{use complementary therapy and meds help})$

Use comp. therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

Multiplicative Rule of Probability

- Find
 $Pr(\text{use complementary therapy and meds help})$

- Use multiplicative rule with
A=event someone uses
complementary therapy and
B=event that meds help,
then
 $Pr(A \cap B) = Pr(A)Pr(B | A)$

Use comp. therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

Multiplicative Rule of Probability

- Find
 $Pr(\text{use complementary therapy and meds help})$

- Use multiplicative rule with
A=event someone uses
complementary therapy and
B=event that meds help,
then
 $Pr(A \cap B) = Pr(A)Pr(B | A)$

► $Pr(A) = \frac{146}{1000} = 0.146$

Use comp. therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

Multiplicative Rule of Probability

- ▶ Find
 $Pr(\text{use complementary therapy and meds help})$

- ▶ Use multiplicative rule with
A=event someone uses
complementary therapy and
B=event that meds help,
then
 $Pr(A \cap B) = Pr(A)Pr(B | A)$

- ▶ $Pr(A) = \frac{146}{1000} = 0.146$

- ▶ $Pr(B | A) = \frac{122}{146} = 0.836$

Use comp. therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

Multiplicative Rule of Probability

- ▶ Find
 $Pr(\text{use complementary therapy and meds help})$

- ▶ Use multiplicative rule with
A=event someone uses
complementary therapy and
B=event that meds help,
then
 $Pr(A \cap B) = Pr(A)Pr(B | A)$

- ▶ $Pr(A) = \frac{146}{1000} = 0.146$
- ▶ $Pr(B | A) = \frac{122}{146} = 0.836$
- ▶ $Pr(A \cap B) =$
 $0.146 \times 0.836 = 0.122$

Use comp. therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

Conditional Probability

- ▶ Given that a baby is preterm, what is the probability the baby is also low birth weight?

Conditional Probability

- ▶ Given that a baby is preterm, what is the probability the baby is also low birth weight?
- ▶ The *conditional probability*, $\Pr(B \mid A)$, is the probability of the event B (low birth weight) given that A (preterm birth) occurs

Conditional Probability

- ▶ Given that a baby is preterm, what is the probability the baby is also low birth weight?
- ▶ The *conditional probability*, $\Pr(B \mid A)$, is the probability of the event B (low birth weight) given that A (preterm birth) occurs
- ▶ *Multiplicative rule of probability*: $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A)$

Conditional Probability

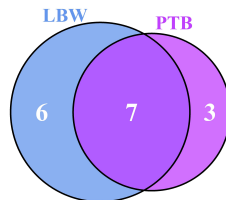
- ▶ Given that a baby is preterm, what is the probability the baby is also low birth weight?
- ▶ The *conditional probability*, $\Pr(B \mid A)$, is the probability of the event B (low birth weight) given that A (preterm birth) occurs
- ▶ *Multiplicative rule of probability*: $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A)$
- ▶ Dividing both sides of the equation by $\Pr(A)$, we find the formula for conditional probability:

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

Conditional Probability

- What is $\Pr(LBW \mid PTB)$?

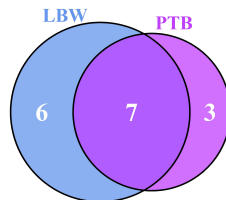
Figure 2: Percent low birth weight and preterm birth from sample of ≈ 5000 births in PIN Study at UNC



Conditional Probability

- ▶ What is $\Pr(LBW \mid PTB)$?
- ▶ Applying the formula with $A=PTB$ and $B=LBW$,
$$\Pr(LBW \mid PTB) = \frac{\Pr(PTB \cap LBW)}{\Pr(PTB)}$$

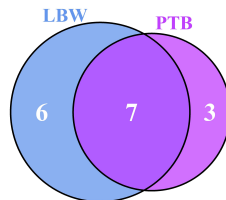
Figure 2: Percent low birth weight and preterm birth from sample of ≈ 5000 births in PIN Study at UNC



Conditional Probability

- ▶ What is $\Pr(LBW \mid PTB)$?
- ▶ Applying the formula with $A=PTB$ and $B=LBW$,
$$\Pr(LBW \mid PTB) = \frac{\Pr(PTB \cap LBW)}{\Pr(PTB)}$$
- ▶ $\Pr(LBW \mid PTB) = \frac{\Pr(PTB \cap LBW)}{0.10}$

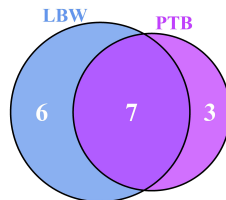
Figure 2: Percent low birth weight and preterm birth from sample of ≈ 5000 births in PIN Study at UNC



Conditional Probability

- ▶ What is $\Pr(LBW \mid PTB)$?
- ▶ Applying the formula with $A=PTB$ and $B=LBW$,
$$\Pr(LBW \mid PTB) = \frac{\Pr(PTB \cap LBW)}{\Pr(PTB)}$$
- ▶ $\Pr(LBW \mid PTB) = \frac{\Pr(PTB \cap LBW)}{0.10}$
- ▶ $\Pr(LBW \mid PTB) = \frac{0.07}{0.10} = 0.70$

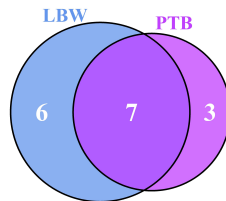
Figure 2: Percent low birth weight and preterm birth from sample of ≈ 5000 births in PIN Study at UNC



Conditional Probability

- ▶ What is $\Pr(LBW \mid PTB)$?
- ▶ Applying the formula with $A=PTB$ and $B=LBW$,
$$\Pr(LBW \mid PTB) = \frac{\Pr(PTB \cap LBW)}{\Pr(PTB)}$$
- ▶ $\Pr(LBW \mid PTB) = \frac{\Pr(PTB \cap LBW)}{0.10}$
- ▶ $\Pr(LBW \mid PTB) = \frac{0.07}{0.10} = 0.70$
- ▶ Using the Venn diagram, you concentrate on the purple circle (PTB) and calculate what fraction of those births are also LBW

Figure 2: Percent low birth weight and preterm birth from sample of ≈ 5000 births in PIN Study at UNC



Conditional Probability

- ▶ When we are considering two events in which the outcome of one event has no relationship with the outcome of the other, then the events are said to be *independent*.

Conditional Probability

- ▶ When we are considering two events in which the outcome of one event has no relationship with the outcome of the other, then the events are said to be *independent*.
- ▶ If A and B are independent events, then $\Pr(A \mid B) = \Pr(A)$ and $\Pr(B \mid A) = \Pr(B)$.

Conditional Probability

- ▶ When we are considering two events in which the outcome of one event has no relationship with the outcome of the other, then the events are said to be *independent*.
- ▶ If A and B are independent events, then $\Pr(A \mid B) = \Pr(A)$ and $\Pr(B \mid A) = \Pr(B)$.
 - ▶ Having breast cancer and having a suspicious mammogram are *not independent events*

Conditional Probability

- ▶ When we are considering two events in which the outcome of one event has no relationship with the outcome of the other, then the events are said to be *independent*.
- ▶ If A and B are independent events, then $\Pr(A \mid B) = \Pr(A)$ and $\Pr(B \mid A) = \Pr(B)$.
 - ▶ Having breast cancer and having a suspicious mammogram are *not* independent events
- ▶ In this special case, the multiplicative rule $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A)$ reduces to $\Pr(A \cap B) = \Pr(A)\Pr(B)$ because $\Pr(B \mid A) = \Pr(B)$

Conditional Probabilities in Asthma Data

Use complementary therapy	Meds Help?		Total
	Yes	No	
Yes	122	24	146
No	757	97	854
Total	879	121	1000

- ▶ What is the probability that a patient uses complementary therapy if he or she reports that conventional medications help?
- ▶ What is the probability that a patient uses complementary therapy if he or she reports that conventional medications do not help?
- ▶ Is reporting that conventional medications help independent of use of complementary therapy? Explain.

Independent vs. Mutually Exclusive

- ▶ Is independent the same as mutually exclusive?

Independent vs. Mutually Exclusive

- ▶ Is independent the same as mutually exclusive?
- ▶ For mutually exclusive events A and B , $\Pr(B \mid A) = 0$, so they are (perfectly) dependent events.

Independent vs. Mutually Exclusive

- ▶ Is independent the same as mutually exclusive?
- ▶ For mutually exclusive events A and B, $\Pr(B \mid A) = 0$, so they are (perfectly) dependent events.
- ▶ If A and B are instead independent events, A's occurrence has no impact on the occurrence of B.

Monty Hall Problem Video

An even simpler solution is to reason that switching loses if and only if the player initially picks the car, which happens with probability $1/3$, so switching must win with probability $2/3$

Still not a believer? Try it yourself on the [Monte Hall Problem Simulator](#).

Law of Total Probability

The law of total probability is given by

$$\begin{aligned}Pr(B) &= Pr(B | A)Pr(A) + Pr(B | \bar{A})Pr(\bar{A}) \\ &= Pr(B \cap A) + Pr(B \cap \bar{A})\end{aligned}$$

If B is the event you are male and A is the event you are a UNC student, then the probability you are male is the sum of the probability that you are male and a UNC student and the probability that you are male and not a UNC student.

Case Study: Internet Addiction

Consider data inspired by a 2009 study of internet addiction published in *Journal of Adolescent Health*. Internet addiction is a disorder characterized by excessive time on the Internet, impaired judgment and decision-making ability, social withdrawal, and depression. The study reported the following.

- ▶ 51.8% of study participants were female and 48.2% were male
- ▶ 13.1% of the females suffered from internet addiction.
- ▶ 24.8% of the males suffered from internet addiction.

Define the following events.

- ▶ F : the event that the selected participant is female
- ▶ I : the event that the selected participant suffers from internet addiction

Case Study: Internet Addiction

Find the following probabilities using the formulas.

1. $Pr(F)$
2. $Pr(\overline{F})$
3. $Pr(I \mid F)$
4. $Pr(I \mid \overline{F})$
5. $Pr(\overline{I} \mid \overline{F})$
6. $Pr(I)$

Case Study: Perceived Risk of Smoking

A Gallup Poll in 2002 examined perceptions of risk associated with smoking. Summary data are below.

Smoking Status	Perceived Risk			
	Very Harmful	Somewhat Harmful	Not too Harmful	Not at All Harmful
Current Smoker	60	30	5	1
Former Smoker	78	16	3	2
Never Smoked	86	10	2	1

Defend or refute the following statements based on the poll data.

1. Current smokers are less likely to view smoking as very harmful than either former smokers or those who have never smoked.
2. Former smokers are more likely to view smoking as very harmful or somewhat harmful than never smokers.

Case Study: Pima Indians

The Pima Indians are often called “pathfinders for health” due to their generosity in participating in health research after NIDDK researchers found an extremely high rate of diabetes in their population in the early 1960's. Below are data on birth defects among 1207 Pima Indian births.

	Birth defects	No defects
Nondiabetic	31	754
Prediabetic	13	362
Diabetic	9	38



Does the prevalence of birth defects among Pima Indians appear to be related to diabetic status? Why or why not?

Alternative Approach: Hypothetical 1000 Tables

Another approach to calculating conditional probabilities, probabilities of unions or intersections, or probabilities of complements is to use the “hypothetical 1000 table” approach. In this approach, probabilities provided are converted into a contingency table with a hypothetical sample size of 1000.

A probability computed from information in a table is a fraction. The key to calculating these correctly is having the right numerator (top) and right denominator (bottom) when you put your fraction together. The denominator for a *conditional* probability will be a row or column total, while the denominator for an *unconditional* probability will be the total for the entire table.

Alternative Approach: Hypothetical 1000 Tables

Montana teenagers were asked whether they had smoked a cigarette last week. The proportion of boys in the survey was 0.487, and the probability that a boy smoked was 0.218, while the probability a girl smoked was 0.173. You are asked to estimate the prevalence of smoking, the probability a respondent was male and smoked, and the probability a respondent was female given that the respondent reported smoking.

Alternative Approach: Hypothetical 1000 Tables

Using the information provided, we can fill in the table, assuming a hypothetical population of 1000 subjects. Information: proportion of boys was 0.487, probability a boy smoked was 0.218, probability a girl smoked was 0.173.

	Smoked		Total
	Yes	No	
Girl			
Boy			0.487×1000
Total			1000

Alternative Approach: Hypothetical 1000 Tables

Using the information provided, we can fill in the table, assuming a hypothetical population of 1000 subjects. Information: proportion of boys was 0.487, probability a boy smoked was 0.218, probability a girl smoked was 0.173.

	Smoked		Total
	Yes	No	
Girl			1000-487
Boy	0.218×487		487
Total			1000

Alternative Approach: Hypothetical 1000 Tables

Using the information provided, we can fill in the table, assuming a hypothetical population of 1000 subjects. Information: proportion of boys was 0.487, probability a boy smoked was 0.218, probability a girl smoked was 0.173.

	Smoked		Total
	Yes	No	
Girl	0.173×513		513
Boy	106		487
Total			1000

Alternative Approach: Hypothetical 1000 Tables

Using the information provided, we can fill in the table, assuming a hypothetical population of 1000 subjects. Information: proportion of boys was 0.487, probability a boy smoked was 0.218, probability a girl smoked was 0.173.

	Smoked		Total
	Yes	No	
Girl	89	513-89	513
Boy	106	487-106	487
Total	89+106	1000-(89+106)	1000

Alternative Approach: Hypothetical 1000 Tables

► $Pr(\text{smoked}) = ?$

	Smoked		Total
	Yes	No	
Girl	89	424	513
Boy	106	381	487
Total	195	805	1000

Alternative Approach: Hypothetical 1000 Tables

- ▶ $Pr(\text{smoked}) = ?$
- ▶ $Pr(\text{male} \cap \text{smoked}) = ?$

	Smoked		Total
	Yes	No	
Girl	89	424	513
Boy	106	381	487
Total	195	805	1000

Alternative Approach: Hypothetical 1000 Tables

- ▶ $Pr(\text{smoked}) = ?$
- ▶ $Pr(\text{male} \cap \text{smoked}) = ?$
- ▶ $Pr(\text{female} \mid \text{smoked}) = ?$

	Smoked		Total
	Yes	No	
Girl	89	424	513
Boy	106	381	487
Total	195	805	1000

Example: Destination Wedding!

A couple from Venezuela had a destination wedding in the Dominican Republic in 2011. Unfortunately, a substantial number of wedding attendees were diagnosed with cholera in the days following the wedding! Attendees were interviewed to identify the cause, and we consider data on whether attendees consumed ice (either in a drink or cooked shrimp served on ice) at the happy event.



Example: Destination Wedding Disaster!

	Cholera		Total
	Yes	No	
Ice	25	72	97
No ice	17	426	443
Total	42	498	540

► $Pr(\text{cholera}) =$



Example: Destination Wedding Disaster!

	Cholera		Total
	Yes	No	
Ice	25	72	97
No ice	17	426	443
Total	42	498	540

► $Pr(\text{cholera}) =$

► $Pr(\text{cholera} \mid \text{ice}) =$



Example: Destination Wedding Disaster!

	Cholera		Total
	Yes	No	
Ice	25	72	97
No ice	17	426	443
Total	42	498	540

► $Pr(\text{cholera}) =$

► $Pr(\text{cholera} \mid \text{ice}) =$

► $Pr(\text{cholera} \mid \text{no ice}) =$



Example: Destination Wedding Disaster!

	Cholera		Total
	Yes	No	
Ice	25	72	97
No ice	17	426	443
Total	42	498	540

- ▶ $Pr(\text{cholera}) =$
- ▶ $Pr(\text{cholera} \mid \text{ice}) =$
- ▶ $Pr(\text{cholera} \mid \text{no ice}) =$
- ▶ Do you think ice is the culprit?



Caution: How Was a Table Constructed?

Case-control study

	Cholera		Total
	Yes	No	
Ice	25	9	34
No ice	17	53	70
Total	42	62	104

A more likely design for the cholera study would be a case-control design, in which all cholera cases and a subset of controls (non-cases) would be asked whether they consumed a drink with ice or shrimp with ice at the wedding. We must be careful with probabilities in this setting.

► $Pr(\text{cholera}) =$

Caution: How Was a Table Constructed?

Case-control study

	Cholera		Total
	Yes	No	
Ice	25	9	34
No ice	17	53	70
Total	42	62	104

A more likely design for the cholera study would be a case-control design, in which all cholera cases and a subset of controls (non-cases) would be asked whether they consumed a drink with ice or shrimp with ice at the wedding. We must be careful with probabilities in this setting.

► $Pr(\text{cholera}) =$

► $Pr(\text{cholera} \mid \text{ice}) =$

Caution: How Was a Table Constructed?

Case-control study

	Cholera		Total
	Yes	No	
Ice	25	9	34
No ice	17	53	70
Total	42	62	104

A more likely design for the cholera study would be a case-control design, in which all cholera cases and a subset of controls (non-cases) would be asked whether they consumed a drink with ice or shrimp with ice at the wedding. We must be careful with probabilities in this setting.

- ▶ $Pr(\text{cholera}) =$
- ▶ $Pr(\text{cholera} \mid \text{ice}) =$
- ▶ Do we get the right estimates of probabilities here?

Reading for Next Time

- ▶ Pagano and Gauvreau, Chapter 6, Section 6.3-6.4