Two-Way Tables

Gaston Sanchez

Two-way tables, crosstables, contingency tables

Example

Body Image

Random sample of 1200 US college students

“What is your percention of your own body?”

- About right
- Underweight
- Overweight

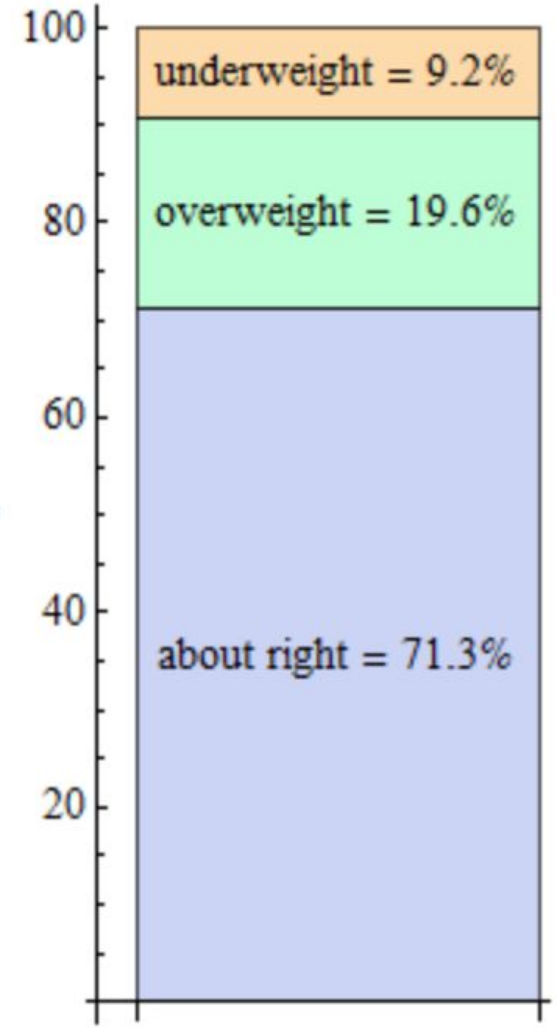
For instance ...

Student	Body Image
...	...
25	overweight
26	about right
27	underweight
28	about right
29	about right
...	...

Frequency Table

Category	Count	Proportion	Percentage
underweight	110	110/1200 = 0.092	9.2%
overweight	235	235/1200 = 0.196	19.6%
about right	855	855/1200 = 0.713	71.3%

Body Image


100
80
60
40
20
underweight = 9.2%
overweight = 19.6%
about right = 71.3%

Gender and Body Image

Student	Gender	Body Image
...	...	...
25	M	overweight
26	M	about right
27	F	underweight
28	F	about right
29	M	about right
...	...	...

Two-Way Table

	About Right	Overweight	Underweight
Male	560	163	37
Female	295	72	73

Two-Way Table

	About Right	Overweight	Underweight	*Row Totals*
Male	560	163	37	*760*
Female	295	72	73	*440*
Column Totals	*855*	*235*	*110*	***1200***

Two-Way Table

	About Right	Overweight	Underweight	*Row Totals*
Male	560	163	37	*760*
Female	295	72	73	*440*
Column Totals	*855*	*235*	*110*	***1200***

Is Body Image related to Gender?

(discussed in class)

Example

Example

Suppose we observe 2 qualitative binary variables:

Gender	male, female
Condition	smoker, non-smoker

Example of crosstable 2x2

Table formed by crossing Gender and Condition

	smoker	non-smoker
male	20	35
female	15	40

Example of crosstable 2x2

Table formed by crossing Gender and Condition

		B	B^c
		smoker	non-smoker
A	male	20	35
A^c	female	15	40

note that these are absolute frequencies

Example of crosstable 2x2

Table formed by crossing Gender and Condition

		B	B^c	
		smoker	non-smoker	*Total*
A	male	20	35	*55*
A^c	female	15	40	*55*
	Total	*35*	*75*	***110*** *grand total*

note that these are absolute frequencies

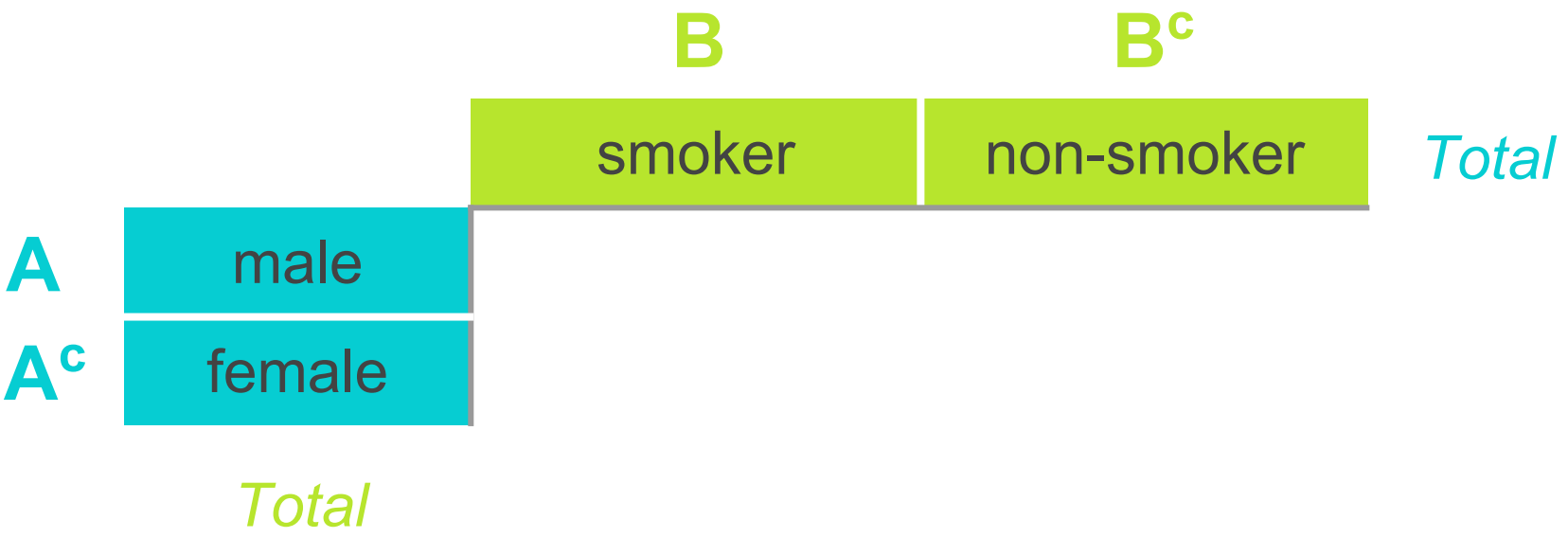
Crosstable 2x2: general case

	B	B^c	*Total*
A	A *and* B	A *and* B^c	*# A*
A^c	A^c *and* B	A^c *and* B^c	*# A^c*
Total	*# B*	*# B^c*	*# N*

In order to get probabilities...

Probability crosstable 2x2

Table formed by crossing Gender and Condition



Probability crosstable 2x2

Table formed by crossing Gender and Condition

		B	B^c	
		smoker	non-smoker	*Total*
A	male	20/110	35/110	*55/110*
A^c	female	15/110	40/110	*55/110*
	Total	*35/110*	*75/110*	***110/110*** *grand total*

note that these are relative frequencies

Probability crosstable 2x2

Table formed by crossing Gender and Condition

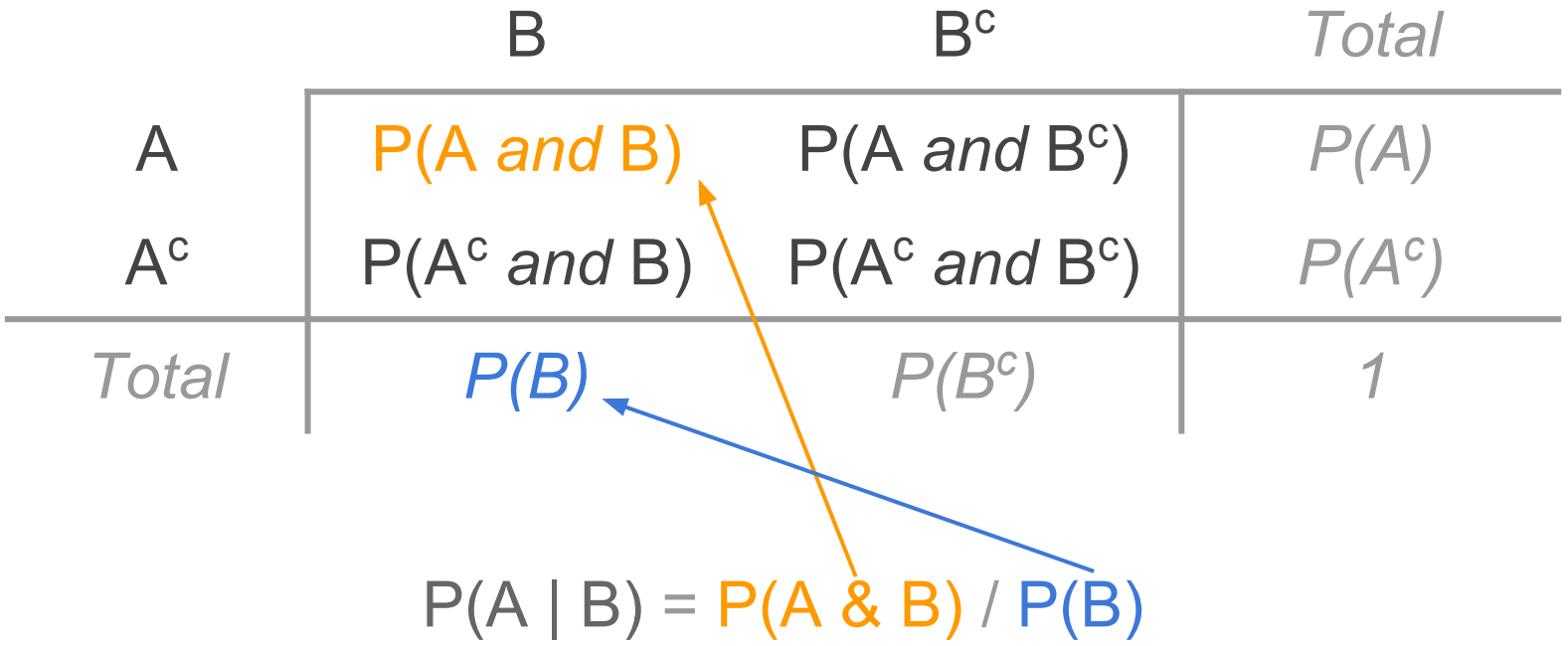
		B	B^c	
		smoker	non-smoker	*Total*
A	male	0.1818	0.3181	*0.5*
A^c	female	0.1363	0.3636	*0.5*
	Total	*0.3181*	*0.6818*	***1.0*** *grand total*

note that these are relative frequencies

Probability Crosstable 2x2

	B	B^c	*Total*
A	P(A *and* B)	P(A *and* B^c)	*P(A)*
A^c	P(A^c *and* B)	P(A^c *and* B^c)	*P(A^c)*
Total	*P(B)*	*P(B^c)*	*1*

Probability Crosstable 2x2



Probability Crosstable 2x2

	B	B^c	*Total*
A	P(A *and* B)	P(A *and* B^c)	*P(A)*
A^c	P(A^c *and* B)	P(A^c *and* B^c)	*P(A^c)*
Total	*P(B)*	*P(B^c)*	*1*

$$P(A^c \mid B) = P(A^c \,\&\, B) \,/\, P(B)$$

Probability Crosstable 2x2

	B	B^c	*Total*
A	P(A *and* B)	P(A *and* B^c)	*P(A)*
A^c	P(A^c *and* B)	P(A^c *and* B^c)	*P(A^c)*
Total	*P(B)*	*P(B^c)*	*1*

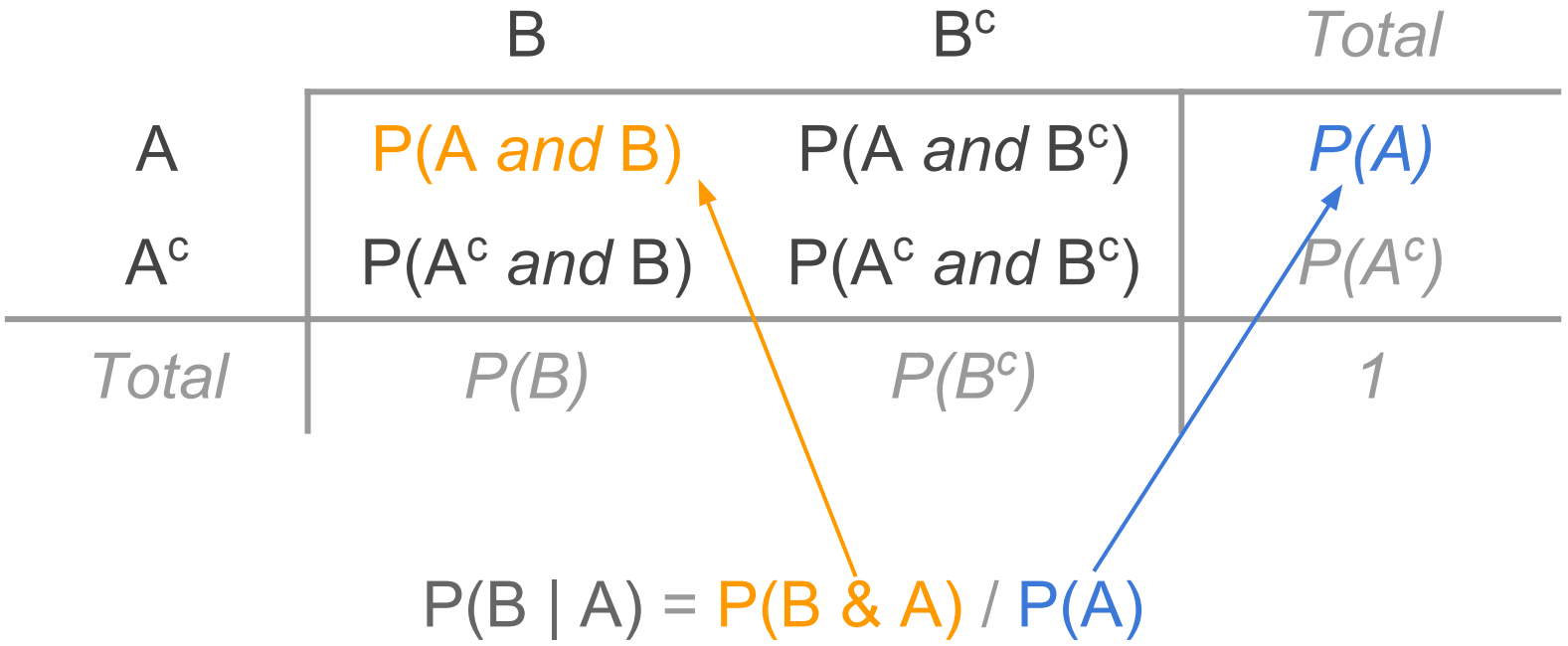
$$P(A \mid B^c) = P(A \,\&\, B^c) \,/\, P(B^c)$$

Probability Crosstable 2x2

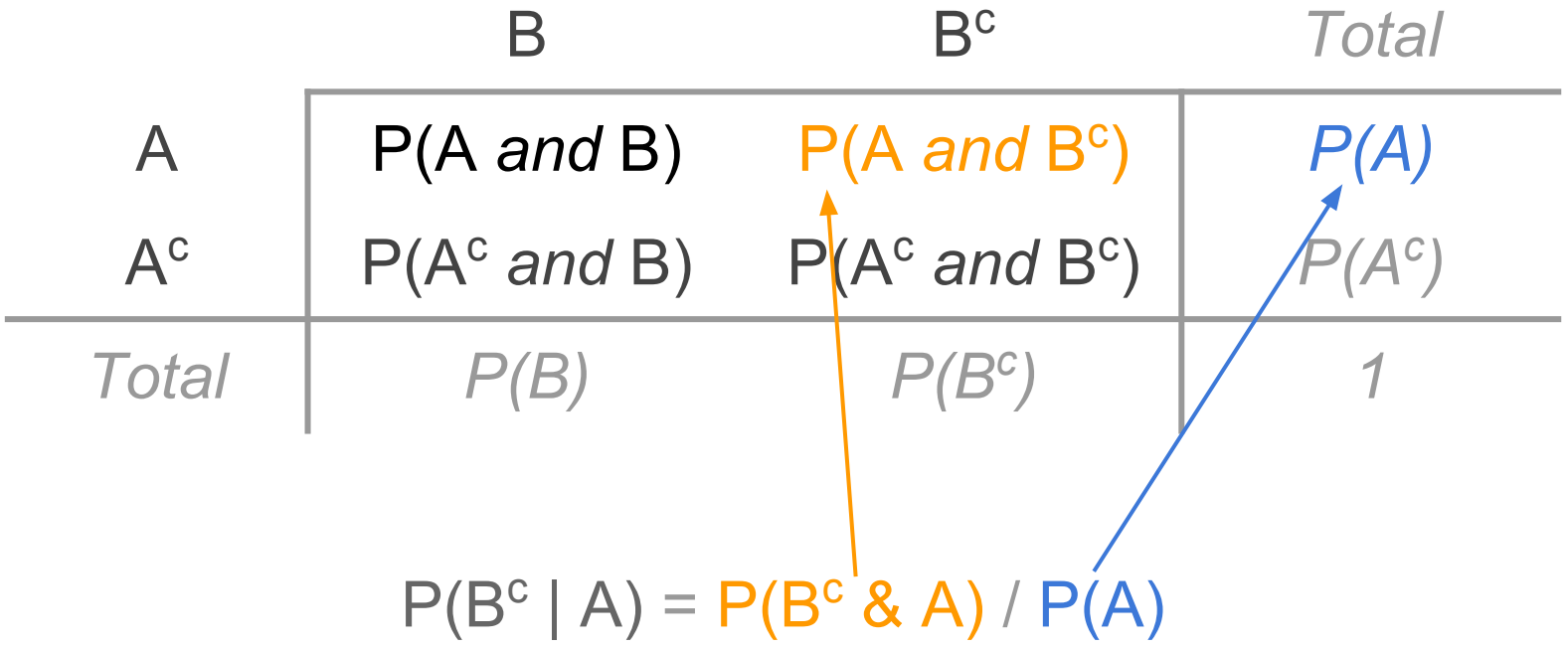
	B	B^c	*Total*
A	P(A *and* B)	P(A *and* B^c)	*P(A)*
A^c	P(A^c *and* B)	P(A^c *and* B^c)	*P(A^c)*
Total	*P(B)*	*P(B^c)*	*1*

$$P(A^c \mid B^c) = P(A^c \,\&\, B^c) \,/\, P(B^c)$$

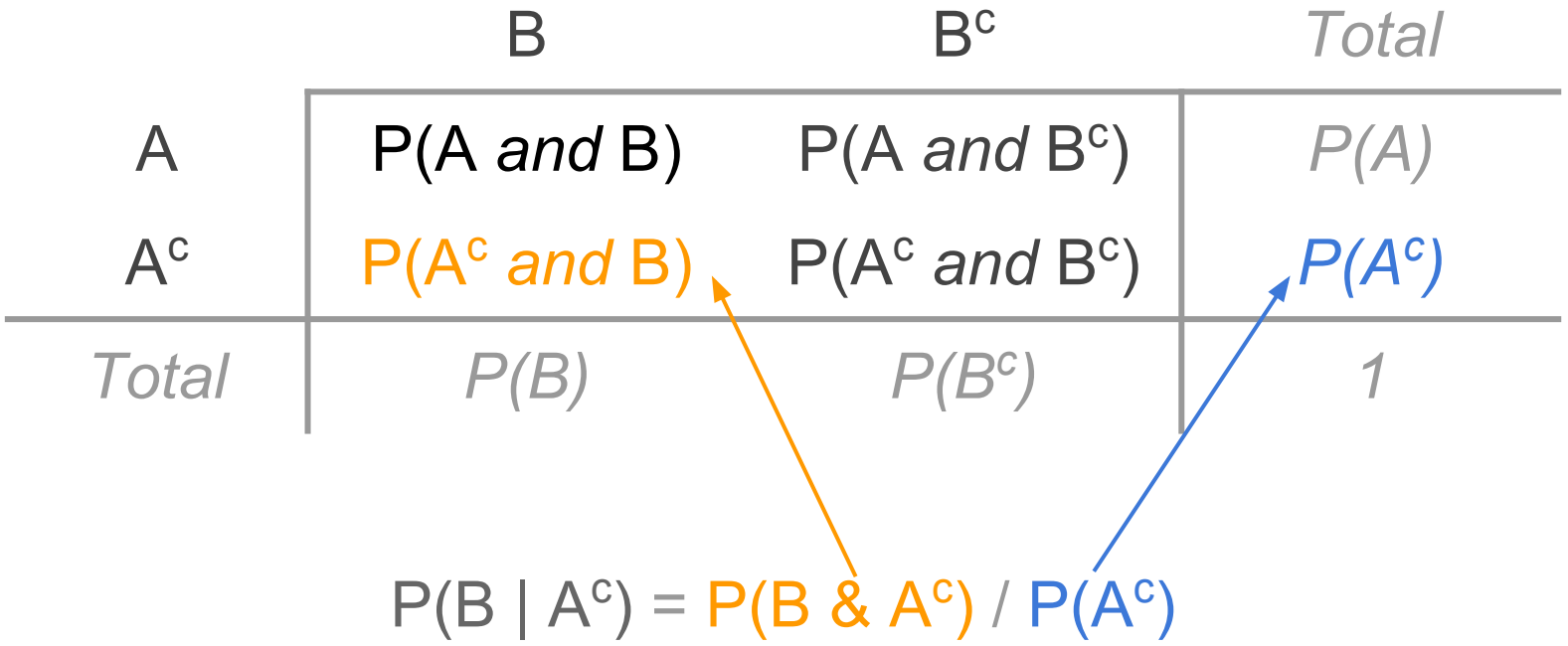
Probability Crosstable 2x2



Probability Crosstable 2x2



Probability Crosstable 2x2



Probability Crosstable 2x2

	B	B^c	*Total*
A	P(A *and* B)	P(A *and* B^c)	*P(A)*
A^c	P(A^c *and* B)	P(A^c *and* B^c)	*P(A^c)*
Total	*P(B)*	*P(B^c)*	*1*

$$P(B^c \mid A^c) = P(B^c \,\&\, A^c) \,/\, P(A^c)$$

General Crosstable *p*x*q*

Crosstables *p*x*q*

We observe 2 qualitative variables *(nominal or ordinal)*

X $A_1, A_2, A_3, \ldots, A_p$

Y $B_1, B_2, B_3, \ldots, B_q$

Crosstable *p*x*q*: general case

	B_1	B_2	...	B_q	*Total*
A_1	A_1 *and* B_1	A_1 *and* B_2	...	A_1 *and* B_q	*# A_1*
A_2	A_2 *and* B_1	A_2 *and* B_2	...	A_2 *and* B_q	*# A_2*
A_3	A_3 *and* B_1	A_3 *and* B_2	...	A_3 *and* B_q	*# A_3*
...					...
A_p	A_p *and* B_1	A_p *and* B_2	...	A_p *and* B_q	*# A_p*
Total	*# B_1*	*# B_2*	...	*# B_q*	*# N*

Example

Crosstable for current enrollment in public and private schools by level of education

	Public	Private	***Total***
Elementary	20%	30%	*50%*
High School	15%	20%	*35%*
College	10%	5%	*15%*
Total	*45%*	*55%*	***100%***

Probability that a student randomly selected is enrolled in Elementary and High School?

P(enrolled in Elementary and HS) = ?

	Public	Private	***Total***
Elementary	20%	30%	*50%*
High School	15%	20%	*35%*
College	10%	5%	*15%*
Total	*45%*	*55%*	***100%***

Independent?
Mutually Exclusive?
None of the above?

P(enrolled in Elementary and HS) = ?

	Public	Private	***Total***
Elementary	20%	30%	*50%*
High School	15%	20%	*35%*
College	10%	5%	*15%*
Total	*45%*	*55%*	***100%***

mutually exclusive events

Keep in mind ...

Mut. Exclusive events

typically has to do with outcomes of same experiment

$\neq$

Independent events

typically has to do with outcomes of different experiments