Contingency Tables

Andy Grogan-Kaylor

7 May 2020

# Last Updated

. display "$S\_TIME $S\_DATE"  
11:37:26 7 May 2020

# Key Concepts and Commands

# Flipping Two Coins

# Setup

. clear all

. set seed 3846

Good value labels are **key** here.

. label define nickel ///  
> 1 "heads for nickel" ///   
> 0 "tails for nickel" // define value label

. label define quarter ///   
> 1 "heads for quarter" ///   
> 0 "tails for quarter" // define value label

. set obs 1000 // 1000 observations  
number of observations (\_N) was 0, now 1,000

. \* curiously it takes around 1000 obs for the proportions  
. \* below to "take hold"

. generate nickel = rbinomial(1, .75) // unfair nickel

. generate quarter = rbinomial(1, .5) // fair quarter

. label values nickel nickel // assign value label

. label values quarter quarter // assign value label

# Crosstabulation

. tabulate nickel quarter, row col  
  
┌───────────────────┐  
│ Key │  
├───────────────────┤  
│ frequency │  
│ row percentage │  
│ column percentage │  
└───────────────────┘  
  
 │ quarter  
 nickel │ tails for heads for │ Total  
─────────────────┼──────────────────────┼──────────  
tails for nickel │ 104 140 │ 244   
 │ 42.62 57.38 │ 100.00   
 │ 21.62 26.97 │ 24.40   
─────────────────┼──────────────────────┼──────────  
heads for nickel │ 377 379 │ 756   
 │ 49.87 50.13 │ 100.00   
 │ 78.38 73.03 │ 75.60   
─────────────────┼──────────────────────┼──────────  
 Total │ 481 519 │ 1,000   
 │ 48.10 51.90 │ 100.00   
 │ 100.00 100.00 │ 100.00

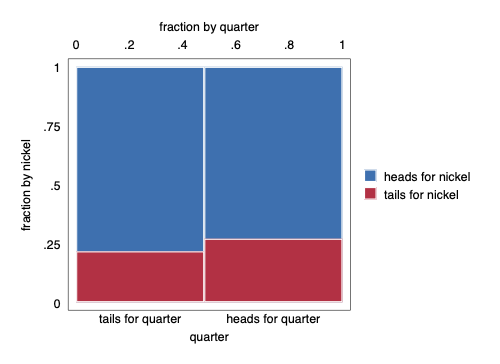
# Graphing (Mosaic Plot)

. \* ssc install spineplot // mosaicplots (spineplots)

. \* ssc install scheme-burd, replace // BuRd graph scheme

. spineplot nickel quarter, scheme(burd)

. graph export nickel-quarter.png, width(500) replace  
(file nickel-quarter.png written in PNG format)



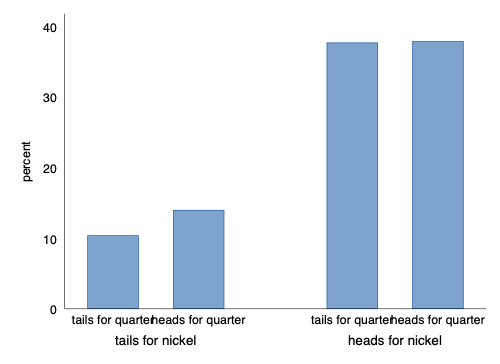
Mosaic Plot

# Bar Chart

Does a bar chart work to visualize these relationships?

. graph bar, over(quarter) over(nickel) scheme(burd)

. graph export nickel-quarter-bar1.png, width(500) replace  
(file nickel-quarter-bar1.png written in PNG format)



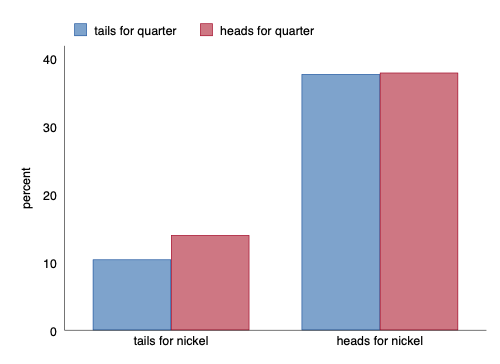
Bar Chart 1

# Bar Chart (2)

Option asyvars adds a crucial color element.

. graph bar, over(quarter) over(nickel) scheme(burd) asyvars

. graph export nickel-quarter-bar2.png, width(500) replace  
(file nickel-quarter-bar2.png written in PNG format)



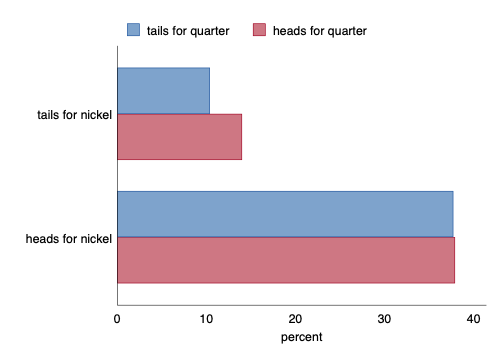
Bar Chart 2

# Horizontal Bar Chart

And hbar may improve legibility even more.

. graph hbar, over(quarter) over(nickel) scheme(burd) asyvars

. graph export nickel-quarter-bar3.png, width(500) replace  
(file nickel-quarter-bar3.png written in PNG format)



Bar Chart 3

# 1961 French Skiiers

. clear all

# Define Matrix

. matrix input FrenchSkiiers = (31, 109 \ 17, 122)

. matrix rownames FrenchSkiiers = Placebo AscorbicAcid

. matrix colnames FrenchSkiiers = Cold NoCold

. matrix list FrenchSkiiers  
  
FrenchSkiiers[2,2]  
 Cold NoCold  
 Placebo 31 109  
AscorbicAcid 17 122

# Theme Music

# Try Making a Data Set From Matrix

. svmat FrenchSkiiers, name(count)  
number of observations will be reset to 2  
Press any key to continue, or Break to abort  
number of observations (\_N) was 0, now 2

. list  
  
 ┌─────────────────┐  
 │ count1 count2 │  
 ├─────────────────┤  
 1. │ 31 109 │  
 2. │ 17 122 │  
 └─────────────────┘

# Enter Data By Hand

There are many alternative commands to do this, but the easiest way is using edit.

I have already done this. Note the structure of the data is different from above.

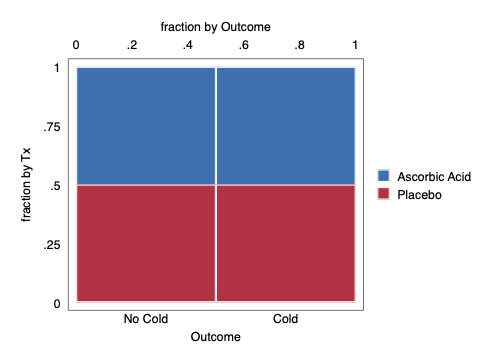
. use "FrenchSkiiers.dta", clear

. list // list the data  
  
 ┌─────────────────────────────────┐  
 │ Tx Outcome Count │  
 ├─────────────────────────────────┤  
 1. │ Ascorbic Acid Cold 17 │  
 2. │ Ascorbic Acid No Cold 122 │  
 3. │ Placebo Cold 31 │  
 4. │ Placebo No Cold 109 │  
 └─────────────────────────────────┘

# Mosaic Plot

. spineplot Tx Outcome, scheme(burd)

. graph export FrenchSkiiers1.png, width(500) replace  
(file FrenchSkiiers1.png written in PNG format)

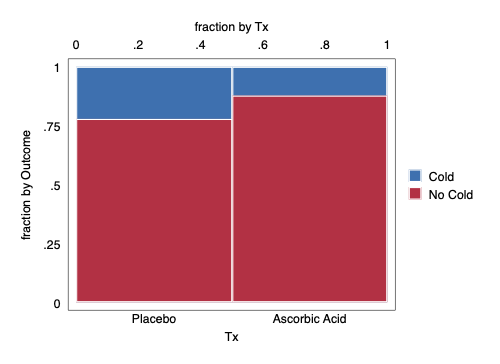


Mosaic Plot Attempt 1

# Mosaic Plot (2)

. spineplot Outcome Tx [fweight=Count], scheme(burd) // order matters to interpretability

. graph export FrenchSkiiers2.png, width(500) replace  
(file FrenchSkiiers2.png written in PNG format)



Mosaic Plot Attempt 2

# Definitions and Notation

## Counts

## Probabilities

# Terms

are *joint* probabilities.

and are *marginal* probabilities.

and are *conditional* probabilities.

# Formulas

## Counts

## Probabilities

## Expected Probabilities and Counts or Frequencies

Observed counts are represented by while expected counts are represented by .

# Fundamental Rule

# Independence (Robert Mare)

If independence is true, then joint probabilities = products of marginal probabilities.

That is, under independence, the conditional distribution equals the marginal distribution.

Under independence, row membership provides no information about the column distribution; and column membership provides no information about the row distribution.

Independence is a model, which is never exactly true in the real world.

# Observed vs. Expected

. scalar N = 31 + 109 + 17 + 122

. scalar A = ((31 + 17)\*(31+109)) / N // expected count

. scalar B = ((31+109)\*(109+122)) / N // expected count

. scalar C = ((31 + 17) \* (17 + 122)) / N // expected count

. scalar D = ((17 + 122) \* (109 + 122)) / N // expected count

. matrix FS = (A, B \ C, D) // matrix of expected values

. matrix rownames FS = Placebo AscorbicAcid // rownames

. matrix colnames FS = Cold NoCold // column names

. matrix list FS  
  
FS[2,2]  
 Cold NoCold  
 Placebo 24.086022 115.91398  
AscorbicAcid 23.913978 115.08602

# Chi-Square Test

. scalar chisquare = (31 - 24.086022)^2 / 24.086022 + ///   
> (109 - 115.91398)^2 / 115.91398 + ///  
> (17 - 23.913978)^2 / 23.913978 + ///   
> (122 - 115.08602)^2 / 115.08602

. scalar list chisquare  
 chisquare = 4.8114124

# Compare With Tabulate

. use "FrenchSkiiers.dta", clear

. tabulate Tx Outcome [fweight = Count], row col chi2  
  
┌───────────────────┐  
│ Key │  
├───────────────────┤  
│ frequency │  
│ row percentage │  
│ column percentage │  
└───────────────────┘  
  
 │ Outcome  
 Tx │ No Cold Cold │ Total  
──────────────┼──────────────────────┼──────────  
 Placebo │ 109 31 │ 140   
 │ 77.86 22.14 │ 100.00   
 │ 47.19 64.58 │ 50.18   
──────────────┼──────────────────────┼──────────  
Ascorbic Acid │ 122 17 │ 139   
 │ 87.77 12.23 │ 100.00   
 │ 52.81 35.42 │ 49.82   
──────────────┼──────────────────────┼──────────  
 Total │ 231 48 │ 279   
 │ 82.80 17.20 │ 100.00   
 │ 100.00 100.00 │ 100.00   
  
 Pearson chi2(1) = 4.8114 Pr = 0.028

# Risk Differences and Risk Ratios (Relative Risk)

Following Viera, 2008:

|  |  |  |
| --- | --- | --- |
|  | Develop Outcome | Do Not Develop Outcome |
| Exposed | a | b |
| Not Exposed | c | d |