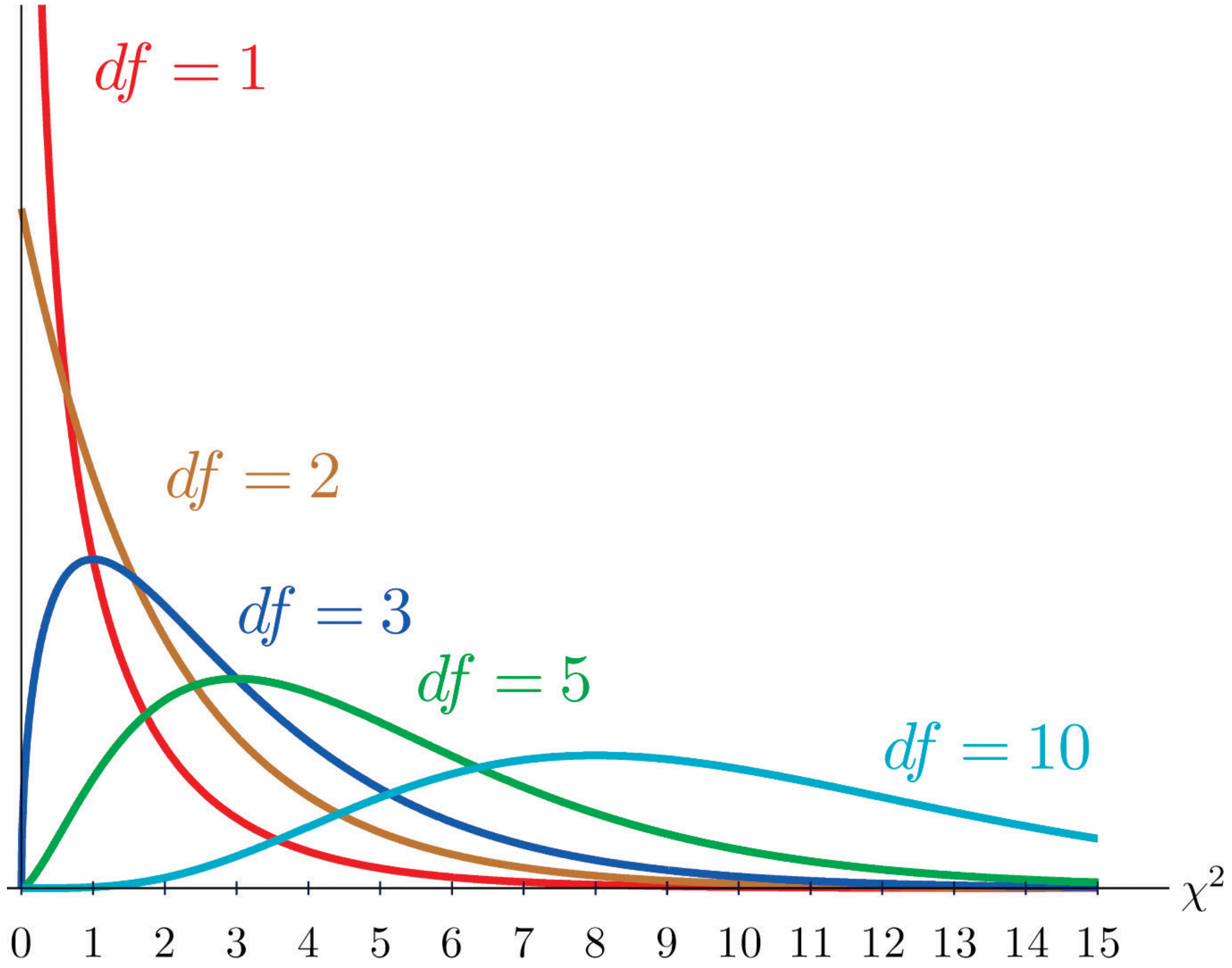
Foundational Statistics

Introduction to Frequency Analysis, Cont.

F1	PS	Ps	pS	ps
PS	PPSS	PPSs	PpSS	PpSs
Ps	PPSs	PPss	PpSs	Ppss
pS	PpSS	PpSs	ppSS	ppSs
ps	PpSs	Ppss	ppSs	ppss

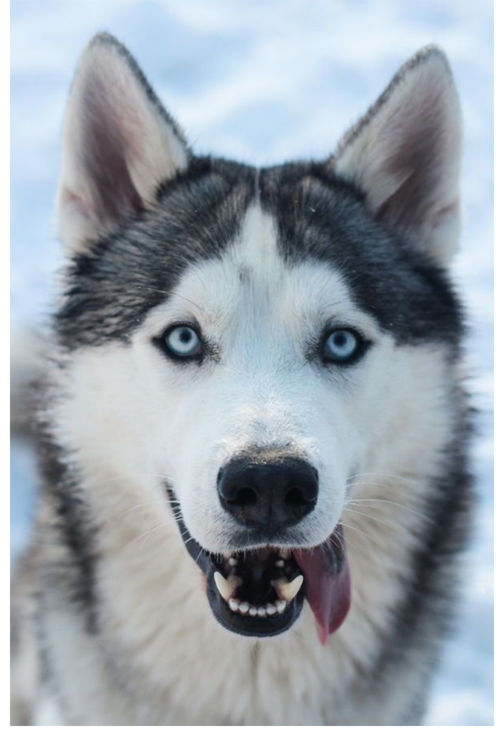
Plantes	Traitement 1	Traitement 2	$N_{k.}$
saines	21	24	44
peu infectées	10	13	23
infectées	11	10	21
très infectées	8	3	11
$N_{.l}$	50	50	100



Tests of independence

Given >1 categorical variable, each with >1 levels, do we observe “biases” in the level combinations?

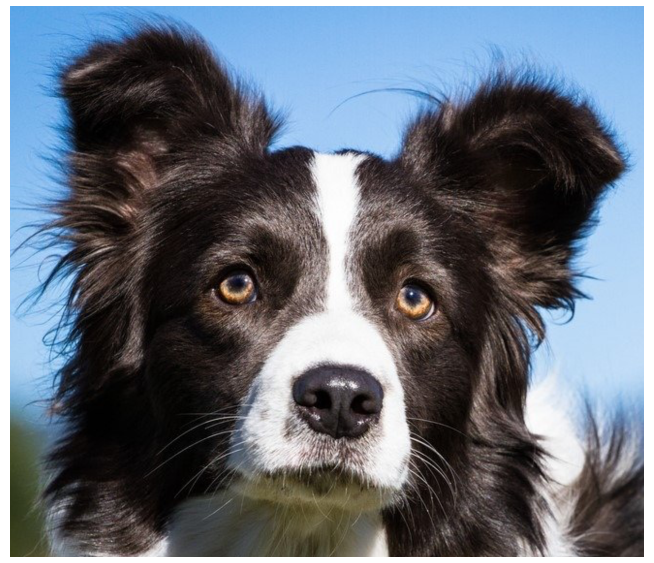
H_0: All categorical variables considered are independent.



prop. blue-eyed

n = 342 n = 187 529

Huskies: 0.647



Border collies: 0.217

n = 90 n = 324 414

432 511

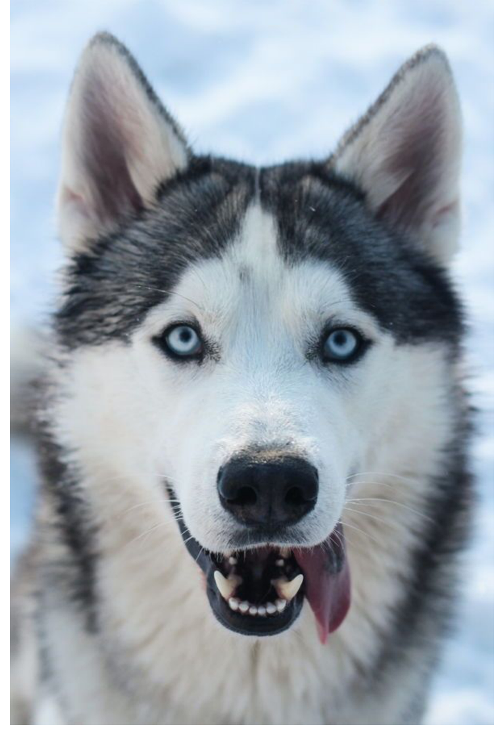
Cross classification (contingency) tables

Variable 2→ Variable 1↓	1	2	Marginal totals variable 1
1	n_{11} π_{11}	n_{12} π_{12}	n_{1j} π_{1j}
2	n_{21} π_{21}	n_{22} π_{22}	n_{2j} π_{2j}
Marginal totals variable 2	n_{i1} π_{i1}	n_{i2} π_{i2}	Grand total n

Tests of independence

Eye color (columns)

Dog breed (rows)



$n = 342$

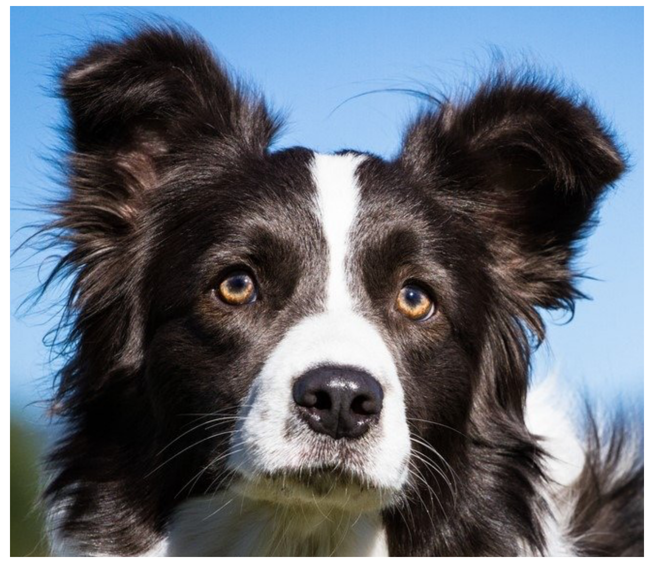
$$\frac{529 * 432}{943} = 242.34$$

$n = 187$

$$\frac{529 * 511}{943} = 286.66$$

$n = 90$

$$\frac{414 * 432}{943} = 189.66$$



$n = 324$

$$\frac{414 * 511}{943} = 224.34$$

row totals

529

414

column totals 432 511

943 = grand total

Expected Counts (Under H_0): $(r_{total} * c_{total}) / grand_{total}$

$$G^2 = 2 \sum o * ln(\frac{o}{e})$$

df:

$$(n_{rows} - 1) * (n_{columns} - 1)$$

Odds ratios: The effect sizes of freq. analysis

Odds: $\dfrac{P(event)}{1 - P(event)}$

***P*(heads) for fair coin is 0.5**

0.5 0.5

$$\frac{0.5}{1 - 0.5} = 1$$

Odds = 1 is called "even odds"
Equally likely to get a head vs. not head

Odds ratios quantify how much more likely an event is in one particular scenario vs another.

Getting heads with a fair coin vs. a "75% tail-heavy" coin:

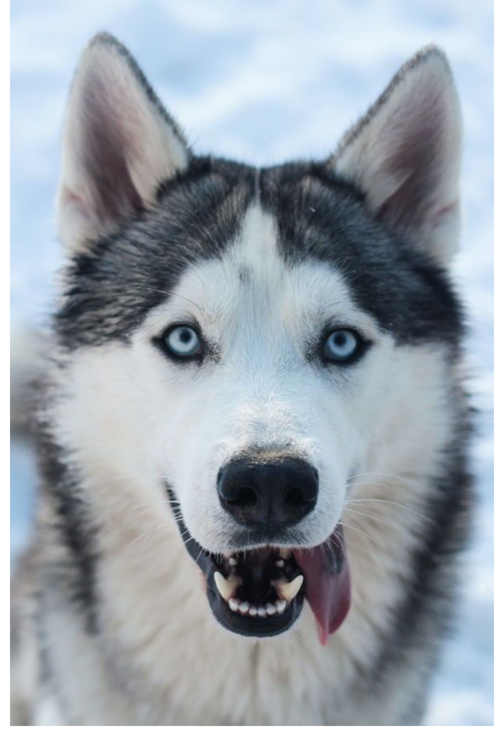
$$\frac{\frac{0.5}{1 - 0.5}}{\frac{0.25}{1 - 0.25}} = \frac{1}{\frac{1}{3}} = 3$$

You are 3x more likely, on average, to get heads with a fair coin than this biased coin

Odds ratios: The effect sizes of freq. analysis

Odds ratios can be calculated from contingency tables:

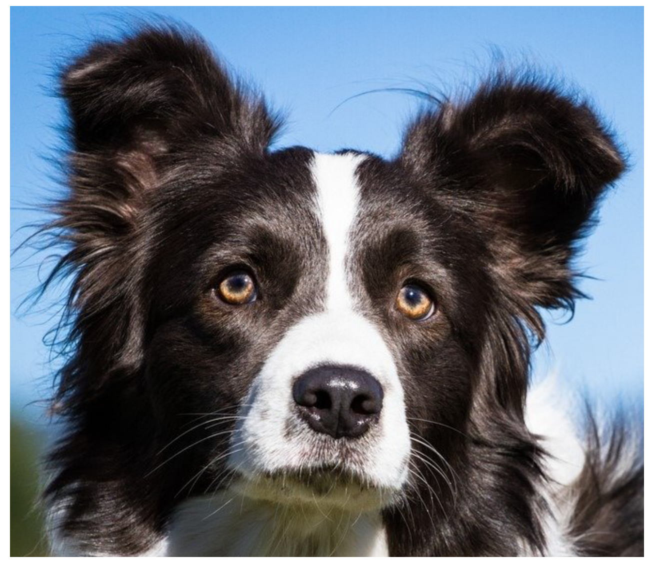
$$\theta = \frac{(cell_{1,1} + 0.5)(cell_{2,2} + 0.5)}{(cell_{1,2} + 0.5)(cell_{2,1} + 0.5)}$$



n = 342

n = 187

n = 90



n = 324

$$\theta = \frac{(342 + 0.5)(324 + 0.5)}{(187 + 0.5)(90 + 0.5)} = 6.55$$

~6.55 times more likely to have blue eyes if a husky than a Border collie

General Statistical Analysis Reporting: Differences, Directionality, and Magnitude

- Emphasize clearly the nature of differences or relationships.
- If you are testing for differences among groups, and you find a significant difference, it is not sufficient to simply report that “groups A and B were significantly different”. How are they different and by how much?
- It is much more informative to say “Group A individuals were 23% larger than those in Group B”, or, “Group B pups gained weight at twice the rate of Group A pups.”
- Report the direction of differences (greater, larger, smaller, etc) and the magnitude of differences (% difference, how many times, etc.) whenever possible.