# CSSS/SOC/STAT 536:
## Logistic Regression and Log-linear Models

# Log-linear Models of Contingency Tables: Multidimensional Tables

Christopher Adolph*
University of Washington, Seattle

March 10, 2005

*Assistant Professor, Department of Political Science and Center for Statistics and the Social Sciences.

# From 2D to 3D tables

The move from 2D to 3D tables adds several issues for loglinear models

- Tables are harder to visualize
- A plethora of interactions are estimable
- New notation is used to cope with growing complexity
- New concepts: *conditional independence*, *conditional association*, *collapsibility*
- Estimation may become difficult if tables have many zeros (curse of dimensionality)

Today, we'll discuss these challenges, but only scratch the surface of the literature

# Log-linear models for 3D tables

Let's start with the model under independence,

$$\mathrm{E}(\mu_{ijk}) = n\hat{\pi}_{i..}\hat{\pi}_{.j.}\hat{\pi}_{..k}$$

Take logs

$$\ln \mathrm{E}(\mu_{ijk}) = \ln n + \ln \hat{\pi}_{i..} + \ln \hat{\pi}_{.j.} + \ln \hat{\pi}_{..k}$$

Rewriting in log-linear form

$$\ln \mathrm{E}(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

# Log-linear models for 3D tables

The saturated model for a 3D table has every two-way *and* three-way interaction

$$\begin{aligned} \ln \mathrm{E}(\mu_{ijk}) &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z \\ &\quad + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ} \end{aligned}$$

(How can we think intuitively about these terms? What is a three-way interaction?)

As before, the saturated model perfectly fits the data, and has 0 degrees of freedom

# Identifying constraints

As in the $I \times K$ case, the parameters of the marginals must sum to 0 for identification

$$
\begin{aligned}
\sum_i \lambda_i^X &= 0 \\
\sum_j \lambda_j^Y &= 0 \\
\sum_k \lambda_k^Z &= 0
\end{aligned}
$$

Alternatively, we drop one element of each set of parameters from the model

That would leave $I-1$ $\lambda^X$'s, $J-1$ $\lambda^Y$'s, and $K-1$ $\lambda^Z$'s

The actual parameter estimates would look different,
but all quantities of interest would stay the same

# Identifying constraints

We also need to put identifying constraints on the interactions.

The interactions must sum to 0 over each *index*:

$$\sum_i \lambda_i^X = \sum_i \lambda_{ij}^{XY} = \sum_i \lambda_{ik}^{XZ} = \sum_i \lambda_{ijk}^{XYZ} = 0$$

$$\sum_j \lambda_j^Y = \sum_j \lambda_{ij}^{XY} = \sum_j \lambda_{jk}^{YZ} = \sum_j \lambda_{ijk}^{XYZ} = 0$$

$$\sum_k \lambda_k^Z = \sum_k \lambda_{ik}^{XZ} = \sum_i \lambda_{jk}^{YZ} = \sum_k \lambda_{ijk}^{XYZ} = 0$$

A different way of saying this is that $\lambda_{\cdot jk}^{XYZ} = 0$, $\lambda_{i \cdot k}^{XYZ} = 0$, etc.

# Identifying constraints

Once again, there are alternative identifying assumptions.

We could drop one row and column from each set of interactions, leaving:

| | |
|---|---|
| $(I-1) \times (J-1)$ | $\lambda^{XY}$ terms |
| $(I-1) \times (K-1)$ | $\lambda^{XZ}$ terms |
| $(J-1) \times (K-1)$ | $\lambda^{YZ}$ terms |
| $(I-1) \times (J-1) \times (K-1)$ | $\lambda^{XYZ}$ terms |

As before, this parameterization yields superficially different estimates,
but identical quantities of interest

# Concepts for 3D tables: Conditional independence

To wrap our minds around the implications of interaction effects, we define

**Conditional independence:**

$X$ and $Y$ are *conditionally independent given* $Z$ if they are independent within any partial table where $Z$ is fixed

We write this $X \perp Y | Z$

Note this is a weaker condition than joint independence (which is written $X \perp Y$)

# Concepts for 3D tables: Collapsibility

If a variable $X$ is jointly or conditionally independent of all other variables, we say the table is *collapsible* over $X$

- Partial tables that sum over $X$ will preserve all conditional relationships among other variables

- Put another way, we don't introduces Simpson's paradoxes by collapsing over $X$

Once again, we can understand this as omitted variable bias, which is harmless when the omitted variable is uncorrelated with other covariates

Collapsibility is very helpful in high dimensional and/or large tables

In a moment, we'll introduce a useful tool for investigating collapsibility

# Concepts for 3D tables: Conditional association

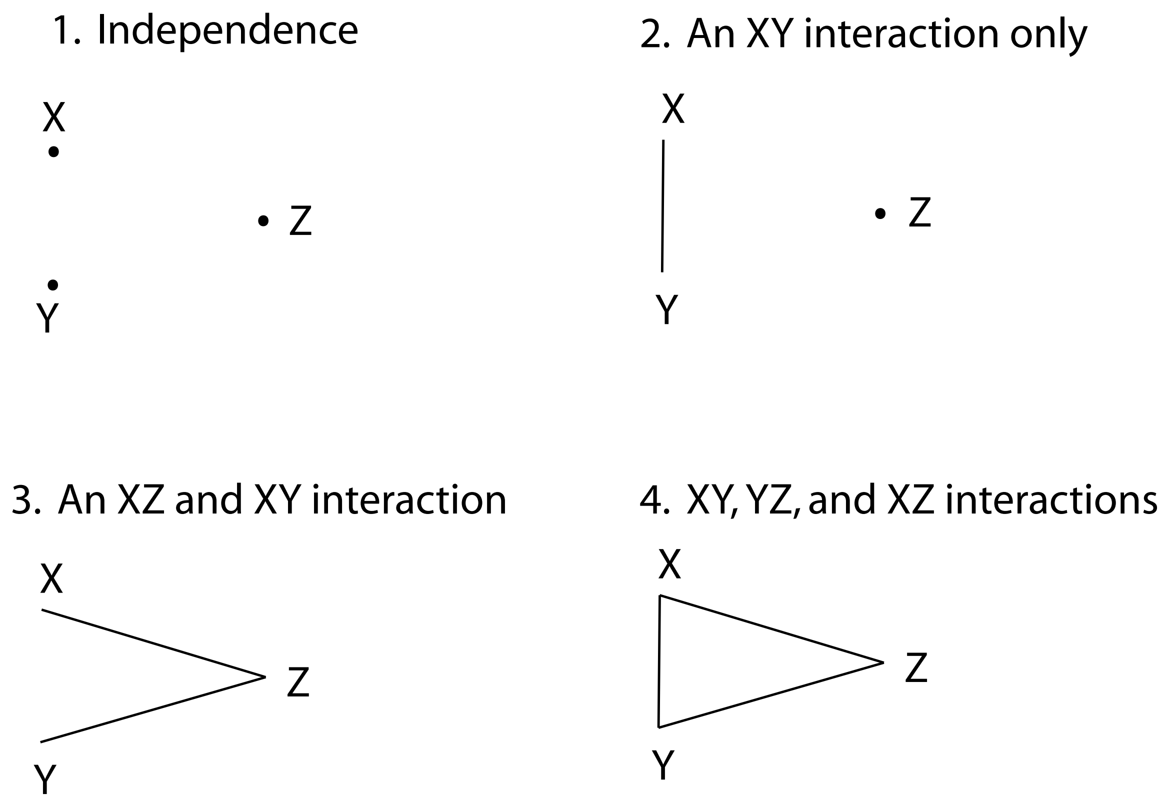
When two variables are *not* conditionally independent, they are *conditionally associated*

$\rightarrow$ That is, for fixed $Z = k$, $X$ and $Y$ are *not* independent

If the association between $X$ and $Y$ is the same for all $Z = k$, then we have homogenous association

There are tests for both conditional and homogenous association; see Agresti

# Concepts for 3D tables: Association Diagrams

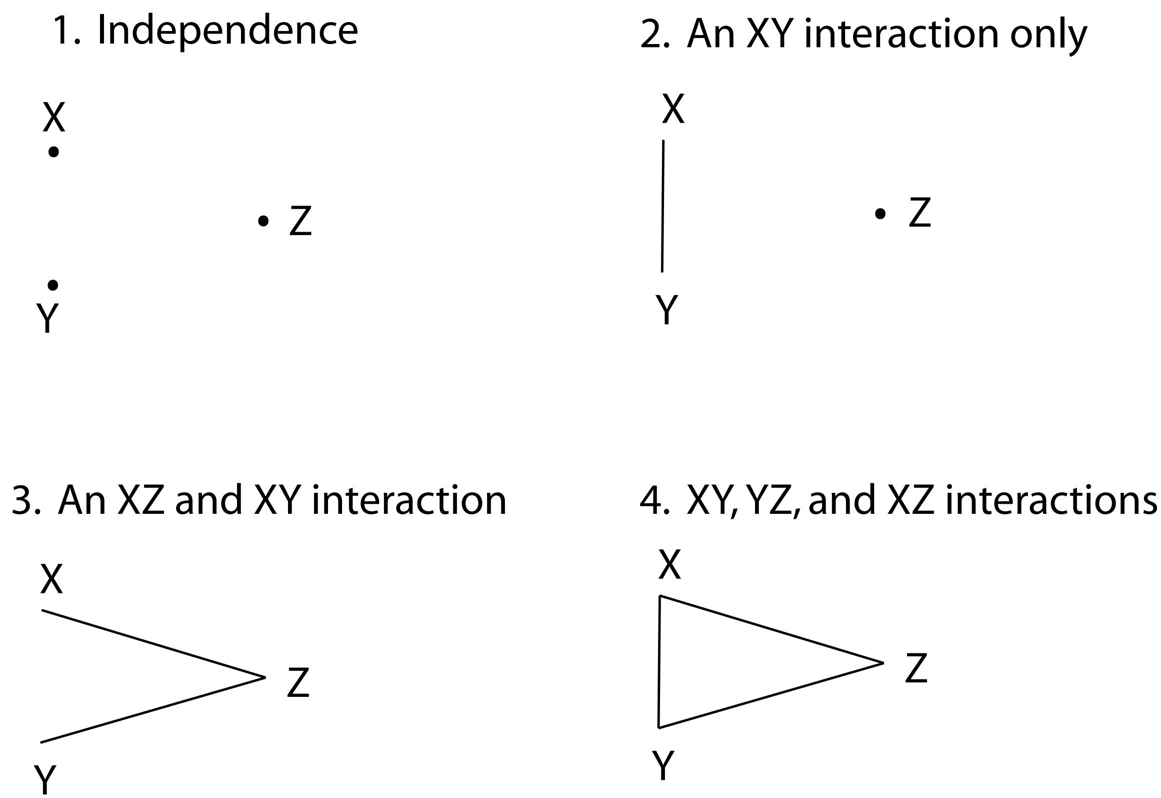



It helps to draw diagrams of which 2-way interactions are present

Variables are

- jointly independent if no *path* connects them
- conditionally associated if directly linked
- conditionally independent if linked by a chain

# Concepts for 3D tables: Association Diagrams

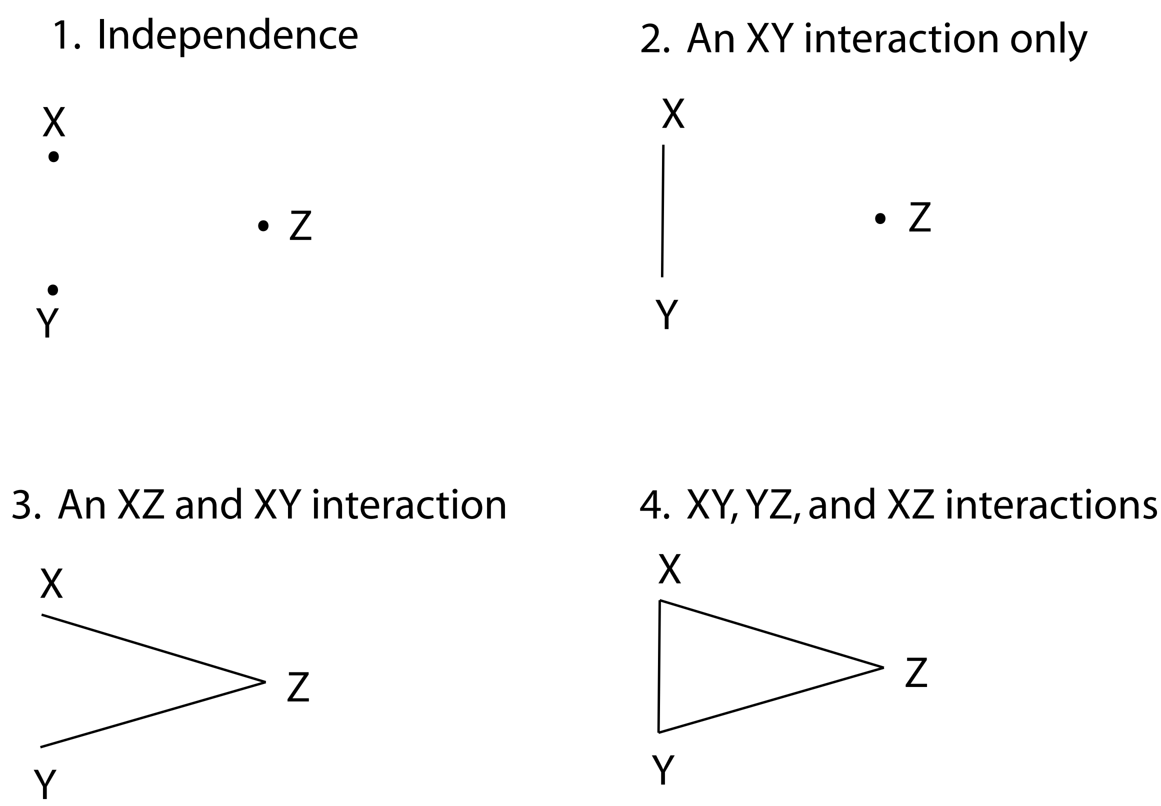



Several models can have the same association diagram

(4) gives both the saturated model,
and the saturated model minus the 3-way interaction

If there are multiple paths between two variables,
the variables are conditionally independent given any of the paths

# Concepts for 3D tables: Association Diagrams




Association diagrams are an easy way to check for collapsibility

- Obviously, singletons are collapsible (joint independence)
- So are variables at the ends of chains (conditional independence)
- But *not* variables between other variables in a chain (cond. assoc.)

# Concepts for 3D tables: Sufficient marginals

Under independence, we have the model

$$\ln \mathrm{E}(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

and the fitted cell probabilities are

$$\mathrm{E}(\pi_{ijk}) = \hat{\pi}_{i..}\hat{\pi}_{.j.}\hat{\pi}_{..k}$$

Now suppose that $X \perp Z$ and $Y \perp Z$, but $X$ and $Y$ are not independent. Then

$$\ln \mathrm{E}(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$$

and the fitted cell probabilities are

$$\mathrm{E}(\pi_{ijk}) = \hat{\pi}_{ij.}\hat{\pi}_{..k}$$

# Concepts for 3D tables: Sufficient marginals

If $X$ and $Y$ are conditionally independent given $Z$, we write the LLM

$$\ln \mathrm{E}(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

and the fitted cell probabilities are

$$\mathrm{E}(\pi_{ijk}) = \hat{\pi}_{i \cdot k} \hat{\pi}_{\cdot jk}$$

Finally, in the saturated model, the fitted cell probabilities are the observed probabilities,

$$\mathrm{E}(\pi_{ijk}) = \hat{\pi}_{ijk}$$

# Concepts for 3D tables: Sufficient marginals

In each of the above examples,
the probabilities on the right-hand-side are the "sufficient marginals"

They are the only proportions we need to observe to estimate the model

We could throw the other cells away, as we did in our $2 \times 2$ example of independence

As the model gets more complicated,
the sufficient marginals include more and more of the actual cells

# Notation for 3D tables

It helps to define a more compact notation for multidimensional LLMs

| Model ($\ln \mathrm{E}(\mu_{ijk}) = \ldots$) | Notation | Assumptions |
|---|---|---|
| $\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$ | $(X, Y, Z)$ | $X \perp Y \perp Z$ |
| $\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$ | $(XY, Z)$ | $X \perp Z, Y \perp Z$ |
| $\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$ | $(XY, XZ)$ | $Y \perp Z \vert X$ |
| $\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{ijk}^{XYZ}$ | $(XYZ)$ | |

Note this notation assumes lower level terms are *included*

Models that nest all lower order terms sometimes referred to as “hierarchical”

This is a helpful simplifying assumption, but not *logically* required

# Fussy points on interaction terms

There is a case where you could omit a lower order term:

1. Theoretically, it should be zero

2. Empirically, it is zero

3. Omitting it improves precision a worthwhile amount

4. You show (1–3) to the referees' satisfaction. . .

A common example from linear regression, if we think fertility depends on GDP/population, but not GDP, we don't insist on including 1/GDP as a "main effect"

Omitting a relevant lower order term will badly bias your result

So will omitting an interaction, though not as badly

(Sidenote: "main effect" is a misleading term. If interactions are present, they are part of the effect)

# Estimating LLM for 3D tables

Estimation of 3+ dimensional tables uses the same MLE as 2D tables

We could use `loglm` on the table directly

Or rewrite as a series of observations with categorical explanatory variables, and estimate with `optim`

Practically, high dimensional tables can be hard to estimate well, because the data are spread out over more parameters (sparseness)

If any marginal statistics have zero sum, the corresponding parameters are not estimable

What does this mean? Under independence, the row totals must all be positive

Given an interaction, some cells may need to be positive as well

Why? Although Poisson *counts* can be 0,
fitted values from $\exp(\lambda)$, $\lambda > -\infty$ must be positive

# Fitting LLM for 3D tables

As with 2D tables, we could run every specification and compare deviances or BICs

This quickly becomes tedious. We could automate the search (usually unwise)...

Or start with a reasonable complex specification and see what terms can be discarded cheaply

But don't be too credulous of $p$-values when stepping through specifications

As with 2D tables, you can often code creative new variables that are potentially more enlightening

# Interpreting parameters

Interpreting the parameters of LLMs with interactions is more difficult

As with 2D tables, we could interpret the parameters as changes in log odds

(See Agresti for examples and discussion)

I would prefer fitted values, first differences, and graphical displays

Key step in modelling a high dimensional data set:
reducing the patterns in the data to simple, mostly accurate statements.

Just fitting an LLM is usually not sufficient;
you also need to extract summary results from it

# Example: Military Deaths

We examine the deaths of US military personnel in 1999 (the example is from Simonoff 2003)

The data come in a $5 \times 4 \times 2$ table:

| Gender | Manner of death | Army | Navy | Air Force | Marines |
|---|---|---|---|---|---|
| Male | Accident | 158 | 96 | 54 | 69 |
| | Illness | 56 | 30 | 21 | 6 |
| | Homicide | 11 | 7 | 2 | 6 |
| | Suicide | 51 | 32 | 13 | 7 |
| | Unknown | 24 | 9 | 13 | 27 |
| Female | Accident | 15 | 9 | 7 | 3 |
| | Illness | 5 | 4 | 4 | 0 |
| | Homicide | 5 | 2 | 0 | 1 |
| | Suicide | 3 | 3 | 1 | 0 |
| | Unknown | 1 | 2 | 2 | 2 |

Note there is no row for combat deaths (there weren't any). Extrapolating from this table to, say, 2004 would be dangerous.

# Example: Military Deaths

We want to know if there are any service or gender related patterns in the different death rates

Clearly, we need the base terms (more men than women in the data, and more Army than Marines)

But what interactions should we include? Let's explore combinations of

($S =$ Service, $G =$ Gender, $M =$ Manner of death)

Note: Simonoff judged models using a corrected AIC$_c$ test
I'm taking his published AICs at face value

We'll also look at BIC

# Example: Military Deaths

The most complicated feasible model is

| Model | df | $G^2$ | BIC | AIC$_c$ |
|---|---|---|---|---|
| $(SM,SG,MG)$ | $12$ | $5.5$ | $-74.1$ | $1.93$ |

Can we simplify this?

# Example: Military Deaths

The next step is to drop each two-way interaction in turn

| Model | df | $G^2$ | BIC | AIC$_c$ |
|---|---|---|---|---|
| $(SM,SG,MG)$ | 12 | 5.5 | −74.1 | 1.93 |
| $(SM,SG)$ | 16 | 14.2 | −92.0 | 1.99 |
| $(SM,MG)$ | 15 | 10.7 | −88.8 | 0.63 |
| $(SG,MG)$ | 24 | 61.4 | −97.8 | 32.25 |

(What do these models mean substantively?)

Our atheoretical exploration has hit a snag:

AIC$_c$ and BIC disagree about the best choice.

Simonoff was using AIC$_c$, and proceeded with $(SM,MG)$

(He made the substantive conclusion that Gender was conditionally indep of Service)

We will follow his path for now

# Example: Military Deaths

Next, we try making one of the variables jointly independent

| Model | df | $G^2$ | BIC | $\text{AIC}_c$ |
|---|---|---|---|---|
| $(M,SG)$ | 28 | 69.2 | 31.83 | $-116.6$ |
| $(G,SM)$ | 19 | 18.6 | 0.00 | $-107.5$ |
| $(S,MG)$ | 27 | 65.7 | 30.40 | $-113.4$ |

Once again, AIC and BIC disagree.

Simonoff concluded $(G,SM)$ was the best model. Implications:

- Men die more often than women, but of the same causes and in similar ratios by service

- Cause of death varies by service (less suicide but more accidents in Marines)

BIC picks $(M,SG)$ by a hair. Were Simonoff's conclusions wrong?

Or do the data even allow us to reject any of the options?

# Example: Military Deaths

Neither: we've made two (related) common mistakes.

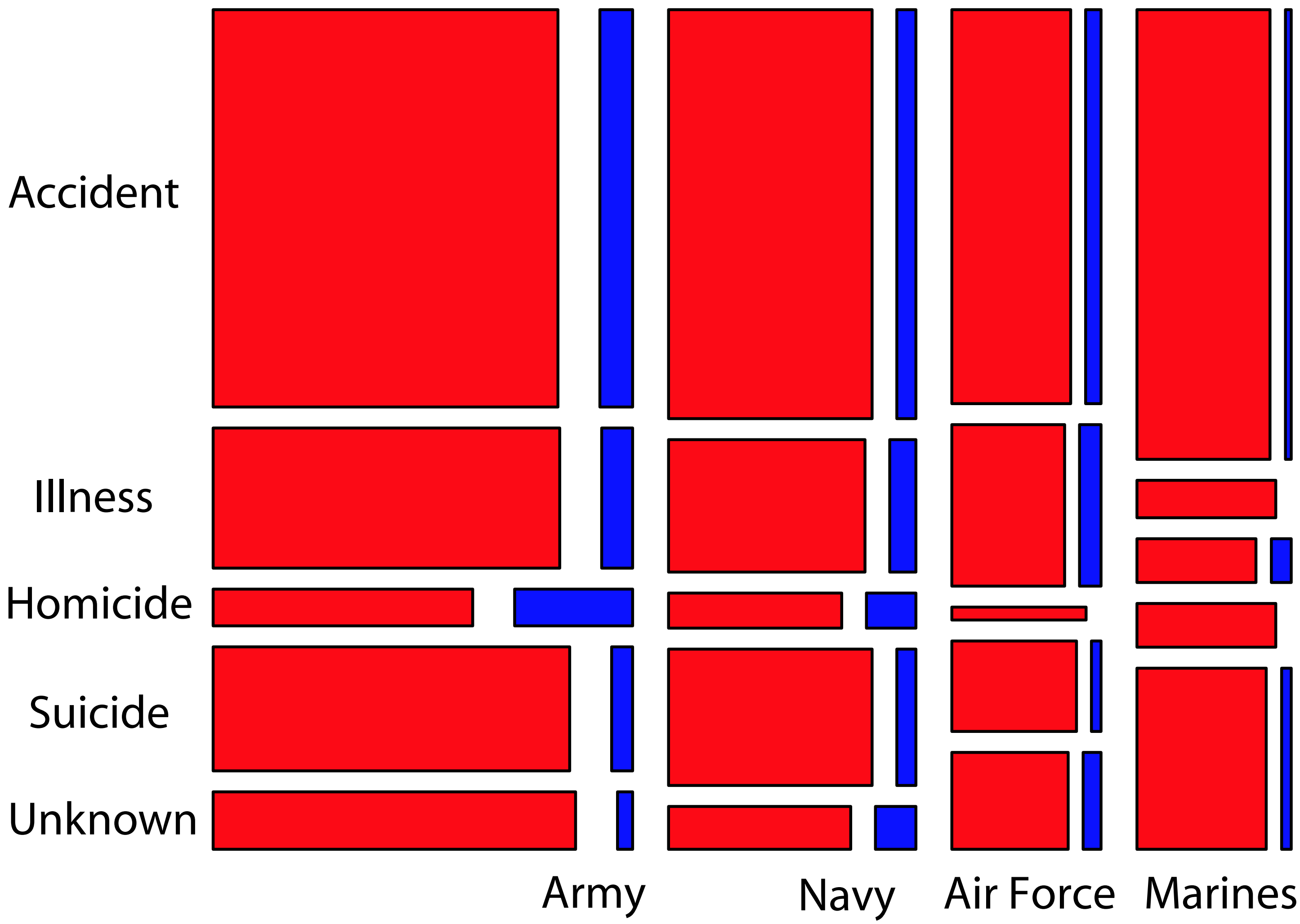
1. We've confused statistical and substantive significance

2. We've dichotomized results into "effect/no effect"

Effects could be confined to particular categories, or be very small but "significant"

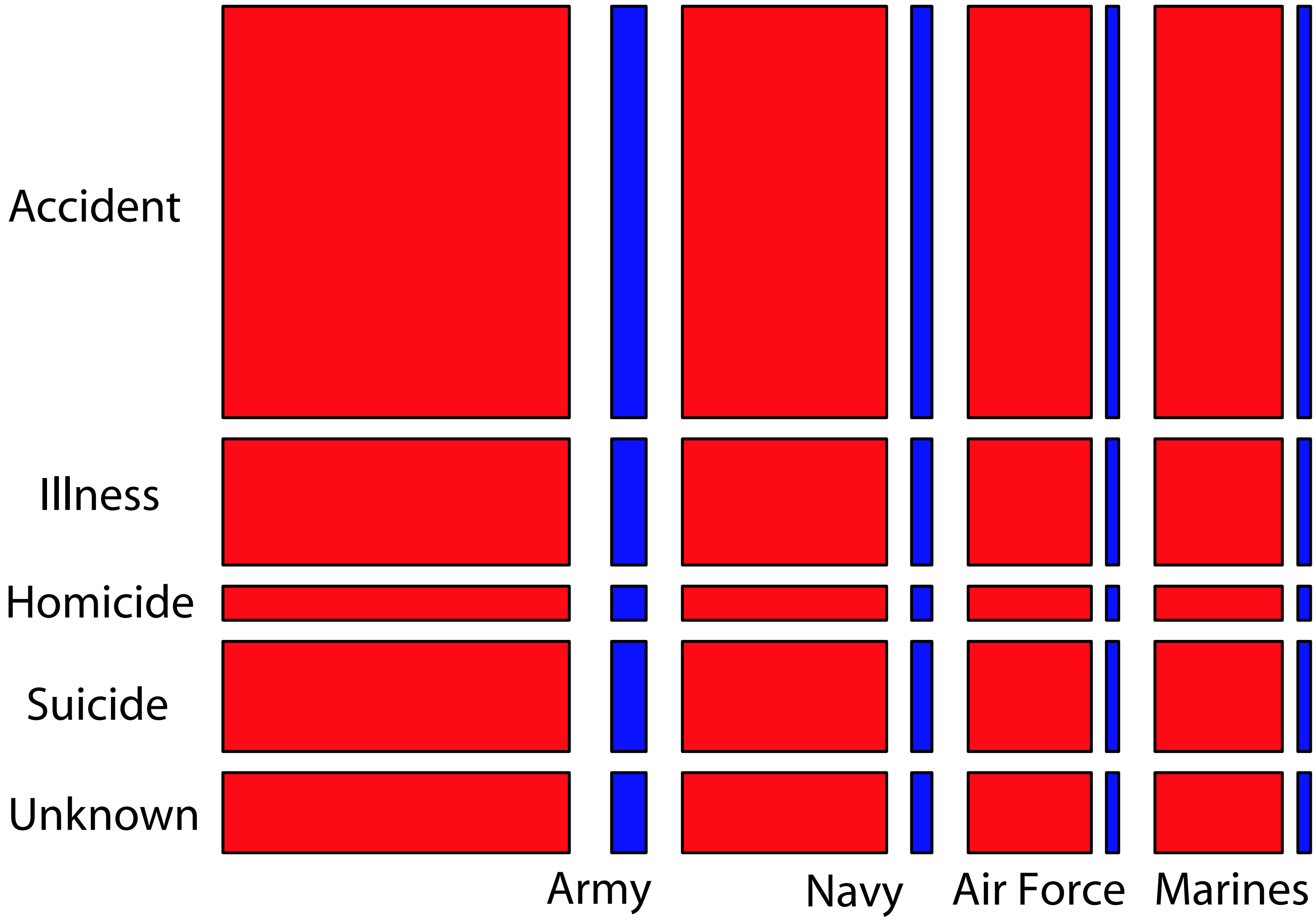
Only way to tell is to look at the model predictions

# Military deaths data (Females in Blue)



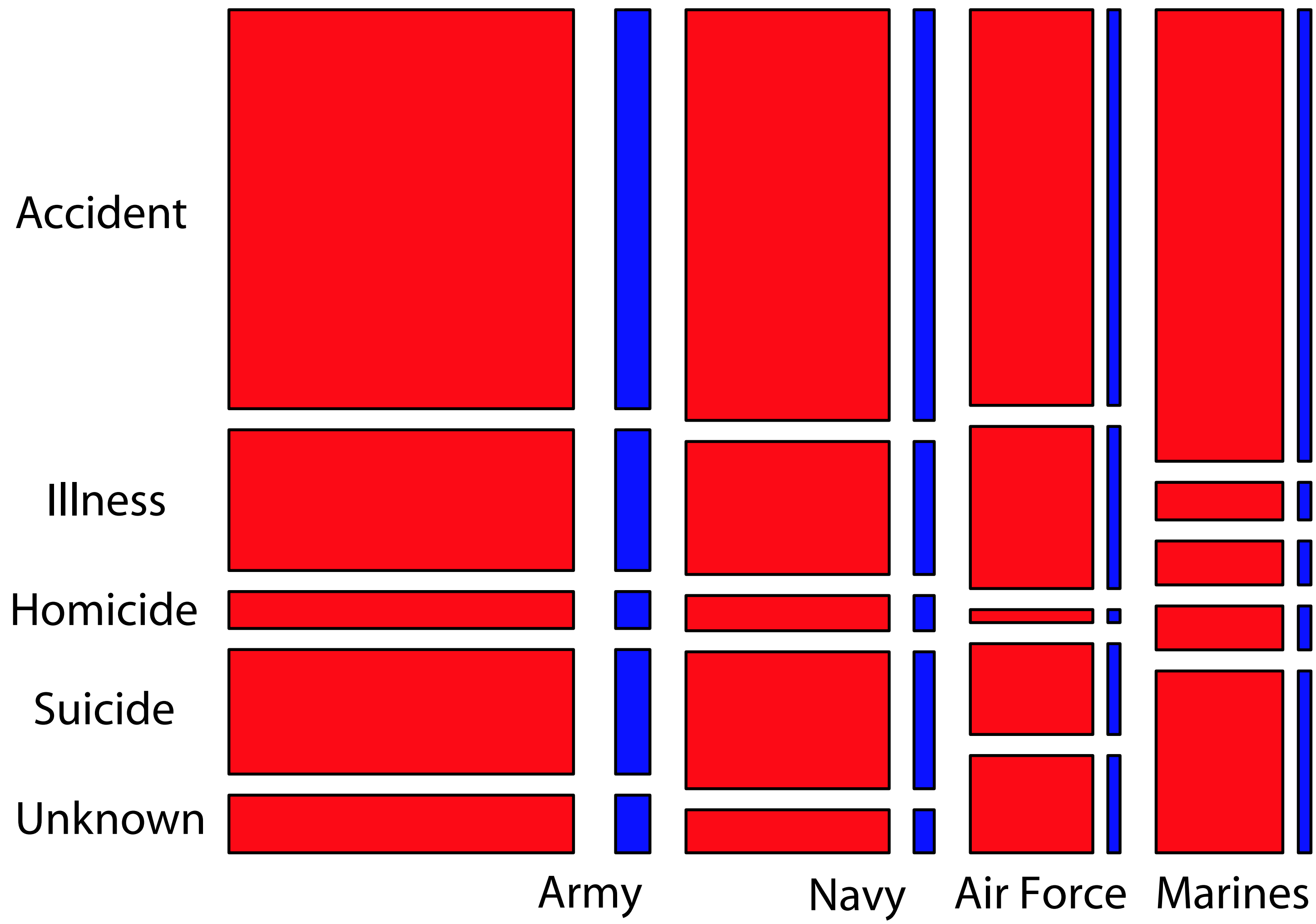

These are the observed counts. We'd like to reproduce interesting patterns here or convince ourselves they were spurious

## Military deaths: Fitted proportions under independence



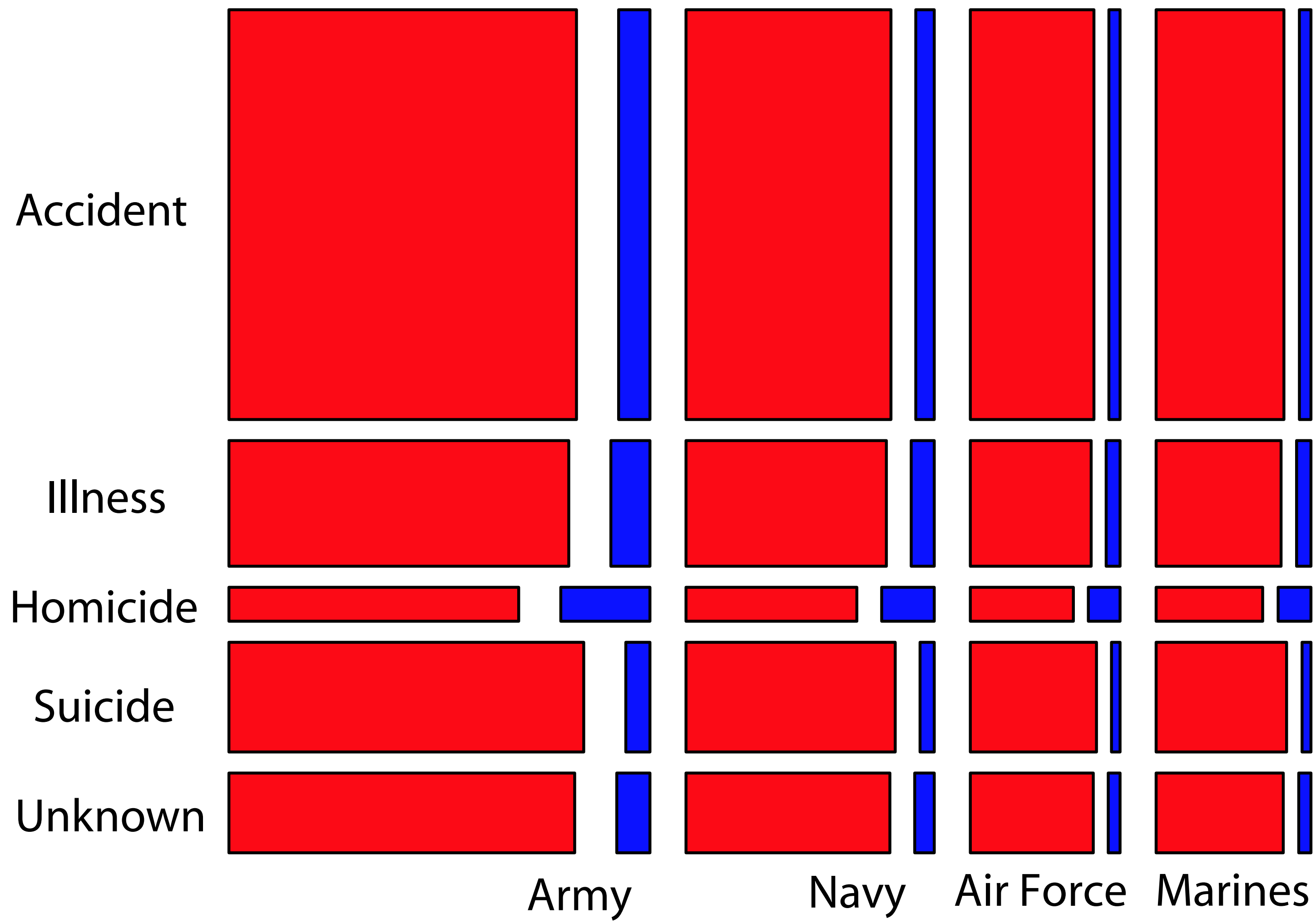

This is the model under independence. It's not horrible (interactions must be weak) but maybe those weak interactions are still non-zero

# Military deaths: Fitted proportions under ($G$,$SM$)



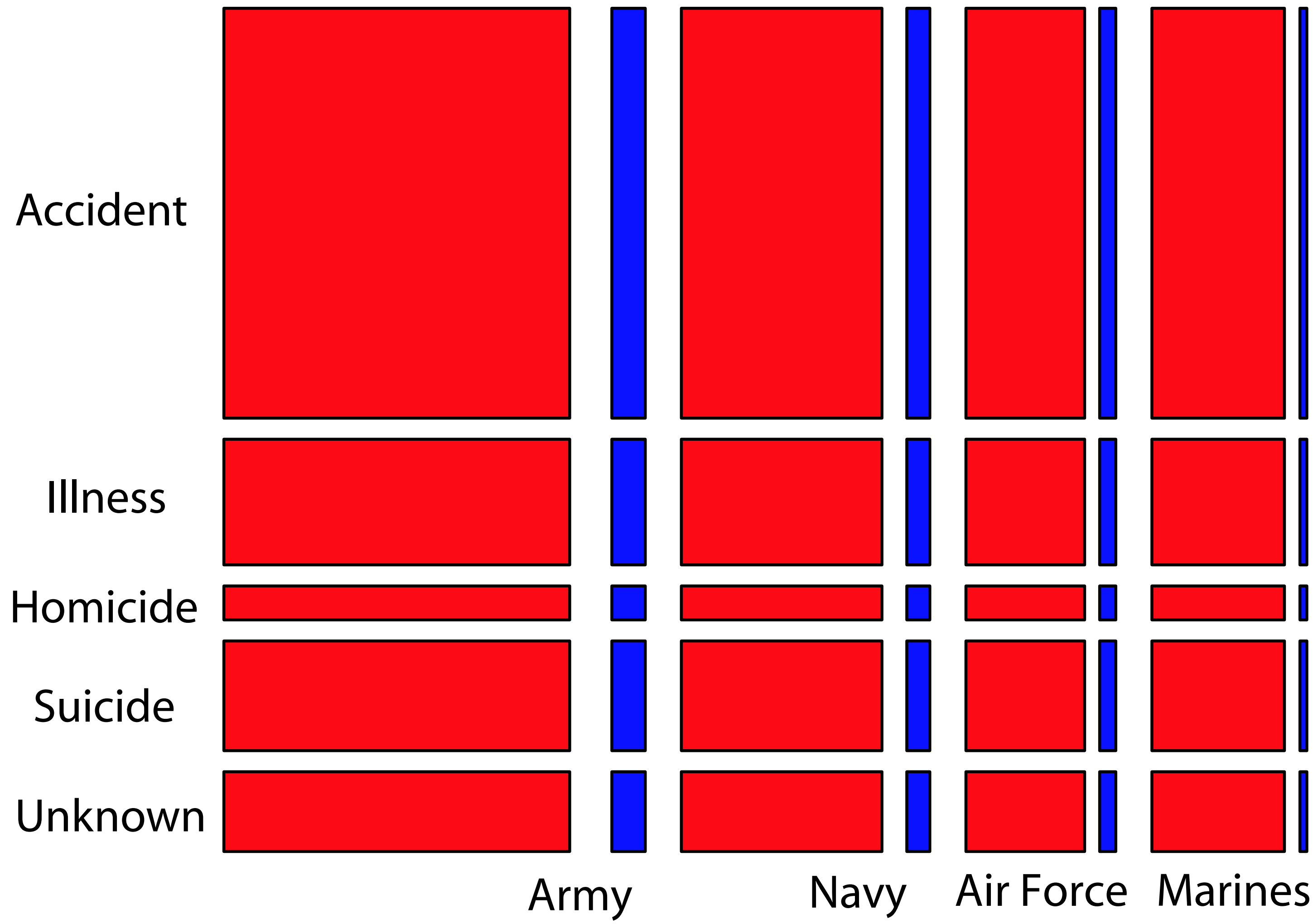

Simonoff's preferred model: Holding Gender independent, we see the Marines have a different pattern of deaths

# Military deaths: Fitted proportions under ($S$,$GM$)




Looks like $(G,SM)$ missed something important: Holding Service independent, we see that women are at higher risk of homicide

# Military deaths: Fitted proportions under ($M$,$GS$)




Holding manner of death independent, we see women are at lower risk in the Marines
Note that no Marine women were victims of homicide

# Example: Military deaths

The fitted models reveal two interesting interactions:

1. gender $\times$ manner: women are relatively more likely to be homicide victims

2. manner $\times$ service: fewer suicides in Marines, and more accidents

Otherwise, gender and service made little difference wrt manner of death

So substantively, $(SM,GM)$ may be the best model, or best place to start

Interestingly, this was the model BIC pushed for in round 2

Probably best to view this exercise as exploratory.

Follow up with models of homicide, accidents, etc,
perhaps getting data by subunits, or time series data

(Elephant in the middle of the room: "Unknown deaths"?
If these are mostly suicides, $MS$ interaction mostly vanishes)

# Concluding thoughts on tabular data

Advantages of tabular presentation:

- Compact presentation of data
- FWIW, allows stepwise investigation of all interactive models
- Compact presentation of result
- Sometimes, ambiguity b/w independent and dependent variables helps

Disadvantages of tabular presentation:

- Not all data fit. Never shoehorn continuous variables in (throwing away data)
- May discourage transformation of data/clear presentation of effect sizes
- Language of LLMs is increasingly unfamiliar in many fields

Remember there is always an equivalent approach using the standard Poisson or NegBin setup

Remember to check for overdispersion