- Example 2.
  - ➢ scenario
    - observe $Y_{\mathbf{x},i} \in r$ categories $\{1, 2, \ldots, r\}$, $i=1, \ldots, K_{\mathbf{x}}$, where $K_{\mathbf{x}}$ is a random variable on the non-negative integers.

do not know how many categorical responses can be collected before data collection stage (example?)

    - assume $K_{\mathbf{x}} \sim Poisson(\lambda_{\mathbf{x}})$ and given (conditioned on) $K_{\mathbf{x}}=k_{\mathbf{x}}$, (a fixed value)
      - □ $Y_{\mathbf{x},1}, Y_{\mathbf{x},2}, \ldots, Y_{\mathbf{x},k_{\mathbf{x}}}$ are independent and identically distributed from $multinomial(1; p_{\mathbf{x},1}, \ldots, p_{\mathbf{x},r})$, where
      - □ $p_{\mathbf{x},j}=P(\text{observing category } j)$, $j=1,\ldots, r$, and $p_{\mathbf{x},1}+\cdots+p_{\mathbf{x},r}=1$.

often used to model the number of some event occurred during a period of time

  - ➢ sufficient statistics

joint → $P(Y_{\mathbf{x},i}\text{'s}, K_{\mathbf{x}}) = \underbrace{P(K_{\mathbf{x}})}_{\text{marginal}} \cdot \underbrace{P(Y_{\mathbf{x},i}\text{'s} \mid K_{\mathbf{x}})}_{\text{conditional}}$

    - the joint pmf of $Y_{\mathbf{x},1}, Y_{\mathbf{x},2}, \ldots, Y_{\mathbf{x},K_{\mathbf{x}}}$ is ← random

$$\frac{e^{-\lambda_x}\lambda_x^{K_x}}{K_x!}\times p_{x,1}^{N_{x,1}}\cdots p_{x,r}^{N_{x,r}} \propto \frac{e^{-\lambda_x p_{x,1}}(\lambda_x p_{x,1})^{N_{x,1}}}{N_{x,1}!}\times\cdots\times\frac{e^{-\lambda_x p_{x,r}}(\lambda_x p_{x,r})^{N_{x,r}}}{N_{x,r}!}$$

$\lambda_x = \lambda_{\mathbf{x}}(p_{\mathbf{x},1}+\cdots+p_{\mathbf{x},r})$

count; Poisson pmf

where $N_{\mathbf{x},j}$ = number of category j in the $K_{\mathbf{x}}$ trials, $j=1, \ldots, r$, and $K_{\mathbf{x}}=N_{\mathbf{x},1}+\cdots+N_{\mathbf{x},r}$.

numerical data; can be treated as a predictor with r levels

can understand how $\mathbf{x}$ influence $\lambda_{\mathbf{x}}$ & $p_{\mathbf{x},i}$'s in analysis.

    - by factorization thm, $(N_{\mathbf{x},1}, \ldots, N_{\mathbf{x},r})$ are the sufficient statistics for the parameters $(\lambda_{\mathbf{x}}p_{\mathbf{x},1}, \ldots, \lambda_{\mathbf{x}}p_{\mathbf{x},r})$ and (dim = r) important information

$p_{\mathbf{x},i} = \dfrac{\lambda_{\mathbf{x}}p_{\mathbf{x},i}}{\lambda_{\mathbf{x}}p_{\mathbf{x},1}+\cdots+\lambda_{\mathbf{x}}p_{\mathbf{x},r}}$

check Notes in LNp.2-26

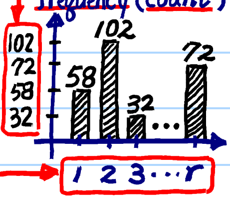
- $N_{\mathbf{x},1}, \ldots, N_{\mathbf{x},r}$ are independent and $N_{\mathbf{x},j} \sim Poisson(\lambda_{\mathbf{x}}p_{\mathbf{x},j})$.

- In the 2 examples, $N_{\mathbf{x}}$'s

apply statistics (+, −, ×, ÷, …) on $N_{\mathbf{x}}$'s, not directly on $Y_{\mathbf{x}}$'s

$Y_{\mathbf{x}}$'s → categorical response (symbols) → counts (numeric) → statistics (numeric calculation) → Information

digitization (check table in LNp.2-4)



Some real-data examples (from Agresti, *Categorical Data Anslysis*)

prospective study

- From Agresti (2002, 2rd ed.)

resp. (M.I) $\boxed{Y}$; expl. (Aspirin) $\boxed{X}$

1st person	Yes	No
2nd	No	⋮
⋮	⋮	Yes
1360th	⋮	⋮

sufficient statistics → $\boxed{N}$ → count

2×2 contingency table

TABLE 3.1 Swedish Study on Aspirin Use and Myocardial Infarction (Yes/No; Yes/No)

	Myocardial Infarction		
	Yes	No	Total
Placebo	28	656	684
Aspirin	18	658	676

28 + 656 = 684; 18 + 658 = 676 } fixed ($k_{\mathbf{x}}$)

from DOE

resp. (M.I)		expl. (Aspirin)
Yes, No		
(N_{11}, N_{12})		No (1)
(N_{21}, N_{22})		Yes (2)

$P(\text{M.I - Yes})$

Note. $P_1 + P_2 \neq 1$

N_{11} ($X=1$) ~ binomial(684, P_1)

N_{21} ($X=2$) ~ binomial(676, P_2)

Q: how X (1→2) influence P_i's?

Both X & Y have 2 categories. does not matter they are nominal, ordinal, or discrete interval

- From Agresti (2013, 3rd ed.)

3-way contingency table

Table 6.9 Clinical Trial Relating Treatment to Response for Eight Centers, with Expected Value and Variance (of Success Count for Drug) Under Conditional Independence

$N_{T,C,1}$ $N_{T,C,2}$ suppose fixed ($k_{T,C}$)

Center	Treatment	Response: Success	Response: Failure		Odds Ratio	μ_{11k}	var(n_{11k})
1	Drug	11	+ 25	= 36	1.19	10.36	3.79
	Control	10	+ 27	= 37			
2	Drug	16	+ 4	= 20	1.82	14.62	2.47
	Control	22	+ 10	= 32			
3	Drug	14	5		4.80	10.50	2.41
	Control	7	12				
4	Drug	2	14		2.29	1.45	0.70
	Control	1	16				
5	Drug	6	11		∞	3.52	1.20
	Control	0	12				
6	Drug	1	10		∞	0.52	0.25
	Control	0	10				
7	Drug	1	4		2.0	0.71	0.42
	Control	1	8				
8	Drug	4	+ 2	= 6	0.33	4.62	0.62
	Control	6	+ 1	= 7			

a 2×2 contingency table

different

$N_{T,C,1} \sim$ binomial($k_{T,C}$, $P_{T,C}$)

fixed known value — parameter

Q: how (T, C), including their main effects & interactions, influence $P_{T,C}$'s?

$\boxed{Y}$ resp. (Success)	2 levels	expl. $\boxed{X}$ (Treatment)	2 levels	(Center)	nominal
1st	Yes		Drug		1
2nd	Yes		Control		1
⋮	⋮		⋮		⋮
⋮	No		Drug		8
⋮	⋮		Control		8

sufficient statistics $\longrightarrow$

count $\boxed{N}$ resp. (Success) Yes, No	2 levels	expl. $\boxed{X}$ (Treatment)	2 levels	(Center)	nominal
(N_{111}, N_{112})		Drug		1	
(N_{211}, N_{212})		Control		1	
⋮		⋮		⋮	
(N_{181}, N_{182})		Drug		8	
(N_{281}, N_{282})		Control		8	