Taking a breath

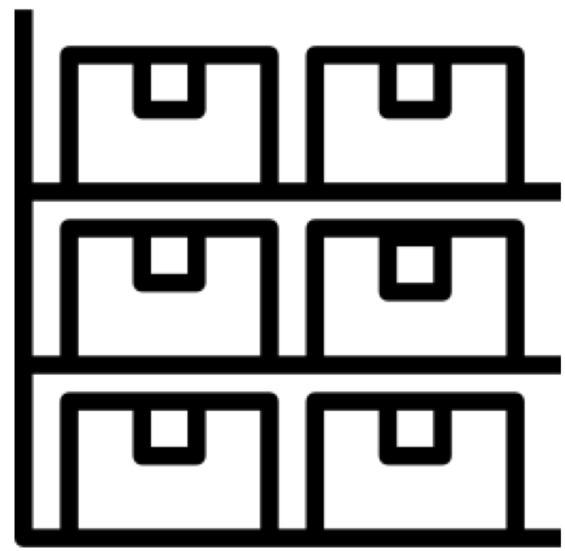
Congratulations!

- You've survived the first month and a half.
- Now you know GLMs!
  - Linear Regression
  - Logistic Regression
  - Poisson Regression
- You can
  - Fit the models and interpret
  - Look at residuals to diagnose potential problems
  - Make inference on the coefficients to answer questions
- One more important GLM that we need to talk about.

Where you are headed

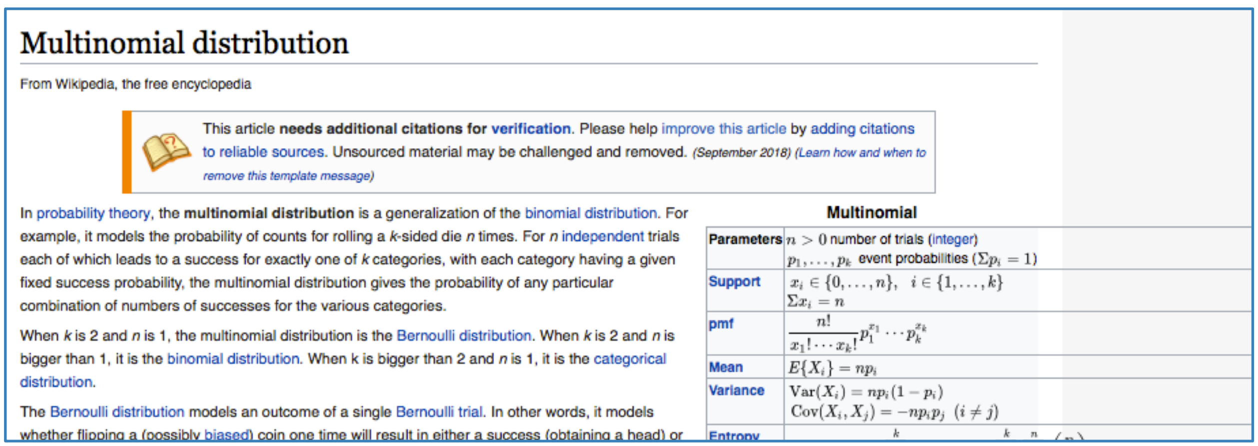
- Stage 1: Understanding
  - Interpret fitted model
  - Identify “significant” coefficients
  - See when and where a fitted model has problem and fix it
- Stage 2: **Using** the fitted models to answer questions
  - Identify the question of interest and map it to visualization / model
  - Use necessary statistics / graphics to make sufficient argument
    - Translate “significance” to meaningful statement.
    - Assess predictive performance
    - Use simulation to answer complex questions
- Stage 3: Incorporate different sources of uncertainty
- Stage 4: Causal inference

Categorical Outcomes



categorical regression

- Generalization of Logistic regression
  - Logistic regression: 0 or 1
  - Multinomial regression: 010000, 000100,..
- Not so easy to interpret.
- Let’s start by reviewing the probability distribution.



Multinomial distribution

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(September 2018) (Learn how and when to remove this template message)*

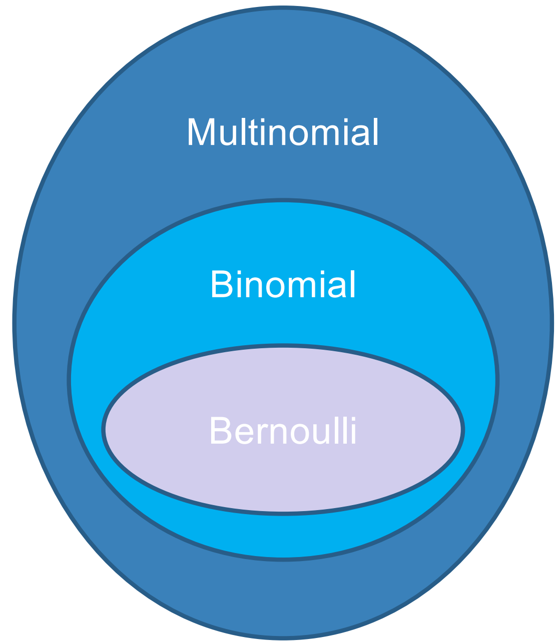
In probability theory, the **multinomial distribution** is a generalization of the binomial distribution. For example, it models the probability of counts for rolling a k-sided die n times. For n independent trials each of which leads to a success for exactly one of k categories, with each category having a given fixed success probability, the multinomial distribution gives the probability of any particular combination of numbers of successes for the various categories.

When k is 2 and n is 1, the multinomial distribution is the Bernoulli distribution. When k is 2 and n is bigger than 1, it is the binomial distribution. When k is bigger than 2 and n is 1, it is the categorical distribution.

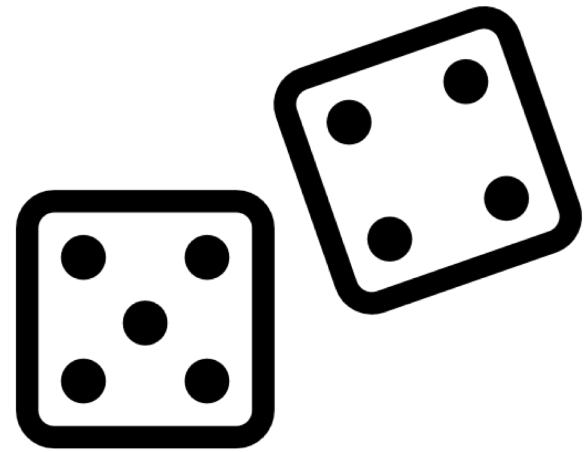
The Bernoulli distribution models an outcome of a single Bernoulli trial. In other words, it models whether flipping a (possibly biased) coin one time will result in either a success (obtaining a head) or

Multinomial

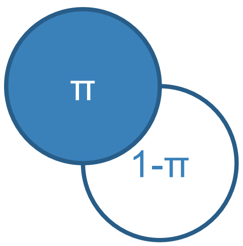
Parameters	$n > 0$ number of trials (integer) $p_1, \ldots, p_k$ event probabilities ($\Sigma p_i = 1$)
Support	$x_i \in \{0, \ldots, n\}, \quad i \in \{1, \ldots, k\}$ $\Sigma x_i = n$
pmf	$\frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$
Mean	$E\{X_i\} = np_i$
Variance	$\mathrm{Var}(X_i) = np_i(1 - p_i)$ $\mathrm{Cov}(X_i, X_j) = -np_i p_j \quad (i \neq j)$
Entropy	

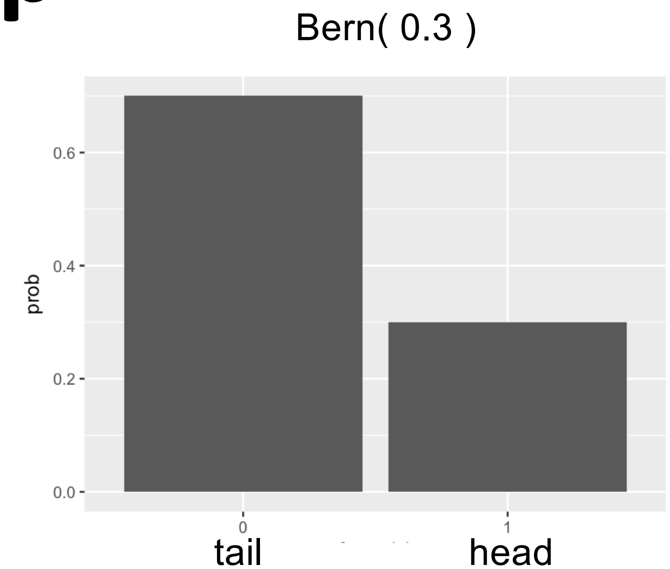


Multinomial Distribution



Bernoulli(π): A coin flip

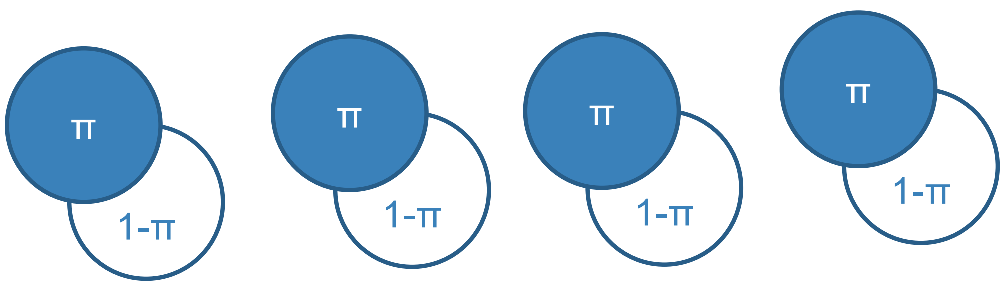


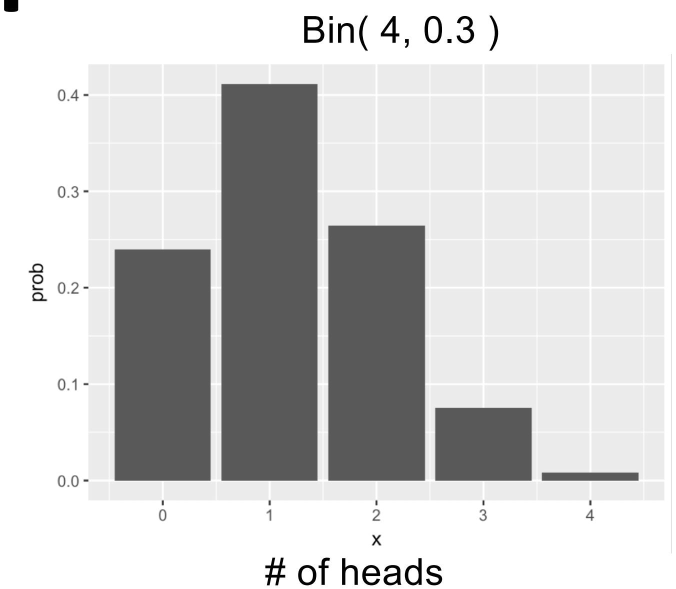


	Trial 1	Trial 2	Trial 3	…
Head	1	1	0	
Tail	0	0	1	

Binomial(n, π): n coin flips

- For n= 4, π=0.3



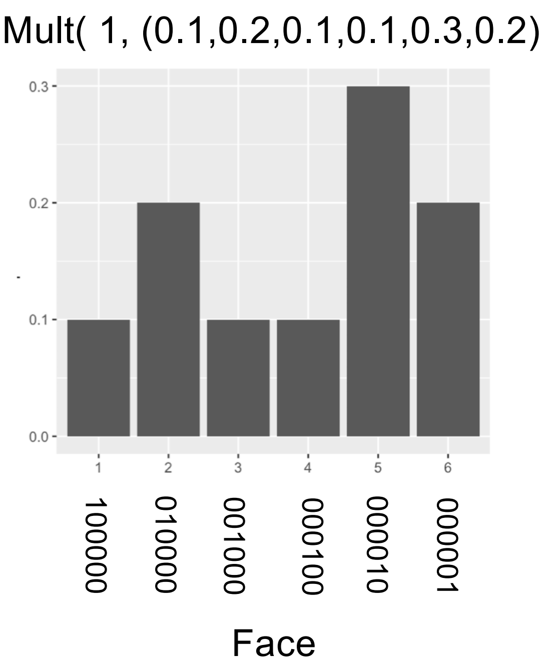


	Trial 1	Trial 2	Trial 3	…
# Head	3	2	0	
# Tail	1	2	4	

Multinomial(1, π): a dice roll

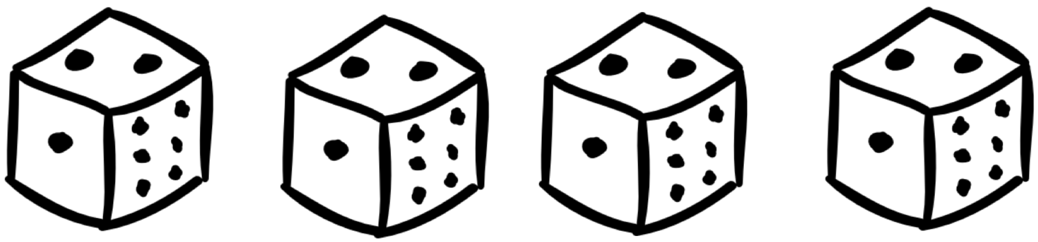
- For a dice roll

	Trial 1	Trial 2	Trial 3	…
1	0	0	0	
2	0	0	1	
3	0	0	0	
4	0	0	0	
5	0	1	0	
6	1	0	0	



Multinomial(n, π): multiple dice rolls

- For n=4

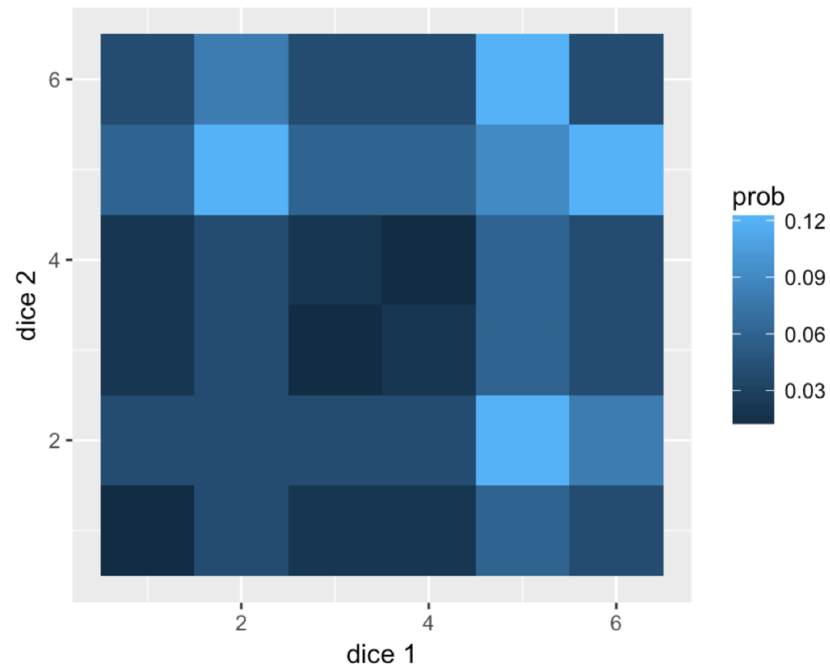


?

	Trial 1	Trial 2	Trial 3	…
1	0	0	1	
2	2	3	1	
3	0	0	0	
4	1	1	1	
5	0	0	1	
6	1	0	0	

Multinomial(n, π): multiple dice rolls

- The difficulty is due to multinomial being a multivariate distribution
- We can barely visualize n=2
- For **π**=(0.1, 0.2, 0.1, 0.1, 0.3, 0.2)

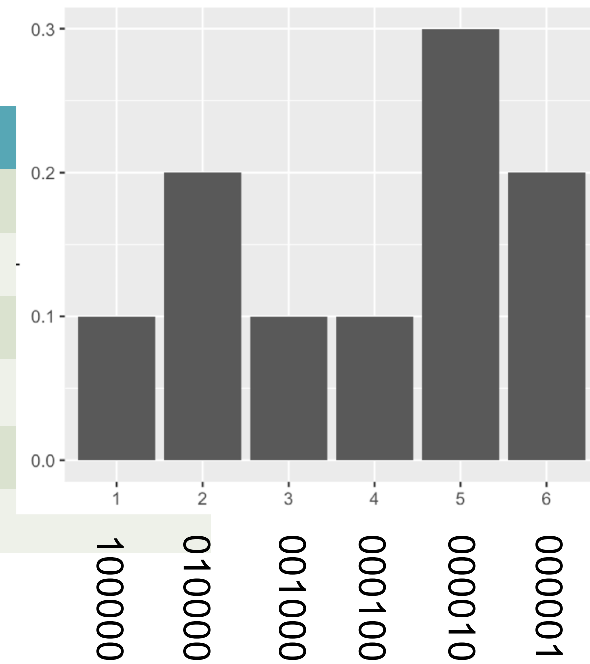


Categorical Distribution(π) = M(1, π)

- For a face of a dice rolled rather than a binary vector

	Trial 1	Trial 2	Trial 3	…
1	0	0	0	
2	0	0	1	
3	0	0	0	
4	0	0	0	
5	0	1	0	
6	1	0	0	

	6	5	1



Multinomial Distribution

- For c category an outcome **y**=(y_1, y_2, … , y_c)

$$f(\mathbf{y}|n, \boldsymbol{\pi}) = \frac{n!}{y_1!y_2!\cdots y_c!}\pi_1^{y_1}\pi_2^{y_2}\cdots\pi_c^{y_c}$$

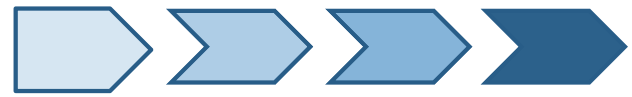
- In R **dmultinom**
- To generate from multinomial **rmultinom**
- Special case when c = 2

$$\begin{aligned} f(\mathbf{y}|n) &= \frac{n!}{y_1!(n-y_1)!}\pi_1^{y_1}(1-\pi_1)^{n-y_1} \\ &= \binom{n}{y_1}\pi_1^{y_1}(1-\pi_1)^{n-y_1} \end{aligned}$$

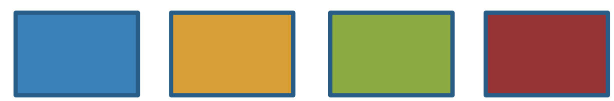
Categorical Regression

Two types of categorical outcome variables

- **Ordinal**: natural *order* among the response categories
  - Number of stars on a restaurant review site
  - Opinion: Yes/Maybe/No
  - Likert Scale: Strongly Agree/Agree/Neutral/Disagree/Strongly Disagree
  - ➢ <u>(Cumulative (proportional odds) / Adjacent Categories/ Continuation Ratio) Logit Model</u>



- **Nominal**: no natural order among the response categories
  - Butterfly types: A, B, C, D
  - Different words: the, and, here, up,..
  - Race/Ethnicity: White, Black, Asian, …
  - ➢ <u>Multinomial / Baseline-category Logit Model</u>



High level understanding

- Ordinal categorical regression
  - Same as logistic regression with different cutoffs.

- Nominal categorical regression
  - Many logistic regression comparing the baseline to each of the categories.

How to fit them in R (discussion)

- Ordered Categorical outcome variable
  - `library(MASS);` **`polr`**`()`
  - `library(VGAM);` **`vglm`**`( formula, family=`**`cumulative,`**`…)`

- Nominal Categorical outcome variable
  - `library(VGAM);` **`vglm`**`( formula, family=`**`multinomial,…)`**
  - `library(nnet);` **`multinom`**`()`
  - `library(mlogit); mlogit()`
  - `Library(mnlogit); mnlogit()`

Ordered Categorical Regression

Example: C-section and infection

Smoking and breathing

- Couple of examples from FKLM

Table 6.1 Data on infections for 251 C-sections

	C-section					
	Planned			Unplanned		
	Infection			Infection		
	I	II	no	I	II	no
Antibiotics						
Risk factor	0	1	17	4	7	87
No risk factor	0	0	2	0	0	0
No antibiotics						
Risk factor	11	17	30	10	13	3
No risk factor	4	4	32	0	0	9
	Head		Tail	Head		Tail

Table 6.2 Pulmonary function test

		Breathing test		
Age	Smoking status	Normal	Borderline	Abnormal
<40	Nonsmoker	577	27	7
	Former smoker	192	20	3
	Current smoker	682	46	11
40–59	Nonsmoker	164	4	0
	Former smoker	145	15	7
	Current smoker	245	47	27
		Tail	Head	

- If we cared only about comparing no vs infected,
- binomial regression seems appropriate.

Logistic regression

- We wanted to know the conditional probability of a head given x

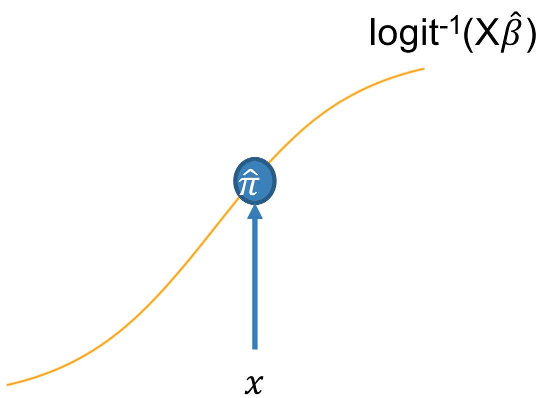


Table 6.2 Pulmonary function test

		Breathing test		
Age	Smoking status	Normal	Borderline	Abnormal
<40	Nonsmoker	577	27	7
	Former smoker	192	20	3
	Current smoker	682	46	11
40–59	Nonsmoker	164	4	0
	Former smoker	145	15	7
	Current smoker	245	47	27

Tail

Head

- Same for the multinomial outcome model, the only problem is that it has multiple probabilities.

Cumulative logit for ordered categorical

- The key is to understand what are 1s and 0s.

0	1
Normal	Borderline, Abnormal
Normal, Borderline	Abnormal

- Let y_i be response for subject i.
- $y_i = j$ means $y_{ij} = 1$ and $y_{ik} = 0$ for $k \neq j$ for K indicators.

Table 6.2 Pulmonary function test

		Breathing test		
Age	Smoking status	Normal	Borderline	Abnormal
<40	Nonsmoker	577	27	7
	Former smoker	192	20	3
	Current smoker	682	46	11
40–59	Nonsmoker	164	4	0
	Former smoker	145	15	7
	Current smoker	245	47	27
		π_1	π_2	π_3

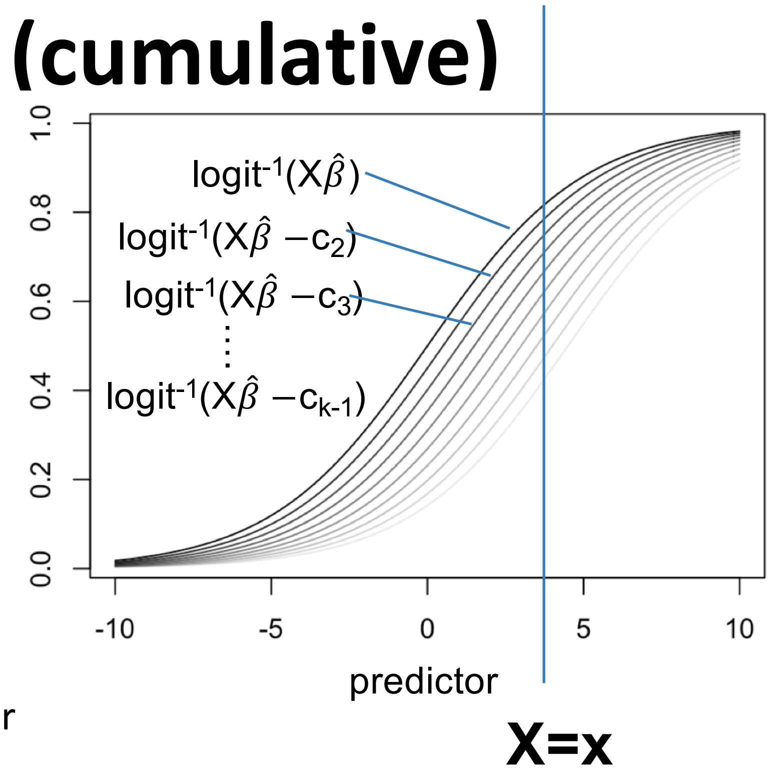
0	1
$P(y_i \le \text{Normal}) = \pi_1$	$P(y_i > \text{Normal}) = \pi_2 + \pi_3$
$P(y_i \le \text{Borderline}) = \pi_1 + \pi_2$	$P(y_i > \text{Borderline}) = \pi_3$

$\xleftrightarrow{\text{logit}}$ $\mathbf{x}_i\hat{\boldsymbol{\beta}}$

Multinomial Ordered Logit (cumulative)

- y takes values in 1, 2, …, K

$$
\begin{aligned}
\Pr(y > 1) &= \operatorname{logit}^{-1}(X\beta) \\
\Pr(y > 2) &= \operatorname{logit}^{-1}(X\beta - c_2) \\
\Pr(y > 3) &= \operatorname{logit}^{-1}(X\beta - c_3) \\
&\cdots \\
\Pr(y > K-1) &= \operatorname{logit}^{-1}(X\beta - c_{K-1}).
\end{aligned}
$$



- $0 = c_1 < c_2 < \cdots < c_{K-1}$, because the probabilities are strictly decreasing by multiplicative effect of $\exp(-c_j)$ for
- β is shared across different levels. This an assumption to simplify the model and is called proportional odds model.

Individual outcome probabilities

- The probabilities of individual outcomes:

$$\begin{aligned}\Pr(y=k) &= \Pr(y>k-1) - \Pr(y>k)\\ &= \operatorname{logit}^{-1}(X\beta - c_{k-1}) - \operatorname{logit}^{-1}(X\beta - c_k).\end{aligned}$$

$P(y_i > \text{Normal}) = \pi_2 + \pi_3$
$P(y_i > \text{Borderline}) = \pi_3$

$P(y_i = \text{Normal}) = 1 - P(y_i > \text{Normal}) = 1 - (\pi_2 + \pi_3) = \pi_1$

$P(y_i = \text{Borderline}) = P(y_i > \text{Normal}) - P(y_i > \text{Borderline})$
$= \pi_2 + \pi_3 - \pi_3 = \pi_2$

$P(y_i = \text{Abnormal}) = P(y_i > \text{Borderline}) = \pi_3$

Table 6.2 Pulmonary function test

		Breathing test		
Age	Smoking status	Normal	Borderline	Abnormal
<40	Nonsmoker	577	27	7
	Former smoker	192	20	3
	Current smoker	682	46	11
40–59	Nonsmoker	164	4	0
	Former smoker	145	15	7
	Current smoker	245	47	27
		π_1	π_2	π_3

Cumulative odds ratio

- Odds ratio of cumulative probabilities is called the cumulative odds ratio.
- For a general cumulative logit model

$$\mathrm{logit}\big(\mathrm{P}(y_i \le j)\big) = \alpha + \boldsymbol{x_i\beta}$$

- Since

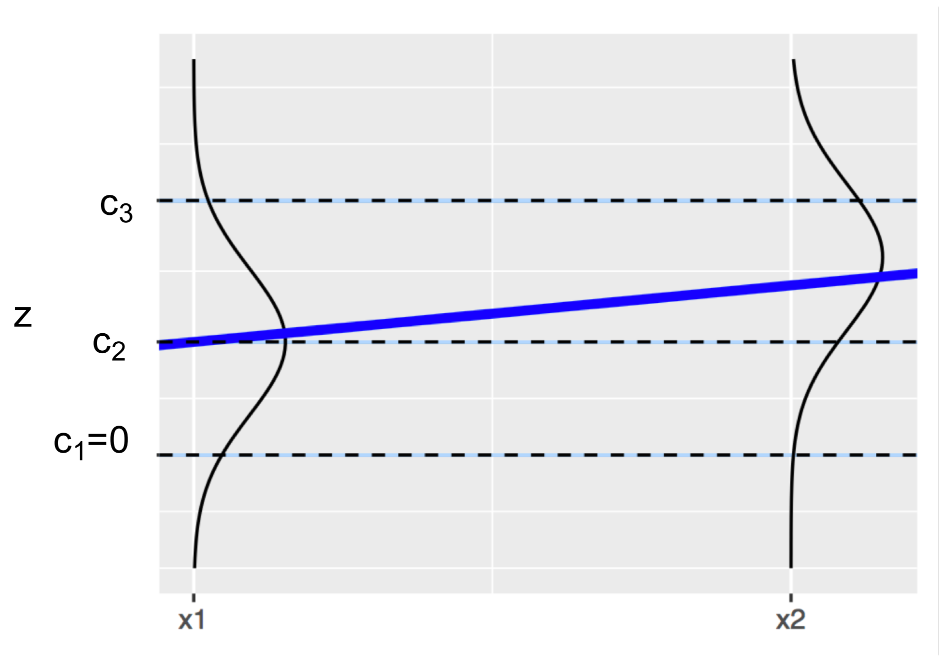
$$\mathrm{Logit}(\mathrm{P}(y_i \le j|\boldsymbol{x}_i = \boldsymbol{u})\,) - \mathrm{logit}\,\big(\mathrm{P}(y_i \le j|\boldsymbol{x}_i = \boldsymbol{v})\big)$$
$$= \log\left(\frac{\mathrm{P}(y_i \le j|\boldsymbol{x}_i = \boldsymbol{u})/\mathrm{P}(y_i > j|\boldsymbol{x}_i = \boldsymbol{u})}{\mathrm{P}(y_i \le j|\boldsymbol{x}_i = \boldsymbol{v})/\mathrm{P}(y_i > j|\boldsymbol{x}_i = \boldsymbol{v})}\right) = (\boldsymbol{u} - \boldsymbol{v})\boldsymbol{\beta}$$

- The odds that response is less than j for **x**=**u** is $\exp((\boldsymbol{u} - \boldsymbol{v})\boldsymbol{\beta})$ times the odds at **x**=**v**.
- When this proportionality is common across all the categories, the is a property called the proportional odds.

Latent variable formulation

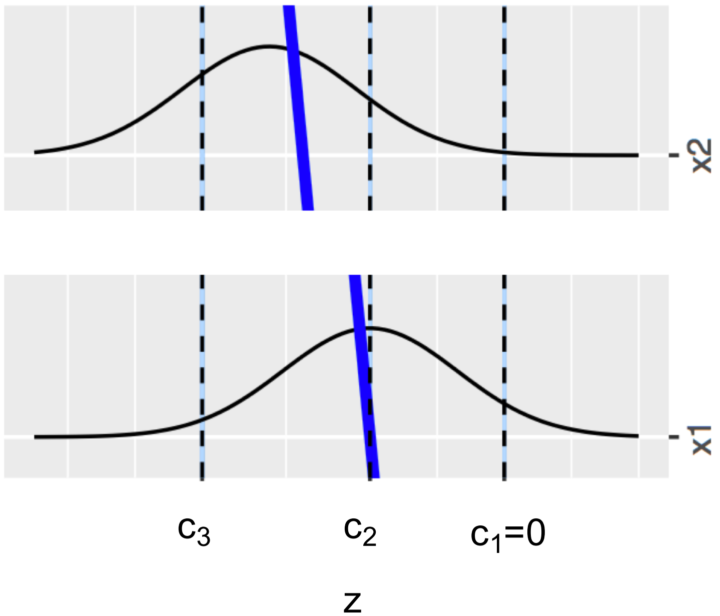
- We can use the latent variable formulation to understand the model as extension of the logistic regression model

$$
\begin{aligned}
y_i &= \begin{cases} 1 & \text{if } z_i < 0 \\ 2 & \text{if } z_i \in (0, c_2) \\ 3 & \text{if } z_i \in (c_2, c_3) \\ & \dots \\ K-1 & \text{if } z_i \in (c_{K-2}, c_{K-1}) \\ K & \text{if } z_i > c_{K-1} \end{cases} \\
z_i &= X_i\beta + \epsilon_i,
\end{aligned}
$$



Latent variable formulation

$$
\begin{aligned}
y_i &= \begin{cases} 1 & \text{if } z_i < 0 \\ 2 & \text{if } z_i \in (0, c_2) \\ 3 & \text{if } z_i \in (c_2, c_3) \\ & \dots \\ K-1 & \text{if } z_i \in (c_{K-2}, c_{K-1}) \\ K & \text{if } z_i > c_{K-1} \end{cases} \\
z_i &= X_i\beta + \epsilon_i,
\end{aligned}
$$



Latent variable formulation

- There are three ways to formulate the model

$$
\begin{aligned}
y_i &= \begin{cases} 1 & \text{if } z_i < c_{1.5} \\ 2 & \text{if } z_i \in (c_{1.5}, c_{2.5}) \\ 3 & \text{if } z_i > c_{2.5} \end{cases} \\
z_i &\sim \text{logistic}(x_i, \sigma^2).
\end{aligned}
$$

$$
\begin{aligned}
y_i &= \begin{cases} 1 & \text{if } z_i < 0 \\ 2 & \text{if } z_i \in (0, c_2) \\ 3 & \text{if } z_i > c_2 \end{cases} \\
z_i &= \alpha + \beta x + \epsilon_i,
\end{aligned}
$$

$$
\begin{aligned}
y_i &= \begin{cases} 1 & \text{if } z_i < c_{1|2} \\ 2 & \text{if } z_i \in (0, c_{2|3}) \\ 3 & \text{if } z_i > c_{2|3} \end{cases} \\
z_i &= \beta x + \epsilon_i,
\end{aligned}
$$

$\varepsilon_i \sim \text{logistic}(0,1)$

Model (6.11)	Model (6.12)	Model (6.13)
$c_{1.5}$	$-\alpha/\beta$	$-c_{1\|2}/\beta$
$c_{2.5}$	$(c_2-\alpha)/\beta$	$-c_{2\|3}/\beta$
σ	$1/\beta$	$1/\beta$

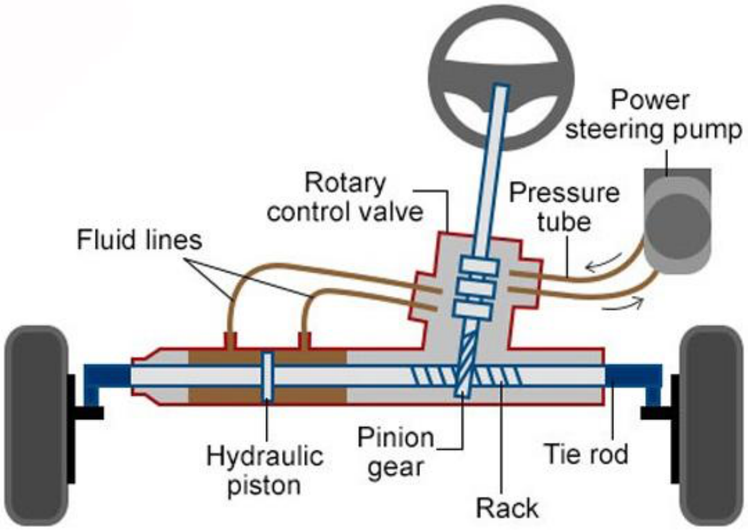
Car preferences (air conditioning and power steering)

- 50 subjects in each of the 6 categories (2 gender and 3 age categories)
- How important is air conditioning and power steering?

Air Conditioning

http://midasofmadison.com/what-to-do-when-your-cars-ac-air-doesnt-smell-right/

Power Steering



https://ekonomi.kompas.com/read/2017/01/31/181626730/bedanya.nyetir.kemudi.elektrik.dengan.hidrauli

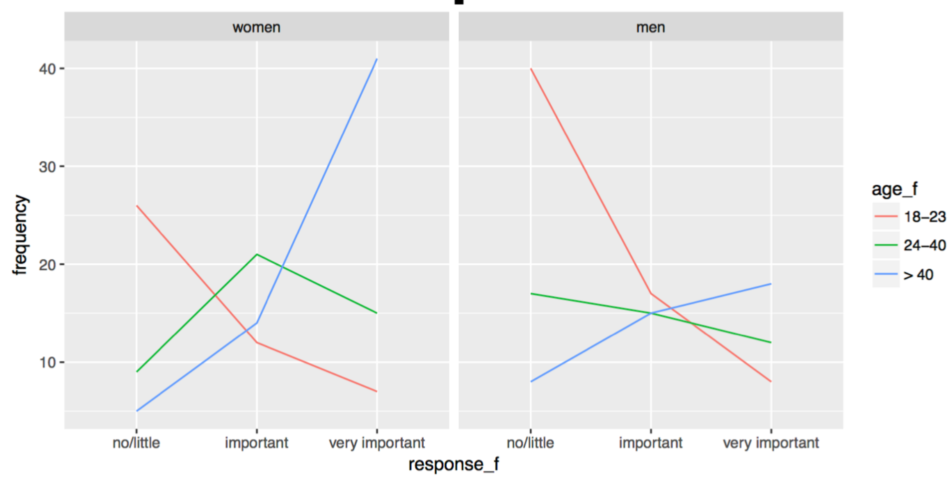
Car preferences (air conditioning and power steering)

- 50 subjects in each of the 6 categories (2 gender and 3 age categories)
- How important is air conditioning and power steering?

Sample data:

sex	age	response	frequency
women	18-23	no/little	26
women	18-23	important	12
women	18-23	very important	7
women	24-40	no/little	9
women	24-40	important	21
women	24-40	very important	15

...



Car preferences (air conditioning and power steering)

```
##       sex   age      response frequency     response_f sex_m age_f
## 1: women 18-23     no/little        26      no/little women 18-23
## 2: women 18-23     important        12      important women 18-23
## 3: women 18-23 very important         7 very important women 18-23
## 4: women 24-40     no/little         9      no/little women 24-40
## 5: women 24-40     important        21      important women 24-40
## 6: women 24-40 very important        15 very important women 24-40
```

Car preferences (air conditioning and power steering)

```
polr(response_f~sex_m+age_f,
                    weights=frequency,
                             data=Car_preferences)
```

Logit P(y > little) $= \log((\pi_2+\pi_3)/\pi_1) = (\beta_1\ \text{men}+\beta_2\ \text{age}_{24-40} + \beta_3\ \text{age}_{>40}\ \boxed{-c_{12}})$

Logit P(y > important) $= \log(\pi_3/\pi_1+\pi_2) = (\beta_1\ \text{men}+\beta_2\ \text{age}_{24-40} + \beta_3\ \text{age}_{>40}\ \boxed{-c_{23}})$

		Value	Std. Error	t value
shared	sex_mmen	-0.58	0.23	-2.55
	age_f24-40	1.15	0.28	4.13
	age_f> 40	2.23	0.29	7.66
	no/little|important	0.04	0.23	0.19
	important|very important	1.65	0.26	6.47

Interpreting the result manually

- Fitted model

$\text{Logit } P(\hat{y} > \text{little}) = \log((\hat{\pi}_2+\hat{\pi}_3)/\hat{\pi}_1) = (-0.58\ \text{men}+1.15\ \text{age}_{24-40} + 2.23\ \text{age}_{>40} - 0.04)$

$\text{Logit } P(\hat{y} > \text{important}) = \log(\hat{\pi}_3/\hat{\pi}_1+\hat{\pi}_2) = (-0.58\ \text{men}+1.15\ \text{age}_{24-40} + 2.23\ \text{age}_{>40} - 1.65)$

- What are the probabilities for female 18-23 ?

	Value	Std. Error	t value
sex_mmen	-0.58	0.23	-2.55
age_f24-40	1.15	0.28	4.13
age_f> 40	2.23	0.29	7.66
no/little|important	0.04	0.23	0.19
important|very important	1.65	0.26	6.47

Interpreting the result manually

$$\log((\hat{\pi}_2+\hat{\pi}_3)/\hat{\pi}_1) = -0.58\ \text{men} + 1.15\ \text{age}_{24-40} + 2.23\ \text{age}_{>40} - 0.04$$

$$\log(\hat{\pi}_3/(\hat{\pi}_1+\hat{\pi}_2)) = -0.58\ \text{men} + 1.15\ \text{age}_{24-40} + 2.23\ \text{age}_{>40} - 1.65$$

- Female: men=0
- Age 18-23: $\text{age}_{24-40} = 0$ $\text{age}_{>40} = 0$

$$(\hat{\pi}_2+\hat{\pi}_3)/\hat{\pi}_1 = \exp(-0.04)$$

$$\hat{\pi}_3/(\hat{\pi}_1+\hat{\pi}_2) = \exp(-1.65)$$

$\Rightarrow$ $\hat{\pi}_1 = 0.51$, $\hat{\pi}_2 = 0.33$ and $\hat{\pi}_3 = 0.16$

- Male: men=1
- Age 18-23: $\text{age}_{24-40} = 0$ $\text{age}_{>40} = 0$

$$(\hat{\pi}_2+\hat{\pi}_3)/\hat{\pi}_1 = \exp(-0.62)$$

$$\hat{\pi}_3/(\hat{\pi}_1+\hat{\pi}_2) = \exp(-2.23)$$

$\Rightarrow$ $\hat{\pi}_1 = 0.65$, $\hat{\pi}_2 = 0.25$ and $\hat{\pi}_3 = 0.10$

Solving using numerical calculator

- Solve

https://www.wolframalpha.com

Input interpretation:

solve $\log\left(\frac{b+c}{a}\right) = -0.04$, $\log\left(\frac{c}{a+b}\right) = -1.65$, $a+b+c=1$ for a, b, c

Open code

$\log(x)$ is the natural logarithm

Enlarge | Data | Customize | Plaintext | Interactive

Result:

More digits

$$a = \frac{\sqrt[25]{e}}{1+\sqrt[25]{e}} \approx 0.509999 \text{ and}$$

$$b = \frac{e^{33/20} - \sqrt[25]{e}}{1+\sqrt[25]{e} + e^{33/20} + e^{169/100}} \approx 0.328892 \text{ and } c = \frac{1}{1+e^{33/20}} \approx 0.161109$$

Download page

POWERED BY THE WOLFRAM LANGUAGE

Assuming "log" is the natural logarithm | Use the base 10 logarithm instead

Alternative: Prediction

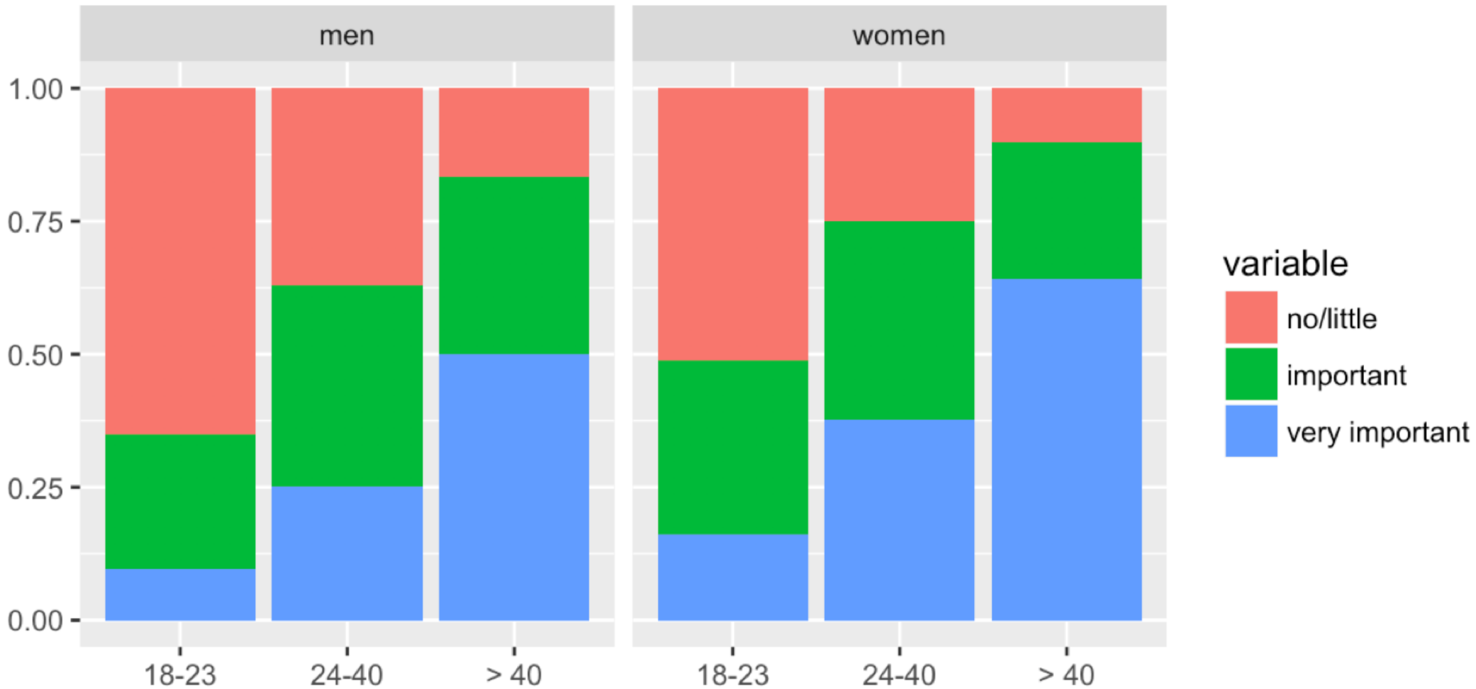
- predict(fit, newdata=X, type="p")
- The key is to create X to show your point.

```
expand.grid(sex_m=c("men","women"),
            age_f=c("18-23","24-40","> 40"))
predict(car_polr,newdata=predx,type = "p")
```

	sex_m	age_f	no/little	important	very important
1	men	18-23	0.6501647	0.2528518	0.09698344
2	women	18-23	0.5108826	0.3286797	0.16043777
3	men	24-40	0.3711383	0.3761309	0.25273073
4	women	24-40	0.2490730	0.3752352	0.37569186
5	men	> 40	0.1662146	0.3334706	0.50031477
6	women	> 40	0.1007497	0.2587618	0.64048855

Visualization

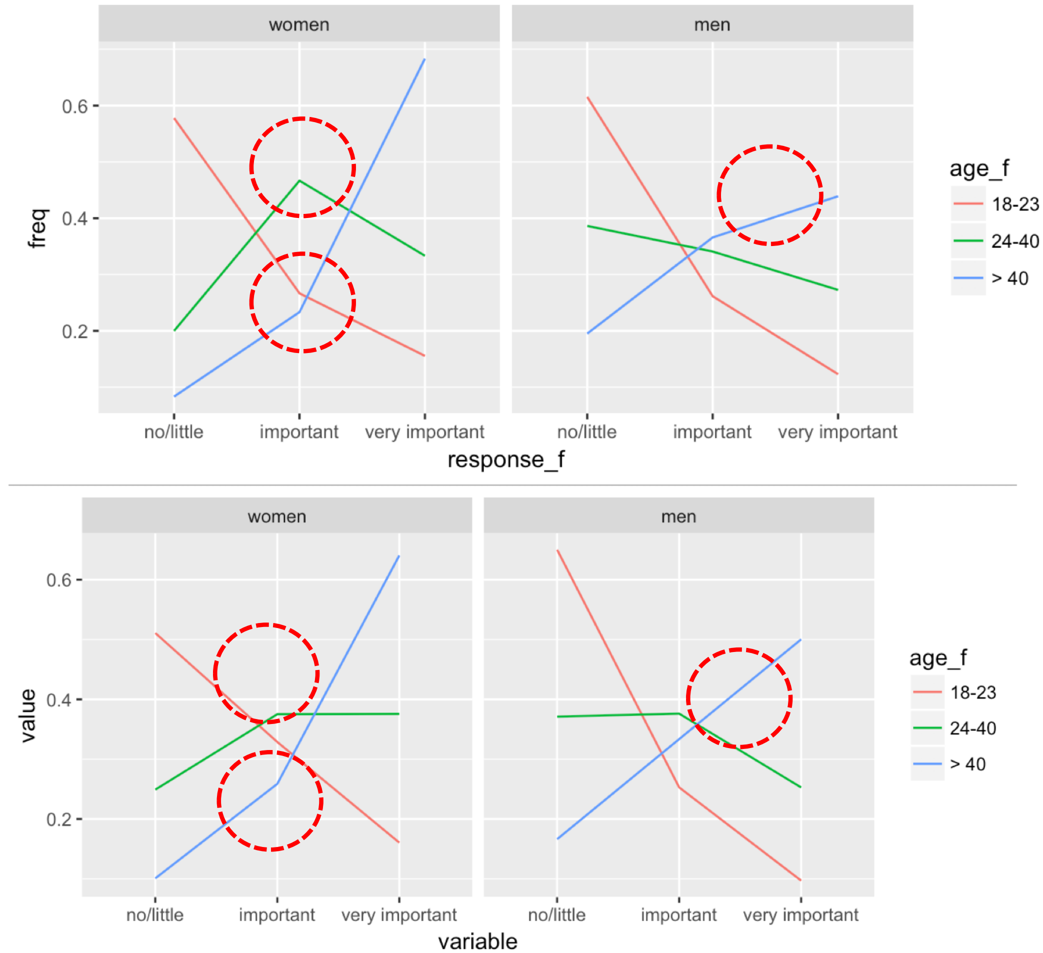
- predict(fit, newdata=, type="p")



Check

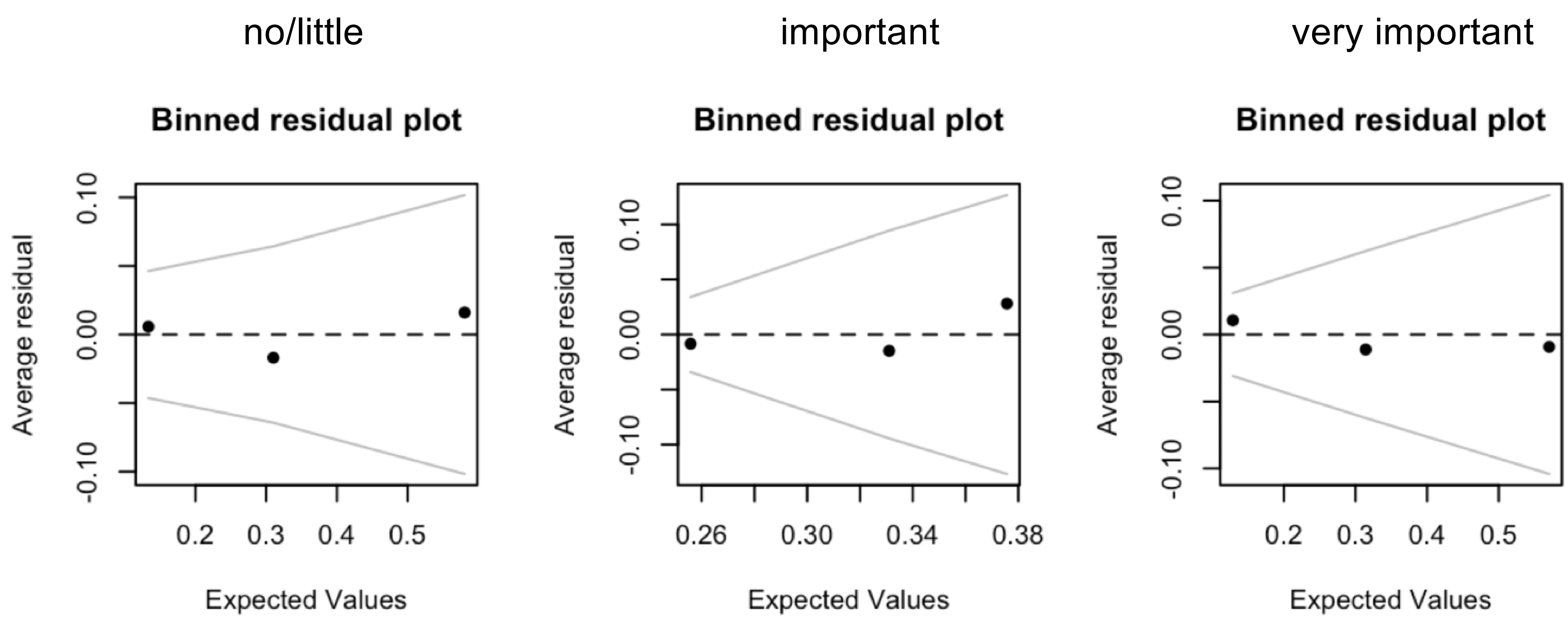
observed

predicted



Residual check

- We have 3 observed proportions and $\hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3$



Confidence Interval

- Profile Likelihood Interval: `confint`

	2.5 %	97.5 %
sex_mmen	-1.02	-0.13
age_f24-40	0.61	1.70
age_f> 40	1.67	2.81

- Wald interval: `confint.default`

	2.5 %	97.5 %
sex_mmen	-1.02	-0.13
age_f24-40	0.60	1.69
age_f> 40	1.66	2.80

Statistics for model comparison

- Deviance (`1-pchisq(deviance(polrfit),df.residual(polrfit))`)

$$D = 2\sum_{i=1}^{N}\sum_{j=1}^{c} n_i y_{ij} \log \frac{n_i y_{ij}}{n_i \hat{\pi}_{ij}} = 2\sum \text{observed} \log \frac{\text{observed}}{\text{fitted}}$$

- Pearson

$$X^2 = 2\sum_{i=1}^{N}\sum_{j=1}^{c} \frac{(n_i y_{ij} - n_i \hat{\pi}_{ij})^2}{n_i \hat{\pi}_{ij}} = \sum \frac{(\text{observed} - \text{fitted})^2}{\text{fitted}}$$

- *With cell frequency > 5* $\chi^2_{N(c-1)-p}$
- AIC

Model comparison

- When you are concerened about adding interaction

```
pchisq(deviance(fit)-deviance(fit2),
        df=df.residual(fit)-df.residual(fit2), lower.tail=FALSE)
```

```
> fit <-polr(response_f~sex_m+age_f,weights=frequency,data=Car_preferences)
> fit2<-polr(response_f~sex_m*age_f,weights=frequency,data=Car_preferences)
> pchisq(deviance(fit)-deviance(fit2), df=df.residual(fit)-df.residual(fit2),
lower.tail=FALSE)
[1] 0.3374261
> AIC(fit);AIC(fit2)
[1] 591.2956
[1] 593.1227
```

vglm uses slightly different parameterization

- You can fit the same model using vglm instead of polr.

```
polr(response_f~sex_m+age_f,
       weights=frequency,
       data=Car_preferences)
```

	Value	Std. Error
sex_mmen	-0.58	0.23
age_f24-40	1.15	0.28
age_f> 40	2.23	0.29
no/little|important	0.04	0.23
important|very important	1.65	0.26

```
vglm(ordered(response_f)~sex_m+age_f,
       family=cumulative(parallel=TRUE)
       weights=frequency,
       data=Car_preferences)
```

	Estimate	Std. Error
(Intercept):1	0.04	0.23
(Intercept):2	1.65	0.25
sex_mmen	0.58	0.23
age_f24-40	-1.15	0.28
age_f> 40	-2.23	0.29

$P(y_i > \text{no/little}) = \pi_2 + \pi_3 = \text{logit}^{-1}(\mathbf{X\beta} - c_1)$

$P(y_i > \text{important}) = \pi_3 = \text{logit}^{-1}(\mathbf{X\beta} - c_2)$

$P(y_i \le \text{no/little}) = \pi_1 = \text{logit}^{-1}(\alpha_1 - \mathbf{X\beta})$

$P(y_i \le \text{important}) = \pi_1 + \pi_2 = \text{logit}^{-1}(\alpha_2 - \mathbf{X\beta})$

Other types of logit

- What you call success and failure defines different models
- *Cumulative Logit:*

  (π_1) vs $(\pi_2$ & π_3 & $\pi_4)$, $(\pi_1$ & $\pi_2)$ vs $(\pi_3$ & $\pi_4)$, $(\pi_1$ & π_2 & $\pi_3)$ vs (π_4),
- Adjacent Categories Logit:
  - π_1 vs π_2, π_2 vs π_3, π_3 vs π_4
- Continuation Ratio Logit:
  - π_1 vs π_2, $(\pi_1$ & $\pi_2)$ vs π_3, $(\pi_1$ & π_2 & $\pi_3)$ vs π_4,...

- All of the above can be fit using vglm function in VGAM package

Cumulative Logit for 10 categories

	1	2	3	4	5	6	7	8	9	10
1	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
2	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
3	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
4	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
5	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
6	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
7	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
8	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
9	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}

1

0

Adjacent Category Logit for 10 categories

	1	2	3	4	5	6	7	8	9	10
1	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
2	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
3	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
4	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
5	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
6	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
7	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
8	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
9	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}

1

0

Continuation Ratio Logit for 10 categories

	1	2	3	4	5	6	7	8	9	10
1	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
2	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
3	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
4	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
5	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
6	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
7	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
8	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
9	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}

0

1

vglm can fit them all

family =

- Partial proportional odds: `cumulative(parallel=FALSE~men)`

  Logit P(y ≤ little) = $\log(\pi_1/(\pi_2+\pi_3))$ = $(a1-\beta_{1(12)}$ men$-\beta_2$ age$_{24-40}$ $-$ β_3 age$_{>40})$

  Logit P(y ≤ important)= $\log((\pi_1+\pi_2)/\pi_3)$= $(a2-\beta_{1(23)}$ men$-\beta_2$ age$_{24-40}$ $-$ β_3 age$_{>40})$

- Adjacent Categories: `acat(reverse=TRUE, parallel=TRUE)`

$$\frac{\pi_1}{\pi_2}, \frac{\pi_2}{\pi_3}, \ldots, \frac{\pi_{J-1}}{\pi_J}. \qquad \log\left(\frac{\pi_j}{\pi_{j+1}}\right) = \mathbf{x}_j^T \boldsymbol{\beta}_j.$$

- Continuation Ratio : `cratio(reverse=FALSE, parallel=TRUE)`

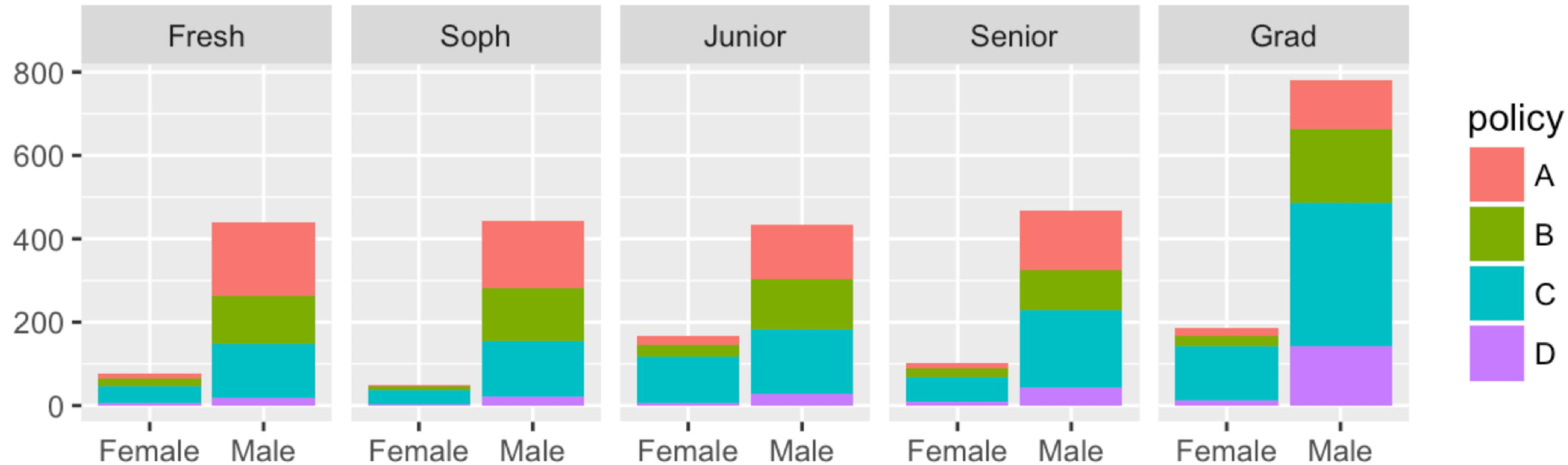
$$\frac{\pi_1}{\pi_2}, \frac{\pi_1+\pi_2}{\pi_3}, \ldots, \frac{\pi_1+\ldots+\pi_{J-1}}{\pi_J} \qquad \log\left(\frac{\pi_j}{\pi_{j+1}+\ldots+\pi_J}\right) = \mathbf{x}_j^T \boldsymbol{\beta}_j$$

newspaper survey on Vietnam War

`data(uncviet)`

- A student newspaper conducted a survey of student opinions about the Vietnam War in May 1967.
- Responses were classified by sex, year in the program and one of four opinions.
  - A (defeat power of North Vietnam by widespread bombing and land invasion)
  - B (follow the present policy)
  - C (withdraw troops to strong points and open negotiations on elections involving the Viet Cong)
  - D (immediate withdrawal of all U.S. troops)
- The survey was voluntary.
- Treat the opinion as the response and the sex and year as predictors.
- Build a proportional odds model, giving an interpretation to the estimates.
- Make sure to check the fit of your model.

EDA

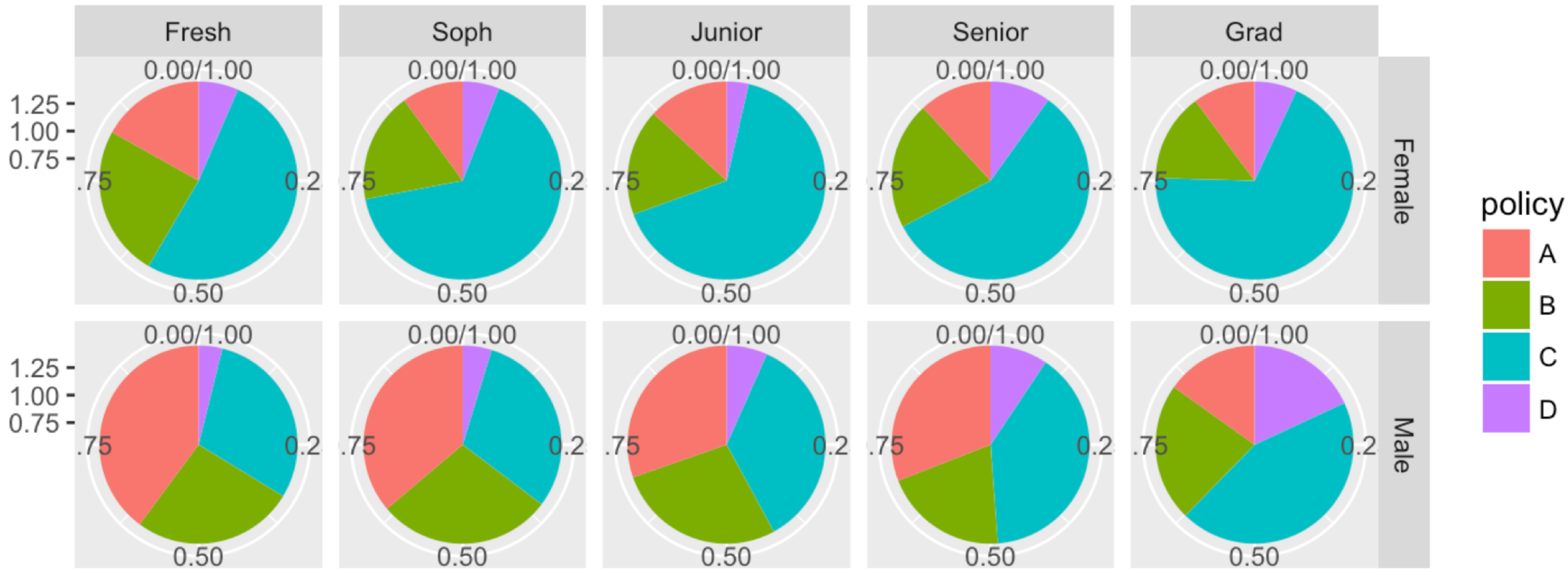


Fresh
Soph
Junior
Senior
Grad
800
600
400
200
0
Female
Male
policy
A
B
C
D

EDA

Female
Male
1.00
0.75
0.50
0.25
0.00
Fresh
Soph
Junior
Senior
Grad
Fresh
Soph
Junior
Senior
Grad
policy
A
B
C
D

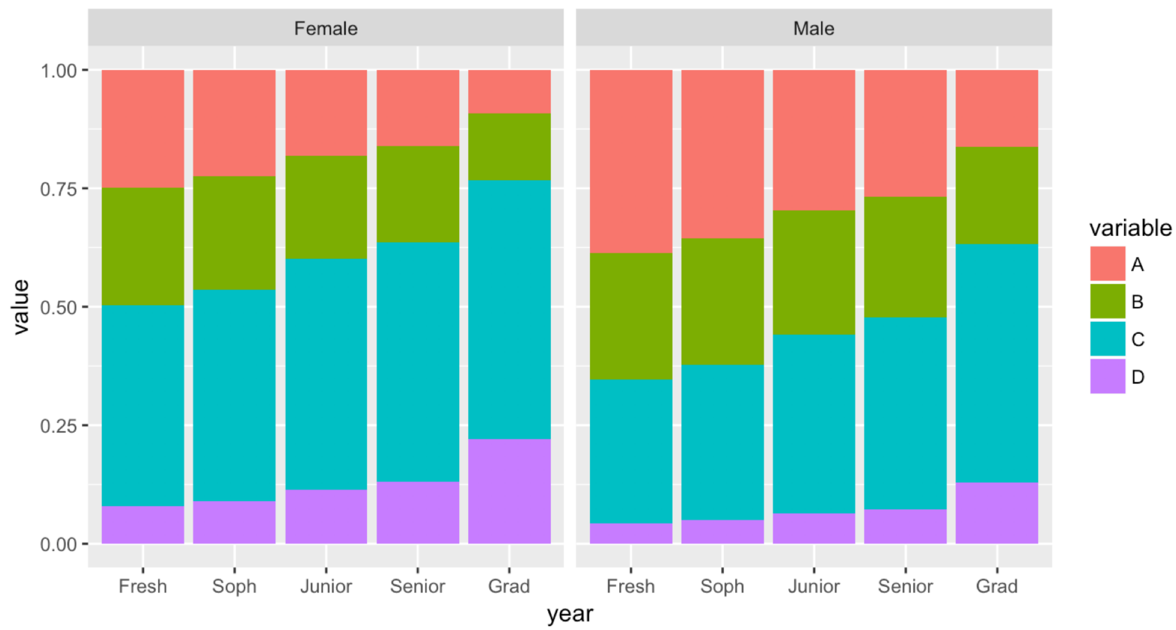
EDA



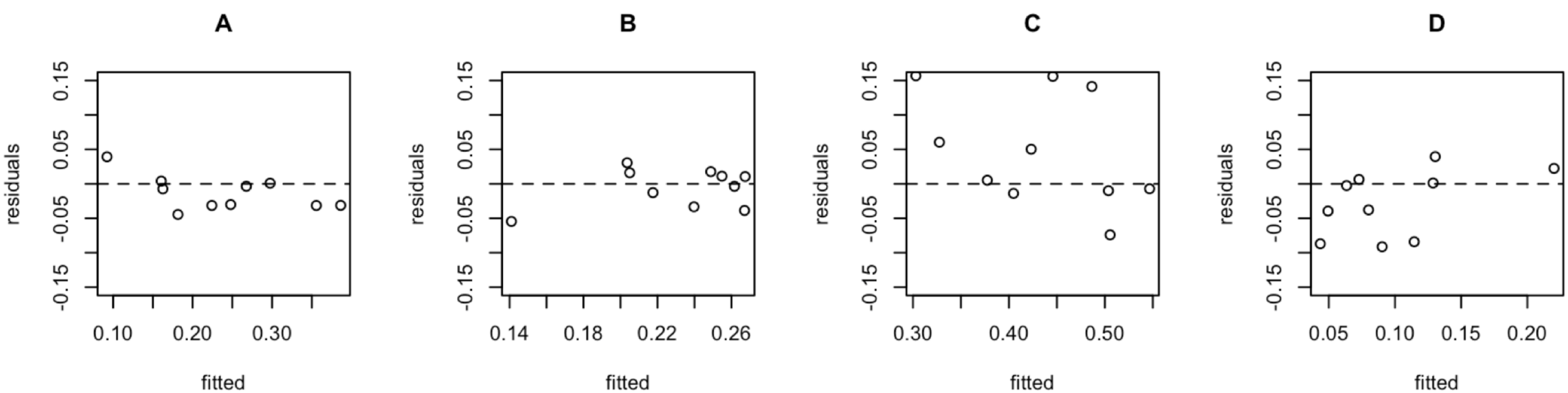
Fresh
Soph
Junior
Senior
Grad
Female
Male
0.00/1.00
0.50
1.25
1.00
0.75
policy
A
B
C
D

Fitted result

```
fit<-vglm(cbind(A,B,C,D)~sex+year, family=cumulative(parallel=TRUE),data=ddat)
```



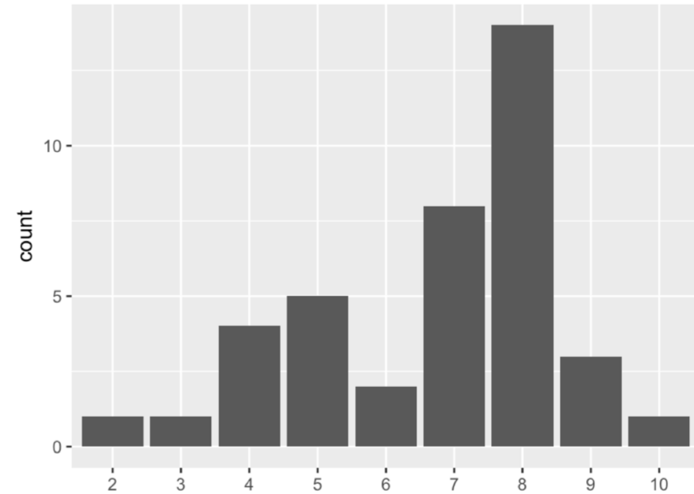
Residuals

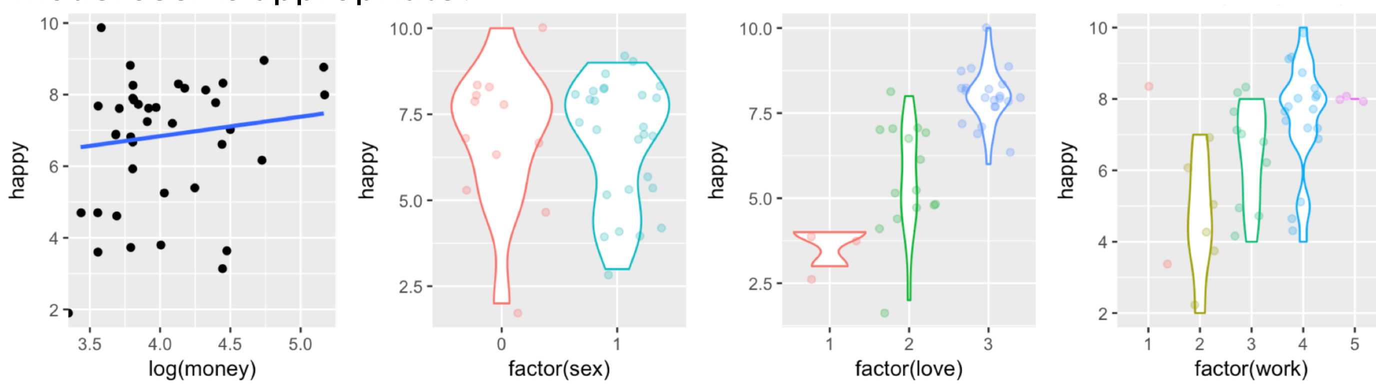

A
B
C
D
residuals
fitted
0.15
0.05
-0.05
-0.15
0.10
0.20
0.30
0.14
0.18
0.22
0.26
0.40
0.50
0.05
0.15
0.20

(HW) Happiness data

library(faraway); data(happy)

- 39 students in a University of Chicago MBA class
  - 4 predictor variables
    - Love
    - Work
    - Sex
    - Money
- What model seems appropriate?





Unordered categorical variable

Unordered Categorical Outcome

- C-section example

Table 6.1 Data on infections for 251 C-sections

	C-section					
	Planned			Unplanned		
	Infection			Infection		
	I	II	no	I	II	no
Antibiotics						
Risk factor	0	1	17	4	7	87
No risk factor	0	0	2	0	0	0
No antibiotics						
Risk factor	11	17	30	10	13	3
No risk factor	4	4	32	0	0	9

- Three categories of outcome
  - type I infection,
  - type II infection, and
  - no infection
- Type I and type II are just different **types** of infection.

(Type I ?? Type II) > worse > No infection

Multinomial Logit Model

- You can NOT assume ordering of π_j for c different categories.
- The probabilities $(\pi_1, \pi_2, \ldots, \pi_c)$ must add up to 1.
  - $\rightarrow$ π_j are not independent.
  - So we can not model π_j as they are.

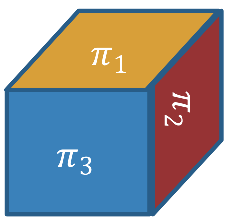


Table 6.1 Data on infections for 251 C-sections

	C-section					
	Planned			Unplanned		
	Infection			Infection		
	I	II	no	I	II	no
Antibiotics						
Risk factor	0	1	17	4	7	87
No risk factor	0	0	2	0	0	0
No antibiotics						
Risk factor	11	17	30	10	13	3
No risk factor	4	4	32	0	0	9

- However, we can speak of odds wrt some baseline category.
- That is we can model π_j / π_c as a function of linear predictors.

$$g\left(\frac{\pi_{ij}}{\pi_{ic}}\right) = \mathrm{f}(x_i)$$

baseline category logit

- baseline category logit transformation of multinomial probabilities

- Multinomial logit model (Baseline-category logit):

$\pi_{1,}\ \pi_{2,}\ \ldots,\ \pi_{j,}\ldots,\ \pi_c \Rightarrow$ Log($\pi_1/\ \pi_c$), log($\pi_2/\ \pi_{c,}$) …, log($\pi_j/\ \pi_{c,}$) ,,…., log($\pi_{c\,-\,1}/\ \pi_{c,}$)

- These (c − 1) equations determine equations for logits with other pairs of response categories

$$\log\frac{\pi_{ij}}{\pi_{ic}} = \mathbf{X}_i\boldsymbol{\beta}_j = \sum_{k=1}^{p}\beta_{jk}x_{ik}, j = 1, \ldots, (c-1)$$

$$\log\frac{\pi_{ia}}{\pi_{ib}} = \log\frac{\pi_{ia}}{\pi_{ic}} - \log\frac{\pi_{ib}}{\pi_{ic}} = \mathbf{X}_i(\boldsymbol{\beta}_a - \boldsymbol{\beta}_b)$$

Multinomial Logit

▪

- We can express response probability as

$$\pi_{ij} = \pi_{ic} exp(\mathbf{X}_i \boldsymbol{\beta}_j)$$

- Since $\sum_{j=1}^{c} \pi_{ij} = 1$

$$\pi_{ic} = \frac{1}{1 + \sum_{h=1}^{c-1} exp(\mathbf{X}_i \boldsymbol{\beta}_h)}$$

- Therefore

$$\hat{\pi}_{ic} = \frac{exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{1 + \sum_{h=1}^{c-1} exp(\mathbf{X}_i \boldsymbol{\beta}_h)}$$

with $\beta_c = 0$

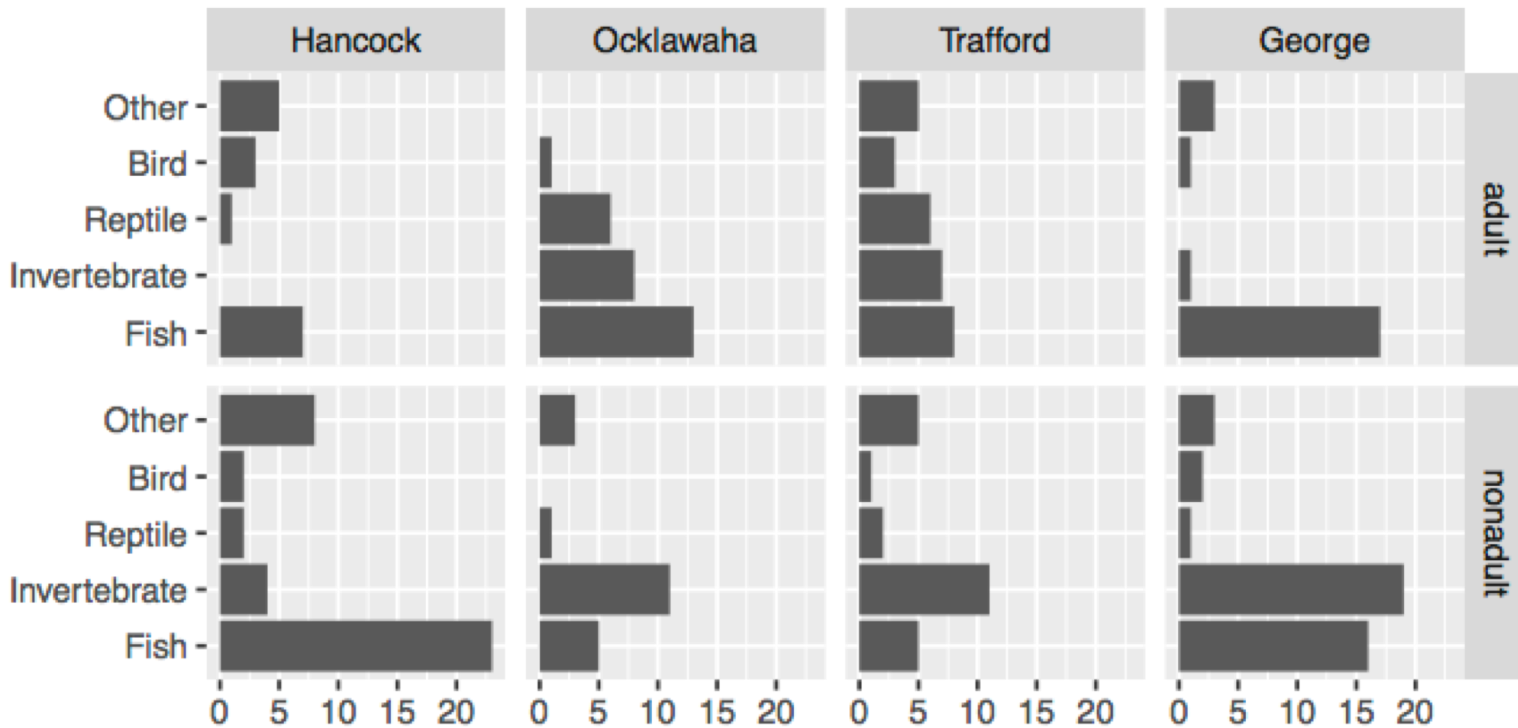
Interpretation is not simple

- Interpretation of effects overall is NOT simple
- If we take derivative wrt x_{ik}

$$\frac{\partial \pi_{ij}}{\partial x_{ik}} = \pi_{ij}(\beta_{jk} - \sum_{j'} \pi_{ij'} \beta_{j'k})$$

- Unlike logistic regression this rate of change does not always have the same sign as β_{jk} .

What does Alligators like to eat?



Fitting the model using vglm

- The model can be fit using vglm

```
gator_fit<-vglm(cbind(Invertebrate, Reptile, Bird, Other,Fish) ~ size+lake,
family=multinomial,data=gator)
```

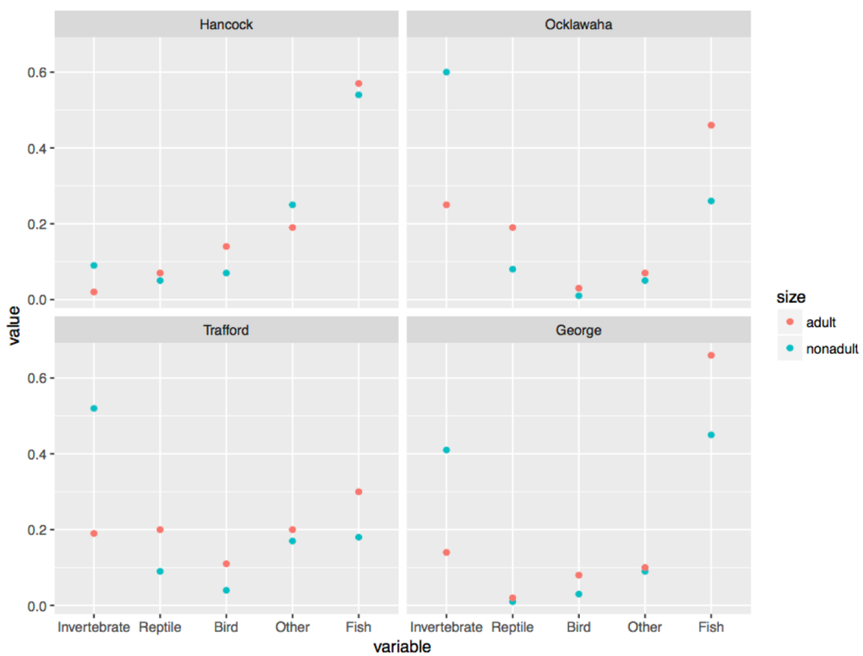
- However interpreting the coefficients is not trivial…

- Model selection:

Model	Deviance	Df	AIC
size+lake	17.08	12	135.03
lake	38.17	16	148.12
size	66.21	24	160.16

```
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept):1         -3.21       0.64   -5.02     0.00
## (Intercept):2         -2.07       0.71   -2.93     0.00
## (Intercept):3         -1.40       0.61   -2.30     0.02
## (Intercept):4         -1.08       0.47   -2.29     0.02
## sizenonadult:1         1.46       0.40    3.68     0.00
## sizenonadult:2        -0.35       0.58   -0.61     0.54
## sizenonadult:3        -0.63       0.64   -0.98     0.33
## sizenonadult:4         0.33       0.45    0.74     0.46
## lakeOcklawaha:1        2.60       0.66    3.93     0.00
## lakeOcklawaha:2        1.22       0.79    1.55     0.12
## lakeOcklawaha:3       -1.35       1.16   -1.16     0.25
## lakeOcklawaha:4       -0.82       0.73   -1.12     0.26
## lakeTrafford:1         2.78       0.67    4.14     0.00
## lakeTrafford:2         1.69       0.78    2.17     0.03
## lakeTrafford:3         0.39       0.78    0.50     0.62
## lakeTrafford:4         0.69       0.56    1.23     0.22
## lakeGeorge:1           1.66       0.61    2.71     0.01
## lakeGeorge:2          -1.24       1.19   -1.05     0.29
## lakeGeorge:3          -0.70       0.78   -0.89     0.37
## lakeGeorge:4          -0.83       0.56   -1.48     0.14
```

Fitted probabilities

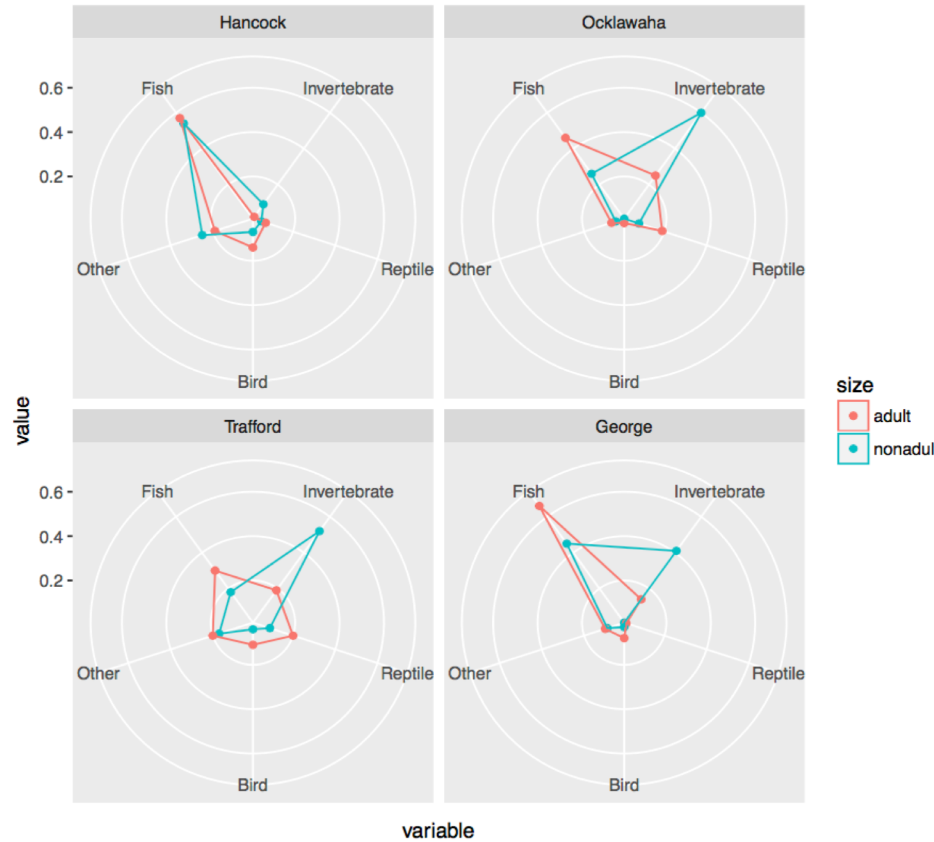


- Looking at the fitted probabilities

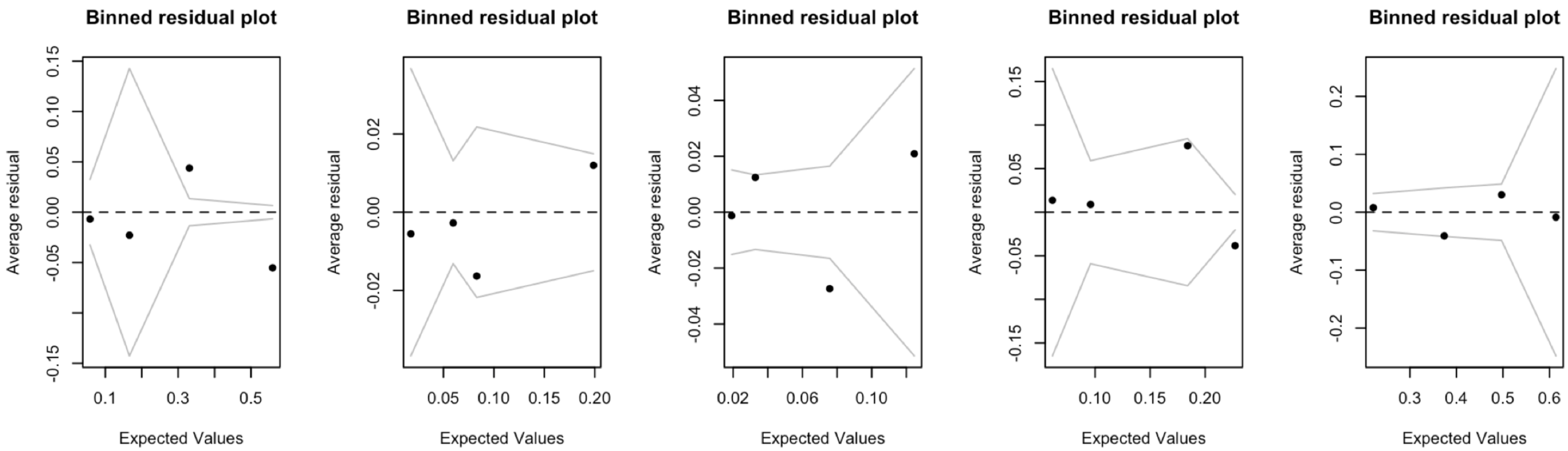
lake	size	Invertebrate	Reptile	Bird	Other	Fish
Hancock	nonadult	0.09	0.05	0.07	0.25	0.54
Hancock	adult	0.02	0.07	0.14	0.19	0.57
Ocklawaha	nonadult	0.60	0.08	0.01	0.05	0.26
Ocklawaha	adult	0.25	0.19	0.03	0.07	0.46
Trafford	nonadult	0.52	0.09	0.04	0.17	0.18
Trafford	adult	0.19	0.20	0.11	0.20	0.30
George	nonadult	0.41	0.01	0.03	0.09	0.45
George	adult	0.14	0.02	0.08	0.10	0.66

Radar plot

- By adding coord_polar()+ geom_polygon(fill=NA)
- you can get a radar plot.



Residual


Binned residual plot
Average residual
Expected Values
0.15
0.10
0.05
0.00
-0.05
-0.15
0.1
0.3
0.5
Binned residual plot
Average residual
Expected Values
0.02
0.00
-0.02
0.05
0.10
0.15
0.20
Binned residual plot
Average residual
Expected Values
0.04
0.02
0.00
-0.02
-0.04
0.02
0.06
0.10
Binned residual plot
Average residual
Expected Values
0.15
0.05
-0.05
-0.15
0.10
0.15
0.20
Binned residual plot
Average residual
Expected Values
0.2
0.1
0.0
-0.1
-0.2
0.3
0.4
0.5
0.6

High School and Beyond `(faraway)`

`data(hsb)`

- The hsb data was collected as a subset of the High School and Beyond study conducted by the National Education Longitudinal Studies program of the National Center for Education Statistics.
- The variables are gender; race; socioeconomic status; school type; chosen high school program type; scores on reading, writing, math, science, and social studies.
- We want to determine which factors are related to the choice of the type of program -- academic, vocational, or general -- that the students pursue in high school.
- The response is multinomial with three levels.v

The data in contingency table format

gender	race	academic	general	vocation
female	african-amer	6	3	4
	asian	4	3	1
	hispanic	5	3	3
	white	43	15	19
male	african-amer	3	2	2
	asian	2	1	0
	hispanic	6	1	6
	white	36	17	15