Multinomial Logistic Regression: Multicateogry Responses

Edps/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology
ILLINOIS

Outline

- Introduction and Extending binary model
- Nominal Responses (baseline model)
- SAS and R
- Inference
- Grouped Data
- Latent variable interpretation
- Discrete choice model ("conditional" model)

Additional References

General References:

- Agresti, A. (2013). *Categorical Data Analysis*, 3rd edition. NY: Wiley.
- Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.
- Powers, D.A. & Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis.* San Diego, CA: Academic Press.

Fitting (Conditional) Multinomial Models using SAS:

- SAS Institute (1995). *Logistic Regression Examples Using the SAS System,* (version 6). Cary, NC: SAS Institute.
- Kuhfeld, W.F. (2001). *Marketing Research Methods in the SAS System,* Version 8.2 Edition, TS-650. Cary, NC: SAS Institute. (reports TS-650A – TS-560I).

Additional References (continued)

Some on my course web-site,

- Anderson, C.J., and Rutkowski, L. (2007). Multinomial logistic regression models. In Osbone (Ed) Best Practices in Quantitative Methods (pp. 309–409). Thousand Oaks: Sage.
- Anderson, C.J. (2009). Categorical data analysis with a psychometric twist. In Milsap & Maydeu-Olvares (Eds) The Sage Handbook of Quantitative Methods in Psychology. (pp. 311–336). Thousand Oaks: Sage.
- Anderson, C.J., Kim, J.S., & Keller, B. (2014). Multilevel modeling of categorical response variables, In Rutkowski, von Davier, & Rutkowski (Eds) A Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis (pp. 481-519).

Situation

- Situation:
  - One response variable Y with J levels.
  - One or more explanatory or predictor variables. The predictor variables may be quantitative, qualitative or both.
- Model: "Multinomial" Logistic regression.
- What if you have multiple predictor or explanatory variables?

  Describe individuals? Descriptors of categories? or Both?

Differences w/rt Binary logistic Regression

There are 3 basic differences.

- Forming logits.
- The Distribution.
- Connections with other models (not mentioned before).

Forming Logits

- When $J = 2$, Y is dichotomous and we can model logs of odds that an event occurs or does not occur. There is only 1 logit that we can form

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$

- When $J > 2$, ...
  - We have a multicategory or "polytomous" or "polychotomous" response variable.
  - There are $J(J-1)/2$ logits (odds) that we can form, but only $(J-1)$ are non-redundant.
  - There are different ways to form a set of $(J-1)$ non-redundant logits.

How to “dichotomized” the response Y?

The most common ones

- Nomnial Y
  - “Baseline” logit models or “Multinomial” logistic regression.
  - “Conditional” or “Multinomial” logit models.
- Ordinal Y
  - Cumulative logits (Proportional Odds).
  - Adjacent categories.
  - Continuation ratios.
  - “Nested” logits

The Multinomial Distribution

- $Y_j \sim$ Mulitnomial$(\pi_1, \pi_2, \ldots, \pi_J)$ where
  - where $\sum_j \pi_j = 1$
  - $Y_j =$ number of cases in the jth category ($Y_j = 0, 1, \ldots, n$).
  - $n = \sum_j Y_j$, the number of "trials".
- Mean: $E(Y_j) = n\pi_j$
- Variance: $\text{var}(Y_j) = n\pi_j(1 - \pi_j)$
- Covariance $\text{cov}(Y_j, Y_k) = -n\pi_j\pi_k$, for $j \neq k$.
- Probability mass function,

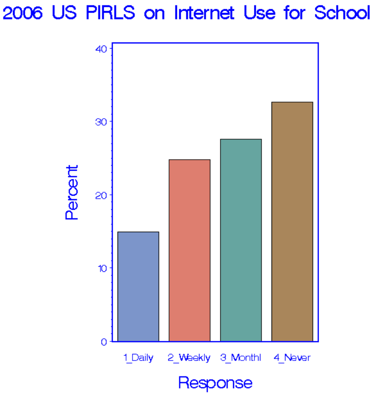
$$P(y_1, y_2, \ldots, y_J) = \left(\frac{n!}{y_1!y_2!\ldots y_J!} \right) \pi^{y_1}\pi^{y_2}\ldots\pi^{y_J}$$

- Binomial distribution is a special case.

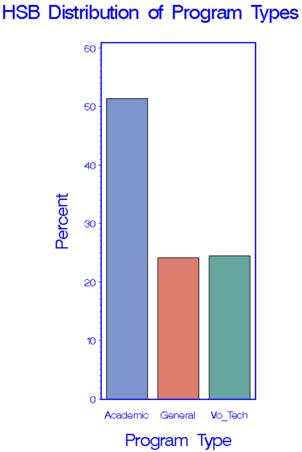
Examples of Multinomial

- High School & Beyond program types
  - General
  - Academic
  - Vo/Tech
- US 2006 Progress in International Reading Literacy Study (PIRLS) responses to item "How often to you use the Internet as a source of information for school-related work" with responses
  - Every day or almost every data ($y_1 = 746$, $p_1 = .1494$)
  - Once or twice a week ($y_2 = 1,240$, $p_2 = .2883$)
  - Once or twice a month ($y_3 = 1,377$, $p_3 = .2757$)
  - Never or almost never ($y_4 = 1,631$, $p_4 = .3266$)

Graph of PIRLS Distribution



Graph of PIRLS Distribution



Connections with Other Models

- Some are equivalent to Poisson regression or loglinear models.
- Some can be derived from (equivalent to) discrete choice models (e.g., Luce, McFadden).
- Some can be derived from latent variable models.
- Those that are equivalent to conditional multinomial models are equivalent to proportional hazard models (models for survival data), which is equivalent to Poisson regression model.
- Some multicategory logit models are very similar to IRT models in terms of their parametric form. The difference between them is that in the IRT models, the predictor is unobserved (latent), and in the model we discuss here, the predictor variable is observed.
- Others.

Multicategory Logit Models for Nominal Responses

- Baseline or Multinomial logistic regression model. Use characteristics of individuals as predictor variables.

  The parameters differ for each category of the response variable.

- Conditional Logit model. Use characteristics of the categories of the response variable as the predictors.

  The model parameters are the same for each category of the response variable.

- Conditional or Mixed logit model. Uses characteristics or attributes of the individuals and the categories as predictor variables.

Confusion

There is not a standard terminology for these models.

- Agresti (90) "Conditional Logit model": "Originally referred to by McFadden as a *conditional logit* model, it is now usually called the *multinomial logit* model."
- Long (97): Refers to the "Baseline or Multinomial logistic regression model" as a "multinomial logit" model and calls "Conditional Logit model" the "conditional logit" model.
- Powers & Xie (00) on the "Conditional" and "Multinomial" models, "However, it is often called a multinominal logit model, leading to a great deal of confusion."
- Agresti (2013) calls all of them "multinomial models" and refers to the Baseline or Multinomial logistic regression model as the "Baseline-category" model.

Further Contribution to Confusion

The models are related (connections):

- Baseline model is a special case of conditional model.
- Conditional Model can be fit as a proportional hazards model (method that at one time had to used, but now there are package(s) specifically these).
- All are special cases of Poisson log-linear models.

Baseline Category Logit Model

The models give a simultaneous representation (summary, description) of the odds of being in one category relative to being in another category for all pairs of categories.

We need a set of $(J-1)$ non-redundant odds (logits). All other can be found from this set.

This model is a special case of the binary logistic regression model.

Consider the HSB data: Program types are General, Academic and Vocational/Technical

Explanatory variables maybe

- Mean of the five achievement test scores, which is numerical/continuous (x_i).
- Socio-economic status, which will be either nominal (β_i^s) or ordinal/numerical (s_i).
- School type, which would be nominal (public, private).

Baseline Category Logit Model: HSB

We could fit a binary logit model to each pair of program types:

$$\begin{aligned}
\log\left(\frac{\text{general}}{\text{academic}}\right) = \log\left(\frac{\pi_1(x_i)}{\pi_2(x_i)}\right) &= \alpha_1 + \beta_1 x_i \\
\log\left(\frac{\text{academic}}{\text{vo/tech}}\right) = \log\left(\frac{\pi_2(x_i)}{\pi_3(x_i)}\right) &= \alpha_2 + \beta_2 x_i \\
\log\left(\frac{\text{general}}{\text{vo/tech}}\right) = \log\left(\frac{\pi_1(x_i)}{\pi_3(x_i)}\right) &= \alpha_3 + \beta_3 x_i
\end{aligned}$$

We can write one of the odds in terms of the other 2,

$$\left(\frac{\text{general}}{\text{vo/tech}}\right) = \left(\frac{\pi_1(x_i)}{\pi_2(x_i)}\right)\left(\frac{\pi_2(x_i)}{\pi_3(x_i)}\right) = \frac{\pi_1(x_i)}{\pi_3(x_i)},$$

Implication for Parameters

We can find the model parameters of one from the other two,

$$\begin{aligned}
\log\left(\frac{\pi_1(x_i)}{\pi_2(x_i)}\right) + \log\left(\frac{\pi_2(x_i)}{\pi_3(x_i)}\right) &= \log\left(\frac{\pi_1(x_i)}{\pi_3(x_i)}\right) \\
(\alpha_1 + \beta_1 x_i) + (\alpha_2 + \beta_2 x_i) &= \alpha_3 + \beta_3 x_i
\end{aligned}$$

Which means that in the *Population*

$$\begin{aligned}
\alpha_1 + \alpha_2 &= \alpha_3 \\
\beta_1 + \beta_2 &= \beta_3
\end{aligned}$$

Parameters & Sample Data

- The estimates from separate binary logit models are *consistent* estimators of the parameters of the model.
- Estimates from fitting separate binary logit models will not yield the equality between the parameters that holds in the population.

$$\hat{\alpha}_1 + \hat{\alpha}_2 \neq \hat{\alpha}_3$$
$$\hat{\beta}_1 + \hat{\beta}_2 \neq \hat{\beta}_3$$

Solution: Simultaneous estimation

- Enforces the logical relationships among parameters.
- Uses the data more *efficiently*, which means that the standard errors of parameter estimates are smaller with simultaneous estimation.

Problem with Simultaneous Estimation

Problem: There are a large number of comparisons and some of them are redundant.

Solution: Choose one of the categories and treat it as a "baseline."
Depending on the study and response variable,

- There maybe a natural choice for the baseline category.
- The choice maybe arbitrary.

Baseline Category Logit Model

For convenience, we'll use the last level of the response variable as the baseline (i.e., the Jth level or category).

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) \qquad \text{for} \quad j = 1, \ldots, J-1$$

The baseline category logit model with one explanatory variable x is

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \alpha_j + \beta_j x_i \qquad \text{for} \quad j = 1, \ldots, J-1$$

- For $J = 2$, this is just regular (binary) logistic regression.
- For $J > 2$, α and β can differ depending on which two categories are being compared.
- The odds for any pair of categories of Y that can be formed are a function of the parameters of the model.

Example: HSB Program Type

- Response variable is High school program (HSP) type where
  1. General
  2. Academic
  3. Vo/Tech
- Explanatory variable is the mean of the five achievement test scores, which is numerical/continuous (x_i).

Example: HSB Program Type

There are $(J-1) = (3-1) = 2$ non-redundant logits (odds):

$$\log\left(\frac{\text{general}}{\text{vo/tech}}\right) = \log\left(\frac{\pi_1}{\pi_3}\right) = \alpha_1 + \beta_1 x$$
$$\log\left(\frac{\text{academic}}{\text{vo/tech}}\right) = \log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x$$

The logit for (1) general and (2) academic equals

$$\begin{aligned}\log\left(\frac{\pi_1}{\pi_2}\right) &= \log\left(\frac{\pi_1/\pi_3}{\pi_2/\pi_3}\right) = \log(\pi_1/\pi_3) - \log(\pi_2/\pi_3) \\ &= (\alpha_1 + \beta_1 x) - (\alpha_2 + \beta_2 x) \\ &= (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)x\end{aligned}$$

The differences $(\beta_1 - \beta_2)$ are known as "contrasts".

Caution

- Programs that explicitly estimate the "baseline" logit model generally either set $\beta_1 = 0$ or set $\beta_J = 0$, and some set the sum $\sum_j \beta_j = 0$.
- Programs that fit the "multinomial" logit model may set $\beta_1 = 0$, $\beta_J = 0$, or $\sum_j \beta_j = 0$.

Estimated Model for HSB

$$\begin{aligned} \text{general/votech:} \quad \hat{\log}(\pi_1/\pi_3) &= -2.8996 + .0599x \\ \text{academic/votech:} \quad \hat{\log}(\pi_2/\pi_3) &= -7.9388 + .1699x \end{aligned}$$

And for comparing general and academic

$$\begin{aligned} \hat{\log}(\pi_1/\pi_2) &= \hat{\log}(\pi_1/\pi_3) - \hat{\log}(\pi_2/\pi_3) \\ &= -2.8996 + .0599x - (-7.9388 + .1699x) \\ &= 5.039 - .110x \end{aligned}$$

If we use either general or academic instead of vo/tech as the baseline category, we get the exact same results.

Interpretation

For a 1 unit change in achievement,

- Odds of General vs Vo/Tech $= \exp(.0599) = 1.06173 \sim 1.062$
- Odds of Academic vs Vo/Tech $= \exp(.1699) = 1.185186 \sim 1.185$
- Odds of General to Academic, $= \exp(-.110) = 0.8958341 \sim 0.896$

For a 10 point change in achievement, yields odds ratios

- General to Votech $= \exp(10(.0599)) = 1.82.$
- Academic to Votech $= \exp(10(.1699)) = 5.47.$
- General to Academic $= \exp(10(-.110)) = .33.$
  (or Academic to General $= 1/.33 = 3.00.$)

Showing that Simultaneous is Better

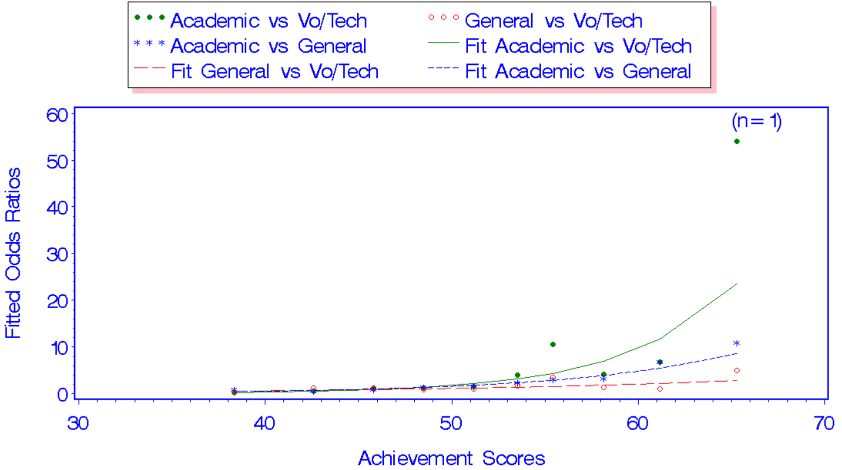
The binary logistic regression model was fit separately to 2 of the 3 possible logits,

$$\log\left(\frac{\pi_1}{\pi_3}\right) = \alpha_1 + \beta_1 x$$

$$\log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x$$

		Simultaneous Fit		Separate Fit	
Parameter		Estimate	ASE	Estimate	ASE
Intercept	(general)	-2.8996	.8156	-2.9656	.8342
	(academic)	-7.9385	.8438	-7.5311	.8572
Achieve	(general)	.0599	.0169	.0613	.0172
	(academic)	.1699	.0168	.1618	.0170

How Well does it Fit?



Computing Probabilities

Just as in logistic regression for $J = 2$, we can talk about (and interpret) baseline category logit model in terms of probabilities.

The probability of a response being in category j is

$$\pi_j = \frac{\exp(\alpha_j + \beta_j x)}{\sum_{h=1}^{J} \exp(\alpha_h + \beta_h x)}$$

Note:

- The denominator $\sum_{h=1}^{J} \exp(\alpha_h + \beta_h x)$ ensures that $\sum_{j=1}^{J} \pi_j = 1$.
- $\alpha_J = 0$ and $\beta_J = 0$ (baseline), which is an identification constraint.

Probabilities and Observed Proportions

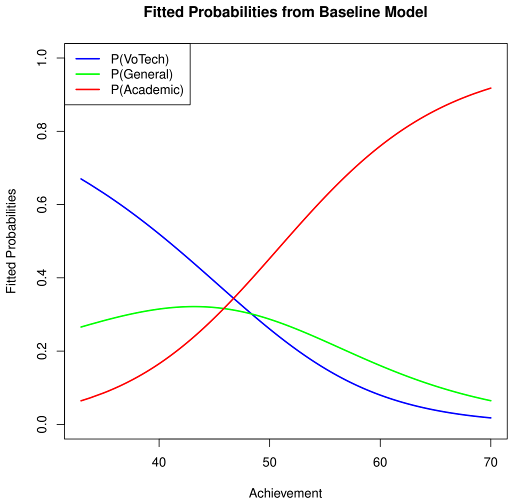
Example: High school and beyond

$$\hat{\pi}_{\text{votech}} = \frac{1}{1 + \exp(-2.90 + .06x) + \exp(-7.94 + .17x)}$$

$$\hat{\pi}_{\text{general}} = \frac{\exp(-2.90 + .06x)}{1 + \exp(-2.90 + .06x) + \exp(-7.94 + .17x)}$$

$$\hat{\pi}_{\text{academic}} = \frac{\exp(-7.94 + .17x)}{1 + \exp(-2.90 + .06x) + \exp(-7.94 + .17x)}$$

Plot of Probabilities



SAS

SAS:

- CATMOD (we'll skip this but code is here)
- GENMOD (data in long form)
- Logistic (my recommendation for most purposes).

R:

- nnet package and function *multinom*
- `relevel` command if you want to change the reference category
- glm for data in long format

SAS: PROC LOGISTC

Input:

```
proc logistic data=hsb;
model hsp = achieve / link=glogit;
```

Output: The LOGISTIC Procedure

```
                Model Information
 Data Set                          WORK.HSB
 Response Variable                 program
 Number of Response Levels         3
 Model                             generalized logit
 Optimization Technique            Newton-Raphson

 Number of Observations Read       600
 Number of Observations Used       600
```

SAS: PROC LOGISTC (continued)

```
      Response Profile
 Ordered                 Total
   Value  program    Frequency
       1  academic         308
       2  general          145
       3  vocation         147
Logits modeled use program='vocation' as the reference
category.
Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.
```

SAS: PROC LOGISTIC (continued)

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1240.134	1091.783
SC	1248.928	1109.371
-2 Log L	1236.134	1083.783

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	152.3507	2	< .0001
Score	138.0119	2	< .0001
Wald	112.7033	2	< .0001

SAS: PROC LOGISTC (continued)

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
achieve	2	112.7033	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	program	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	academic	1	-7.9388	0.8439	88.5061	< .0001
Intercept	general	1	-2.8996	0.8156	12.6389	0.0004
achieve	academic	1	0.1699	0.0168	102.7046	< .0001
achieve	general	1	0.0599	0.0168	12.7666	0.0004

SAS: PROC LOGISTIC (continued)

Odds Ratio Estimates

Effect	program	Point Estimate	95% Wald Confidence Limits	
achieve	academic	1.185	1.147	1.225
achieve	general	1.062	1.027	1.097

SAS: PROC GENMOD

Trick to use SAS/GENMOD: re-arrange the data. (works in R too using *glm*

Consider the data as a 2–way, (Student × Program type) table:

		Program Type			
		general	academic	vo/tech	
	1	1	0	0	1
	2	1	0	0	1
Student	3	0	1	0	1
	⋮	⋮	⋮	⋮	⋮
	600	0	0	1	1

The saturated loglinear model for this table is

$$\log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P + \lambda_{ij}^{SP}$$

SAS: PROC GENMOD (continued)

Associated with each row/student is a numerical variable, "achieve". Consider "Student" as being ordinal and fit a nominal by ordinal loglinear model where the achievement test scores x_i are the category scores:

$$\log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P + \beta_j^* x_i$$

We can convert the nominal by ordinal loglinear model into a logit model. For example, comparing General (1) and Vo/Tech (3):

$$\begin{aligned}\log\left(\frac{\mu_{i1}}{\mu_{i3}}\right) &= \log(\mu_{i1}) - \log(\mu_{i3}) \\ &= (\lambda_1^P - \lambda_3^P) + (\beta_1^* - \beta_3^*)x_i \\ &= \alpha_1 + \beta_1 x_i\end{aligned}$$

SAS: PROC GENMOD (continued)

```
data hsp2;
input student hsp count achieve;
datalines;
   1   1   1   41.32
   1   2   0   41.32
   1   3   0   41.32
   ⋮   ⋮   ⋮       ⋮
 600   1   0   43.44
 600   2   0   43.44
 600   3   1   43.44
proc genmod;
class student hsp;
model count = student hsp hsp*achieve / link=log dist=Poi;
```

SAS: PROC GENMOD (continued)

```
proc genmod;
class student hsp;
model count = student hsp hsp*achieve / link=log dist=Poi;
```

- "Student" ensures that the sum of each row of the fitted values equals 1 (fixed by design) — the λ_i^S's or "nuisance" parameters.
- "HSP" ensures that the program type margin is fit perfectly — the λ_j^P's which gives us the α_j's in the logit model.
- "HSP*achieve" — the β_j^* which gives the parameter estimates for the β_j's in the logit model.

SAS: PROC GENMOD (continued)

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
...								
student	596	1	0.2231	1.4145	-2.5492	2.9954	0.02	0.8747
student	597	1	-0.7416	1.4171	-3.5190	2.0358	0.27	0.6007
student	598	1	-1.0972	1.4203	-3.8809	1.6865	0.60	0.4398
student	599	1	-0.2319	1.4145	-3.0042	2.5405	0.03	0.8698
student	600	0	0.0000	0.0000	0.0000	0.0000	.	.
program	Academic	1	-7.9388	0.8439	-9.5927	-6.2848	88.51	<.0001
program	General	1	-2.8996	0.8156	-4.4982	-1.3010	12.64	0.0004
program	votech	0	0.0000	0.0000	0.0000	0.0000	.	.
achieve*program	Academic	1	0.1699	0.0168	0.1370	0.2027	102.70	<.0001
achieve*program	General	1	0.0599	0.0168	0.0271	0.0928	12.77	0.0004
achieve*program	votech	0	0.0000	0.0000	0.0000	0.0000	.	.

SAS: PROC GENMOD (continued)

SAS/GENMOD sets $\lambda_3^P = 0$ and $\beta_3^* = 0$, you get the correct ASE errors for the α_j's and β_j's:
Since

$$\alpha_j = (\lambda_j^P - \lambda_3^P) = \lambda_j^P$$

the ASE of α_j simply equals the ASE of λ_j^P.
Since

$$\beta_j = (\beta_j^* - \beta_3^*) = \beta_j^*$$

the ASE of β_j simply equals ASE of β_j^*.

SAS: PROC CATMOD

For sake of completeness. . .

```
proc catmod data=hsb;
response logits;
direct achieve ;
model hsp = achieve ;
title 'PROC CATMOD';
run;
```

R input

```
# for baseline multinomial logistic regression
library(nnet)
hsb ← read.table("hsb_data.txt",header=TRUE)
hsb$hsp ← as.factor(hsb$ hsp)
summary( mlogit ← multinom(hsp ∼ achieve, data = hsb) )
# Change the reference category to VoTech, which is hsp=3.
# This will give same parameters that are in the lecture
#notes
hsb$hsp2 ← relevel(hsb$hsp, ref = "3")
summary( mlogit <- multinom(hsp2 ∼ achieve, data = hsb) )
```

R output

```
# weights:  9 (4 variable)
initial value 659.167373
iter 10 value 541.891723
iter 10 value 541.891723
final value 541.891723
converged
Call:
multinom(formula = hsp2 ~ achieve, data = hsb)
Coefficients:
    (Intercept)     achieve
 1    -2.899594  0.05994778
 2    -7.938918  0.16985890
 Std.  Errors:
                (Intercept)     achieve
 1                0.8156199  0.01677800
 2                0.8438590  0.01676059
Residual Deviance:  1083.783
AIC: 1091.783
```

R using long format

```
hsb$id ← as.factor( seq(1:length(hsb$ses)
t ← table(hsb$id,hsb$hsp)
hsb.hsp ← as.data.frame(t)
names(hsb.hsp) ← c("id","program","y")
long <- merge(hsb, hsb.hsp, by=c("id"))
long$id ← as.factor(long$id)
long$program ← as.factor(long$program)
# -- take a little longer but works fine
summary(long.mod ← glm(y ∼ id + program + program*achieve,
data=long, family=poisson))
```

R using long format: output

```
                     Estimate  Std. Error  z value  Pr(>|z|)
 (Intercept)        -0.627712    1.003060   -0.626  0.531448
 id2                -0.091960    1.414257   -0.065  0.948155
 ⋮                      ⋮            ⋮          ⋮        ⋮
 id597              -2.373473    1.441443   -1.647  0.099641    .
 id598              -2.729093    1.449542   -1.883  0.059737    .
 id599              -1.863725    1.431436   -1.302  0.192918
 id600              -1.631872    1.427559   -1.143  0.252989
 program1           -2.899618    0.815618   -3.555  0.000378  ***
 program2           -7.938763    0.843852   -9.408   < 2e-16  ***
 achieve                   NA          NA       NA        NA
 program1:achieve    0.059948    0.016778    3.573  0.000353  ***
 program2:achieve    0.169856    0.016760   10.134   < 2e-16  ***
 ---
Null deviance:  1318.3 on 1799 degrees of freedom
Residual deviance:  1083.8 on 1196 degrees of freedom
AIC: 3491.8
```

Statistical Inference

There are 2 kinds of tests we'll talk about here:

1. Test whether an explanatory variable is related to the response variable.
2. Test whether the parameters for two (or more) categories of the response variable are the same.

Both of these tests can be done using either Wald or likelihood ratio (LR) tests. We'll talk about LR tests here; see Long (1997) for the Wald tests.

LR Test on Regression Parameters

Test whether an explanatory/predictor variable is not related to the response; that is,

$$H_o : \beta_{k1} = \ldots = \beta_{kJ} = 0$$

for the kth explanatory variable.

Example of LR test: Consider HSB example but now include SES as a nominal variable and then as an ordinal variable.

Model	-2Log(like)	Δdf	ΔG^2	p-value
achieve, nominal SES	1064.666	—	—	—
achieve, ordinal SES	1068.240	2	3.57	.16
achieve	1083.783	2	15.54	$< .001$

Wald Tests on Regression Parameters

Test whether an explanatory/predictor variable is not related to the response; that is,

$$H_o : \beta_{k1} = \ldots = \beta_{kJ} = 0$$

for the kth explanatory variable.

$$LR = 1083.783 - 1064.666 = 19.117 \qquad df = 4 \qquad p - value < .001$$

Parameters from model with SES as qualitative/nominal variable

	Standard	Wald					
Parameter		program	DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept		academic	1	-7.4105	0.8683	72.8340	<.0001
Intercept		general	1	-3.1096	0.8541	13.2538	0.0003
achieve		academic	1	0.1611	0.0173	86.8168	<.0001
achieve		general	1	0.0654	0.0174	14.0527	0.0002
ses	1	academic	1	-0.3297	0.1887	3.0517	0.0807
ses	1	general	1	0.2220	0.1868	1.4119	0.2347
ses	2	academic	1	-0.2806	0.1560	3.2351	0.0721
ses	2	general	1	-0.2477	0.1656	2.2385	0.1346

Test Two Responses the Same

Do two (or more) response categories have the same parameter estimates (i.e., can they be combined?).

If two response categories, j and j', are indistinguishable with respect to the variables in the model, then

$$H_o : (\beta_{1j} - \beta_{1j'}) = \ldots = (\beta_{Kj} - \beta_{Kj'}) = 0$$

for the K explanatory variables.

Why don't we have to consider the α's?

There are two LR tests that can be used:

- I. Fit the model with no restrictions on the parameters, and then fit the model restricting the parameters to be equal.
- II. Fit a binary logistic regression model to the two response categories in question.

Method I

Example: Consider the model with just mean achievement as the explanatory variable.

Method I:

Multinomial baseline model	G^2	Δdf	ΔG^2	p-value
No restrictions	1083.7834	—	—	—
$\hat{\beta}_1 = \hat{\beta}_3$	1097.0522	1	13.27	$< .001$

Not surprising Wald for $H_o : \beta_{general} = 0 = \beta_{votech}$ is significant.

Notes regarding Method I:

- This can be done using SAS/GENMOD,but not LOGISTIC or CATMOD.
- Also work in R using glml.
- The trick is to create a new variable that is used to impose the equality restriction.

Method I in SAS

```
data hsbI;
set expand;
* Create a new dummy variable for equating parameters for votech and
general;
if program="general" or program="votech" then xhsp=0;
else xhsp=1;
run;
title 'Full Model (no restrictions)';
proc genmod data=hsbI;
class student program;
model Y = student program program*achieve / link=log dist=poi;
run;
```

Method I & SAS (continued)

```
title 'Equate the slope parameters for votech and general';
proc genmod data=hsbl;
class student program;
model Y = student program xhsp*achieve / link=log dist=poi;
run;
```

In the model, “program” is categorical and “xhsp” is numerical.

Method I in R

Create "long" data set are previously described (around page 46)

```
# no restrictions
summary( logit.3slopes ← glm(y ∼ id + program +
program*achieve, data=long, family=poisson))

# restrictions
long$program2 ← ifelse(long$program=="2", 0, 1)
table(long$program2,long$program)

summary( logit.2slopes ← - glm(y ∼ id + program +
program2*achieve, data=long, family=poisson))

anova(logit.2slopes,logit.3slopes,test="LRT")
```

Advantages of Method I

- You can use this method to check whether a sub-set of or specific parameters are equal.
- You can use this trick to see if the parameters for more than two response categories are the same.
- It uses all the data, which is different from. . .

Method II

Using the binary logistic regression model to test

$$H_o : (\beta_{1j} - \beta_{2j'}) = \ldots = (\beta_{Kj} - \beta_{Kj'}) = 0$$

for the K explanatory variables.

1. Create a new data set that only contains the observations from response categories j and j'.
2. Fit the binary logistic regression model to the new data set.
3. Compute the likelihood ratio statistic that all the slope coefficients (β_k's) are simultaneously equal to 0 — not the intercept term..

Example: We have

$$LR = 13.76 \text{ with } df = 1,\ p < .001.$$

Method I and II

Notes regarding Methods I and II:

- In this case, both methods yield same conclusion and similar test statistics (13.76 vs 13.27).
- Method I is more flexible in terms of the range of possible tests that can be performed.
- Method I uses all of the data.
- The Method II is much easier. Just how easy this is,

```
data hsbGV;
set hsb;
if program="academic" then delete;
proc logistic data=hsbGV;
model program = achieve / link=glogit;
```

Method II in SAS

```
data hsb_GV;
set hsb;
if program="Academic" then delete;
run;

title 'Method II';
proc logistic data=hsb_GV;
model program = achieve / link=glogit;
run;
```

Method II in R

```
hsb.VG ← sample(hsb[(hsb$hsp != "2"),])

summary(VGmod ← glm(hsp ∼ achieve, data=hsb.VG,
family=binomial))

anova(VGmod,test="LRT")
```

Baseline Logit model & Grouped Data

NYLS Example from Powers & Xie (2000) *Statistical Methods for Categorical Data Analysis* (1st edition). page 236=238.

$n = 978$ of 20-22 year old men from NYLS.

		Employment Status		
	Father's	In school	Working	Inactive
Race	education	1	2	3
White/other	≤ 12 yr	204	195	131
Black	≤ 12 yr	100	53	67
White/other	> 12 yr	78	90	28
Black	> 12 yr	12	5	9

NYLS Example of Grouped Data

All log-linear models should include the Race × Father's education interaction.

E = Employment status, F = Father's education, R = Race

Model as a Loglinear	Logit	df	G^2	p-value
(RF,E)	null	6	35.7151	$< .01$
(RF,RE)	(R)	4	12.4426	$.01$
(RF,FE)	(F)	4	23.8428	$< .01$
(RF,RE,FE)	(R,F)	2	3.6659	$.16$
(RFE)	(R,E,RE)	0	0	—

Look at paramters...

Using Log-linear/Logit Connection

The best log-linear model is

$$\log(\mu_{ijk}) = \lambda + \lambda_i^E + \lambda_j^R + \lambda_k^F + \lambda_{jk}^{RF} + \lambda_{ij}^{ER} + \lambda_{ik}^{EF}$$

The corresponding logit model taking “not working” as baseline,

$$\begin{aligned} \log\left(\frac{\mu_{ijk}}{\mu_{3jk}}\right) &= (\lambda_i^E - \lambda_3^E) + (\lambda_{ij}^{ER} - \lambda_{3j}^{RE}) + (\lambda_{ik}^{EF} - \lambda_{3k}^{EF}) \\ &= \alpha_i + \beta_{ij}^R + \beta_{ik}^F \end{aligned}$$

for $i = 1$ (in school) and $i = 2$ (working).

If last category (baseline) has parameter $= 0$, then ASE of logit will be same as in log-linear.

Log-linear & Logit parameters

Dummy coding: $F = 1$ for father's education > 12, 0 for ≤ 12
$R = 1$ for Black, 0 for White/other.

Log-linear Model (RF,RE,FE)			Logit Model (R,F)			
Parameter	Est.	s.e.	Parameter	Est.	s.e.	odds ratio
λ	4.8577	0.0868				
λ_1^F	-1.4474	0.1854				
λ_1^R	-0.6196	0.1425				
λ_{11}^{RF}	-0.8846	0.2090				
λ_1^E	0.4529	0.1102	α_1	0.4529	0.1102	
λ_2^E	0.4346	0.1111	α_2	0.4346	0.1111	
λ_3^E	0.0000	0.0000				
λ_{11}^{ER}	-0.0706	0.1796	β_1^R	-0.0706	0.1796	0.93
λ_{21}^{ER}	-0.7769	0.2026	β_2^R	-0.7769	0.2026	0.46
λ_{31}^{ER}	0.0000	0.0000				
λ_{11}^{EF}	0.5130	0.2160	β_1^E	0.5130	0.2160	1.67
λ_{21}^{EF}	0.6117	0.2186	β_2^E	0.6117	0.2186	1.83
λ_{31}^{EF}	0.0000	0.0000				

Logistic Regression as Latent Variable Model

The baseline multinomial (and binary) logistic regression models can be derived as a Random Utility Model or Discrete Choice Model.

A simple version...

- Let ψ_{ij} be the underlying value of person i's utility of option j.
- We assume

$$\psi_{ij} = \beta_{1j}x_{1i} + \beta_{2j}x_{2i} + \ldots + \beta_{pj}x_{pi} + \epsilon_{ij}$$

- There are J *utility functions*
- Observed variable depends on ψ_{ij},

$$y_{ij} = j \qquad \text{if} \quad \psi_{ij} > \psi_{ij'} \qquad \text{for all} \quad j \neq j'$$

That is, choose j if it has the larger ψ_{ij} — maximize utility.

Logistic Regression as Latent Variable Model

Assumptions for ϵ_{ij} are independent and

- If $\epsilon_{ij} \sim N(0, \sigma^2)$, then have a Thurstonian model.
- If $\epsilon_{ij} \sim$ Gumbel (extreme value) distribution, then Y_{ij} follows a baseline multinomial model.

Conditional Logistic Regression Model

- In Psychology, this is either Bradley & Terry (1952) or the Luce (1959) choice model.
- In business/economics, this is McFadden's (1974) conditional logit model.

Situation: Individuals are given a set of possible choices, which differ on certain attributes. We would like to model/predict the probability of choices using the attributes of the choices as explanatory/predictor variables.

Examples

- Subjects are given 8 chocolate candies and asked which one they like the best where the explanatory variables are type of chocolate, texture, and whether includes nuts.
- Individuals must choose which of 5 brands of a product that they prefer where the explanatory variable is the price of the product. The company presents different combinations of prices for the different brands to see how much of an effect this has on choice behavior.
- The classic example: choice of mode of transportation (eg, train, bus, car). Characteristics or attributes of these include time waiting, how long it takes to get to work, and cost.

Conditional Logistic Regression Model

- The coefficients of the explanatory variables are the same over the categories (choices) of the response variable.
- The values of the explanatory variables differ over the outcomes (and possibly over individuals).

$$\pi_j(x_{ij}) = \frac{\exp[\alpha + \beta x_{ij}]}{\sum_{j\epsilon C_i} \exp[\alpha + \beta x_{ij}]}$$

where

- x_{ij} is the value of the explanatory variable for individual i and response choice j.
- The summation in the denominator is over response options/choices that individual i is given.

Properties of the Model

- The odds that individual i chooses option j versus k is a function of the difference between x_{ij} and x_{ik}:

$$\log\left(\frac{\pi_j(x_{ij})}{\pi_k(x_{ik})}\right) = \beta(x_{ij} - x_{ik})$$

- The odds of choosing j versus k does not depend on any of the other options in the choice set or the other options' values on the attribute variables.

  Property of "Independence from Irrelevant Alternatives".
- The multinomial/baseline model can be written in the same form as the conditional logit model model (see Agresti, 2013; Anderson & Rutkowski, 2008; Anderson, 2009). Implications...
- This model can incorporate attributes or characteristics of the decision maker/individual.
- It can be written as a proportional hazard model. Implications....

Example 1: Choice of Chocolates

Hypothetical: SAS Logistic Regression examples, 1995; Kuhfeld, 2001.
The model that was fit is

$$\pi_j(c_j, t_j, n_j) = \frac{\exp[\alpha + \beta_1 c_j + \beta_2 t_j + \beta_3 n_j]}{\sum_{h=1}^{8}(\exp[\alpha + \beta_1 c_h + \beta_2 t_h + \beta_3 n_h])}$$

where

- Type of chocolate is dummy coded:

$$c_j = \begin{cases} 1 & \text{if milk} \\ 0 & \text{if dark} \end{cases}$$

- Texture is dummy coded:

$$t_j = \begin{cases} 1 & \text{if hard} \\ 0 & \text{if soft} \end{cases}$$

- Nuts is dummy coded:

$$n_j = \begin{cases} 1 & \text{if no nuts} \\ 0 & \text{if nuts} \end{cases}$$

Example 1: Odds

In terms of Odds:

$$\frac{\pi_j(c_j, t_j, n_j)}{\pi_k(c_k, t_k, n_k)} = \exp[\beta_1(c_j - c_k)]\exp[\beta_2(t_j - t_k)]\exp[\beta_3(n_j - n_k)]$$

parameter	df	value	ASE	Wald	p	$\exp\beta$	
α	1	-2.88	1.03	7.78	.01	—	
Type of chocolate							
milk	1	-1.38	.79	3.07	.08	.25	or $(1/.25) = 4.00$
dark	0	0.00					
Texture							
hard	1	2.20	1.05	4.35	.04	9.00	
soft	0	0.00					
Nuts							
no nuts	1	-.85	.69	1.51	.22	.43	or $(1/.43) = 2.33$
nuts	0	0.00					

Example 1: Ranking

Use $\exp\beta$ for interpretation.

The predicted probabilities.

Rank	Dark	Soft	Nuts	$\hat{pi}$
1	dark	hard	nuts	0.50400
2	dark	hard	no n	0.21600
3	milk	hard	nuts	0.12600
4	dark	soft	nuts	0.05600
5	milk	hard	no n	0.05400
6	dark	soft	no n	0.02400
7	milk	soft	nuts	0.01400
8	milk	soft	no n	0.00600

Estimation

In SAS:

- PHREG (proportional hazard model)
- GENMOD
- MDC (multinomial discrete choice model)

In R:

- glm as a poisson regression
- mlogit package
- mnlogit package (alternative to mlogit; however, it was recently archived.)
- others

Long data format (for SAS)

```
data chocs;
    title 'Chocolate Candy Data';
    input subj choose dark soft nuts @@;
     t=2-choose;
     if dark=1 then drk='dark'; else drk='milk';
     if soft=1 then sft='soft'; else sft='hard';
     if nuts=1 then nts='nuts'; else nts='no nuts';
    datalines;
 1  0  0  0  0      1  0  0  0  1   ...
 1  1  1  0  0      1  0  1  0  1
 2  0  0  0  0      2  0  0  0  1
 2  0  1  0  0      2  1  1  0  1
 ⋮                  ⋮
```

Long data format (for R)

head(c)

subj	choice	dark	soft	nuts
1	0	0	0	0
1	0	0	0	1
1	0	0	1	0
1	0	0	1	1
1	1	1	0	0
1	0	1	0	1
1	0	1	1	0
1	0	1	1	1
2	0	0	0	0
2	0	0	0	1
2	0	0	1	0
2	0	0	1	1
⋮				

Using GLM

```
summary( glm.choc ← glm(choice ∼ dark + soft + nuts,
                        data=choc, family=poisson) )

beta ← glm.choc$coefficients

#For the probabilities:
pi0 ←- exp(beta[1]+ beta[2]*choc$dark[1:8]
+beta[3]*choc$soft[1:8] + beta[4]*choc$nuts[1:8])

(pi.hat ← pi0/sum(pi0))

# For easier of knowing what's what
probs ← cbind(choc[1:8,],pi.hat)
(pi.hat ← probs[order(probs$pi.hat),])
```

Proportional hazard model

- It's typically used for modeling survival data; that is, modeling the time until death (or other event of interest).
- It's equivalent to a Poisson regression for the number of deaths and to a negative exponential for survival times.
- For more details see Agresti (2013).

Using SAS PROC PHREG:

```
proc phreg data=chocs outest=betas;
 strata subj;
 model t*choose(0)=dark soft nuts;
run;
```

Relevant Output from PHREG

```
             Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.
            Model Fit Statistics

               Without                       With
Criterion   Covariates                 Covariates

-2 LOG L        41.589                     28.727
AIC             41.589                     34.727
SBC             41.589                     35.635
```

Relevant Output from PHREG

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
dark	1	1.38629	0.79057	3.0749	.0795
soft	1	-2.19722	1.05409	4.3450	.0371
nuts	1	0.84730	0.69007	1.5076	.2195

Using GENMOD

```
proc genmod data=chocs;
class subj dark soft nuts;
model choose = dark soft nuts /link=log dist=poi obstats;
ods output ObStats=ObStats;
run;
proc sort data=ObStats;
by subj pred;
run;
title 'Predicted probabilities for different chocolates';
proc print data=ObStats;
where subj="1";
var dark soft nuts pred ;
run;
```

Relevant Output from GENMOD

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Limits			Wald Chi Square
Intercept		1	-2.8824	1.0334	-4.9078	-0.8570	7.78	0.0053
dark	0	1	-1.3863	0.7906	-2.9358	0.1632	3.07	0.0795
dark	1	0	0.0000	0.0000	0.0000	0.0000	.	.
soft	0	1	2.1972	1.0541	0.1312	4.2632	4.35	0.0371
soft	1	0	0.0000	0.0000	0.0000	0.0000	.	.
nuts	0	1	-0.8473	0.6901	-2.1998	0.5052	1.51	0.2195
nuts	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

Using PROC MDC

Documentation is not under the STAT, but under ETS (econometrics).

```
proc mdc data=chocs;
model choose = dark soft nuts / type=clogit nchoice=8 covest=hessian;
id subj;
run;
```

Output:

Conditional Logit Estimates
Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx $Pr > \|t\|$
dark	1	1.3863	0.7906	1.75	0.0795
soft	1	-2.1972	1.0541	-2.08	0.0371
nuts	1	0.8473	0.6901	1.23	0.2195

R using glm

```
summary( glm.choc ← glm(choice ∼ dark + soft + nuts,
data=choc, family=poisson) )

# For the probabilities:
beta ← glm.choc$coefficients

pi0 ← exp(beta[1]+ beta[2]*choc$dark[1:8]
+beta[3]*choc$soft[1:8] + beta[4]*choc$nuts[1:8])
(pi.hat ← pi0/sum(pi0))

# For easier of knowing what's what
probs ←cbind(choc[1:8,],pi.hat)
(pi.hat ← probs[order(probs$pi.hat),])
```

Example 2: Brand and price

Five brands that differ in terms of price where price is manipulated. For each of the 8 combinations of brand and price included in the study.
The data:

```
data brands;
   input p1-p5 f1-f5;
datalines;
 5.99  5.99  5.99  5.99  4.99  12  19  22  33  14
 5.99  5.99  3.99  3.99  4.99  34  26   8  27   5
 5.99  3.99  5.99  3.99  4.99  13  37  15  27   8
 5.99  3.99  3.99  5.99  4.99  49   1   9  37   4
 3.99  5.99  5.99  3.99  4.99  31  12   6  18  33
 3.99  5.99  3.99  5.99  4.99   4  29  16  42   9
 3.99  3.99  5.99  5.99  4.99  37  10   5  35  13
 3.99  3.99  3.99  3.99  4.99  16  14   5  51  14
```

Example 2: Brand and price (continued)

In all models that we fit, we assume (i.e., fit a parameter) for brand preference.

The two models that are fit:

1. The effect of price does not depend on brand ($G^2 = 2782.4901$)
2. The effect of price depends on the brand; that is, the strength of brand loyalty depends on price ($G^2 = 2782.4901$)..

LR statistic for testing whether effect of price depends on brand:

$$G^2 = 2782.4901 - 2782.0879 = .4022, \qquad df = 3, \qquad p = .94$$

Example 2: The models

The simpler model. . .

$$\pi_j(b_{1j}, b_{2j}, b_{3j}, b_{4j}, p_j) = \frac{\exp[\alpha + \beta_1 b_{1j} + \beta_2 b_{2j} + \beta_3 b_{3j} + \beta_4 b_{4j} + \beta_5 p_j]}{\sum_{h=1}^{5} \exp[\alpha + \beta_1 b_{1h} + \beta_2 b_{2h} + \beta_3 b_{3h} + \beta_4 b_{4h} + \beta_5 p_h]}$$

- Brands are dummy coded. Eg,

$$b_{1j} = \begin{cases} 1 & \text{if brand is 1} \\ 0 & \text{otherwise} \end{cases}$$

- Price is a numerical variable, p_j.

Or in terms of odds:

$$\frac{\pi_j(b_{1j}, b_{2j}, b_{3j}, b_{4j}, p_j)}{\pi_k(b_{1k}, b_{2k}, b_{3k}, b_{4k}, p_k)} = \exp[\beta_1(b_{1j} - b_{1k})] \exp[\beta_2(b_{2j} - b_{2k})] \exp[\beta_3(b_{3j} - b_{3k})] \exp[\beta_4(b_{4j} - b_{4k})] \exp[\beta_5(p_j - p_k)]$$

Example 2: The Estimates

Variable		DF	Parameter Estimate	Standard Error	Chi-Square	p	$\exp \hat{\beta}$
brand1	β_1	1	0.66727	0.12305	29.4065	$< .0001$	1.95
brand2	β_2	1	0.38503	0.12962	8.8235	0.0030	1.47
brand3	β_3	1	-0.15955	0.14725	1.1740	0.2786	.85
brand4	β_4	1	0.98964	0.11720	71.2993	$< .0001$	2.69
brand5	—	0	0	.	.	.	1.00
price	β_5	1	0.14966	0.04406	11.5379	0.0007	1.16

- Which brand is the most preferred?
- Which brand is least preferred?
- What is the effect of price?

  How would you interpret $\exp[.1497] = 1.16$?

Estimation using GENMOD

Format of data needed for input to GENMOD:

```
data brands2;
   input combo brand price choice @@;
   datalines;

1  1  5.99  12     1  2  5.99  0      1  3  5.99  0
1  1  5.99  0      1  2  5.99  19     1  3  5.99  0
1  1  5.99  0      1  2  5.99  0      1  3  5.99  22
⋮
```

Estimation using GENMOD (continued)

No interaction

```
proc genmod;
 class combo brand ;
 model choice = combo brand /link=log dist=poi;
run;
```

With an interaction

```
proc genmod;
 class combo brand ;
 model choice = combo brand brand*price /link=log dist=poi;
run;
```

Estimation using R glm

```
# ...and to match SAS output
brands ← within(brands, brand ← relevel(brand, ref =
"5"))

# In models, combo is just a nuisance variable
summary( mdc1 ← glm(choice ∼ combo + brand + price,
data=brands, family=poisson) )
anova(mdc1,test="LRT")

# with interaction
summary( mdc2 ← glm(choice ∼ combo + brand + price +
brand*price, data=brands, family=poisson) )
anova(mdc2,test="LRT")

# Test between models
anova(mdc1,mdc2)
```

Estimation using MDC

Format of data needed for input to MDC:

brand1	brand2	brand3	brand4	br	price	Y	case
1	0	0	0	1	5.99	1	1
0	1	0	0	2	5.99	0	1
0	0	1	0	3	5.99	0	1
0	0	0	1	4	5.99	0	1
0	0	0	0	5	4.99	0	1
1	0	0	0	1	5.99	1	2
0	1	0	0	2	5.99	0	2
⋮							

Estimation using MDC (continued)

Using dummy codes:

```
title 'MDC for the brands and price';
proc mdc data=mdcdata;
model y = brand1 brand2 brand3 brand4 price
  / type=clogit nchoice=5 covest=hessian;
id case;
run;
```

Using Class (default are effect codes):

```
title 'MDC for the brands and price';
proc mdc data=mdcdata;
class br;
model y = br price / type=clogit nchoice=5 covest=hessian;
id case;
run;
```

mlogit in R

```
# Choice to be logical
longer$choice ← longer$Y==1

# To make mlogit happy
long.m ← dfidx(longer, shape="long", idx= list("case",
"alternative"), choice="choice")

# The model formula format:
# logical variable for choice ∼
generic alternative specific | individual specific | alternative
specific

# main effects
fm1 ← formula(choice ∼ brand1 + brand2 + brand3 + brand4
              + price | 0 | 0 )
```

mlogit in R (continued)

```
summary(model.1← mlogit(fm1,long.m) )

# Estimated odds ratios
(exp.beta ← exp(model.1$coefficients[1:5]) )

# with an interaction interaction
fm2 <- formula(choice ~ brand1 + brand2 + brand3 + brand4 +
price + brand1*price + brand2*price + brand3*price +
brand4*price | 0 | 0 )

summary(model.2 ← mlogit(fm2,long.m) )
```

See R script on course web-site.

Using PHREG

It's a real pain in this case. If you really want to know how to do this, see SAS code on the course web-site. The data manipulation is non-trivial.

Example 3: Modes of Transportation

From Powers & Xie (2000).
The Response variable is mode of transportation:
$j = 1$ for train, 2 for bus, and 3 for car.
Explanatory Variables are:

- $t_{ij} =$ time waiting in Terminal.
- $v_{ij} =$ time spent in the Vehicle.
- $c_{ij} =$ Cost of time spent in vehicle.
- $g_{ij} =$ Generalized cost measure $= c_{ij} + v_{ij}(\text{value}_{ij})$ where value equals subjective value of respondent's time for each mode of transportation.

The multinomial logit model that appears to fit the data is

$$\pi_{ij} = \frac{\exp[\beta_1 t_{ij} + \beta_2 v_{ij} + \beta_3 c_{ij} + \beta_4 g_{ij}]}{\sum_{h=1}^{3} \exp[\beta_1 t_{ih} + \beta_2 v_{ih} + \beta_3 c_{ih} + \beta_4 g_{ih}]}$$

Example 3: Modes of Transportation (continued)

The odds of choosing mode j versus mode k for individual i,

$$\frac{\pi_{ij}}{\pi_{ik}} = \exp[\beta_1(t_{ij} - t_{ik})]\exp[\beta_2(v_{ij} - v_{ik})]\exp[\beta_3(c_{ij} - c_{ik})]\exp[\beta_4(g_{ij} - g_{ik})]$$

The odds of choosing mode j versus mode k for individual i,

$$\frac{\pi_{ij}}{\pi_{ik}} = \exp[\beta_1(t_{ij} - t_{ik})]\exp[\beta_2(v_{ij} - v_{ik})]\exp[\beta_3(c_{ij} - c_{ik})]\exp[\beta_4(g_{ij} - g_{ik})]$$

Example 3: Interpretation

Variable	Parameter	Value	ASE	Wald	p-value	e^{β}	$1/e^{\beta}$
terminal, t_{ij}	β_1	$-.002$	.007	.098	.75	.99	1.002
vehicle, v_{ij}	β_2	$-.435$	.133	10.75	.001	.65	1.55
cost, c_{ij}	β_3	$-.077$	.019	15.93	$< .001$	.03	1.08
generalized cost, g_{ij}	β_4	$.431$	.133	10.48	.001	1.54	.65

Odds of choosing a particular mode of transportation decreases as

- Time waiting in terminal increases.
- Time spent in vehicle increases.
- Cost increases.

Odds of choosing a particular model of transportation increases as

- Generalized cost (value of individual's time) increases

Example 3: SAS

Only PROC MDC.

```
data transport;
input mode ttme invc invt gc hinc psize tasc basc casc id;
 hincb=basc*hinc;
 hincc=casc*hinc;
label mode='Mode of transportation choosen'
 ttime='Time in terminal'
 invc='Time in vehicle'
 gv='Generalized cost'
 hinc='Household income';
datalines;
 0  34  31  372  71  35  1  1  0  0  1
 0  35  25  417  70  35  1  0  1  0  1
 1   0  10  180  30  35  1  0  0  1  1
 0  44  31  354  84  30  2  1  0  0  2
```

Example 3: SAS

Code:

```
title 'Attributes of modes of transportation';
proc mdc data=transport;
 model mode = ttme invc invT gc / type=clogit nchoice=3
  covest=hessian;
 id ID;
run;
```

Example 3: R

Set up the data

```
trans ← read.table("transportation_data.txt",header=TRUE)
n ← length(unique(trans$id))
trans$alt ← rep(c("Bus","Train","Car"),n)
# To make mlogit happy
tm ← dfidx(trans, shape="long", idx=list("id","alt"),
choice="mode")
```

Take a look at the data

Example 3: data

```
head(tm, n=6)
```

∼∼∼∼∼∼∼

first 6 observations out of 456

∼∼∼∼∼∼∼

	mode	ttme	invc	invt	gc	hinc	psize	tasc	basc	casc	hincb h	incc	idx
1	FALSE	34	31	372	71	35	1	1	0	0	0	0	1:Bus
2	TRUE	0	10	180	30	35	1	0	0	1	0	35	1:Car
3	FALSE	35	25	417	70	35	1	0	1	0	35	0	1:rain
4	FALSE	44	31	354	84	30	2	1	0	0	0	0	2:Bus
5	TRUE	0	11	255	50	30	2	0	0	1	0	30	2:Car
6	FALSE	53	25	399	85	30	2	0	1	0	30	0	2:rain

Example 3: data

~~~ indexes ~~~

	id	alt
1	1	Bus
2	1	Car
3	1	Train
4	2	Bus
5	2	Car
6	2	Train

indexes: 1, 2
~~~

Example 3: run mlogit

```
fm.b ← formula(mode ~ ttme + invc + invt + gc | hinc | 0 )
summary(model.a ← mlogit(fm.b,tm))
```

The Mixed Model

The conditional multinomial model that incorporates attributes of the categories (choices) and of the decision maker.

This model is a combination of the multinomial and conditional multinomial modela.

Suppose

- Response variable Y has J categories/levels.
- Explanatory variables
    - x_i that is a measure of an attribute of individual i
    - w_j that is a measure of an attribute of alternative j.
    - z_{ij} that is a measure of an attribute of alternative j for individual i.

The Mixed Model

The "Mixed" Model:

$$\pi_j(x_i, w_j, z_{ij}) = \frac{\exp[\alpha_j + \beta_{1j}x_i + \beta_2 w_j + \beta_3 z_{ij}]}{\sum_{h=1}^{J} \exp[\alpha_h + \beta_{1h}x_i + \beta_2 w_h + \beta_3 z_{ih}]}$$

The odds of individual i choosing category j versus category k,

$$\begin{aligned} \frac{\pi_j(x_i, w_j, z_{ij})}{\pi_k(x_i, w_k, z_{ik})} &= \exp[\alpha_j - \alpha_k] \exp[(\beta_{1j} - \beta_{1k})x_i] \\ & \quad \exp[\beta_2(w_j - w_k)] \exp[\beta_3(z_{ij} - z_{ik})] \end{aligned}$$

Example 3 Continued

Explanatory Variables are:

t_{ij} = time waiting in Terminal.

v_{ij} = time spent in the Vehicle.

c_{ij} = Cost of time spent in vehicle.

g_{ij} = Generalized cost measure $= c_{ij} + v_{ij}(\text{value}_{ij})$ where value equals subjective value of respondent's time for each mode of transportation.

h_i = Household income.

The mixed model that appears to fit the data is

$$\pi_{ij} = \frac{\exp[\beta_1 t_{ij} + \beta_2 v_{ij} + \beta_3 c_{ij} + \beta_4 g_{ij} + \alpha_j + \beta_{5j} h_i]}{\sum_{h=1}^{3} \exp[\beta_1 t_{ih} + \beta_2 v_{ih} + \beta_3 c_{ih} + \beta_4 g_{ih} + \alpha_h + \beta_{5h} h_i]}$$

Example 3: The Odds

The odds of choosing mode j versus mode k for individual i,

$$\frac{\pi_{ij}}{\pi_{ik}} = \exp[\beta_1(t_{ij} - t_{ik})]\exp[\beta_2(v_{ij} - v_{ik})]\exp[\beta_3(c_{ij} - c_{ik})]\exp[\beta_4(g_{ij} - g_{ik})] \exp[(\alpha_j - \alpha_k)]\exp[(\beta_{5j} - \beta_{5k})h_i]$$

The odds of choosing mode j versus mode k for individual i,

$$\frac{\pi_{ij}}{\pi_{ik}} = \exp[\beta_1(t_{ij} - t_{ik})]\exp[\beta_2(v_{ij} - v_{ik})]\exp[\beta_3(c_{ij} - c_{ik})]\exp[\beta_4(g_{ij} - g_{ik})] \exp[(\alpha_j - \alpha_k)]\exp[(\beta_{5j} - \beta_{5k})h_i]$$

Example 3: Parameter Estimates

Parameter Estimates:

Variable	Parameter	Value	ASE	Wald	p-value	e^{β}	$1/e^{\beta}$
Terminal, t_{ij}	β_1	$-.074$	.017	19.01	$< .001$	.93	1.08
Vehicle, v_{ij}	β_2	$-.619$	.152	16.54	$< .001$	.54	1.86
Cost, c_{ij}	β_3	$-.096$	.022	19.02	$< .001$	.91	1.10
Generalized cost, g_{ij}	β_4	$.581$	.150	15.08	$< .001$	1.79	.56
Bus							
Intercept,	α_1	-2.108	.730	6.64	.01		
Income, h_i	β_{51}	.031	.021	1.97	.16	1.03	.97
Car							
Intercept	α_2	-6.147	1.029	35.70	$< .001$		
Income, h_i	β_{52}	.048	.023	7.19	.01	1.05	.95

Example 3: Interpretation

Effect of household income:

- The odds of choosing a bus versus a train given household income increases from h_i to $h_i + 100$ units is $\exp(100(.031)) = 22.2$ times.
- The odds of choosing a car versus a train given household income increases from h_i to $h_i + 100$ units is $\exp(100(.048)) = 121.5$ times.
- The odds of choosing a car versus a bus given household income increases from h_i to $h_i + 100$ units is $\exp(100(.048 - .031)) = \exp(1.7) = 5.5$ times.

Example 3: SAS

Mostly the same, but a little twist,

```
hincb=basc*hinc;
hincc=casc*hinc;
title 'Mixed';
proc mdc data=transport;
model mode = ttme invc invT gc basc hincb casc hincc
        / type=clogit nchoice=3 covest=hessian;
id ID;
run;
```

Next up

Multi-category logit model ordinal response variables.