Accounting for Overdispersion

Plot Pearson residuals vs. fitted values

\[ nhat \leftarrow \text{fitted( fit )} \]

\[ \text{plot( nhat, presid^2, xlab="Fitted mean response", ylab="Squared Pearson residuals")} \]

\[ \text{abline( h=1, col="red", lwd=2 )} \]

\[ \text{sfit} \leftarrow \text{loess( presid^2 \sim nhat )} \]

\[ \text{lines( sort(sfit$x), sfit$fitted[order(sfit$x)])} \]

\[ \text{abline(h=phihat, lty=2, col="red", lwd=2)} \]

• Again, it looks as though the smoother is consistently above the \( y=1 \) line, indicating overdispersion

Introduction
Multinomial Logistic Regression Model
Development
Parameter estimation
Parameter interpretation
Example - Greene (1995)
General Comments
Nominal data

Definition

- **Nominal data** are data that take values which do not have a natural ordering

- **Examples:**
  1. Choice of health plan (HMO, Blue Cross, United Health)
  2. Political party affiliation (Democrat, Republican, Independent)
  3. Choice of automobile (Toyota, Ford, Chevy, etc)
Nominal data

Notes

1. Polytomous response data are most commonly found in the econometrics literature

   ▶ McFadden (1974) is credited with introducing *multinomial* or *discrete choice* model

2. The actual values of a response are irrelevant, although they are typically coded numerically (eg. 0,1,2) no ordinal meaning is associated with the coding
Regression Models for Nominal Response Data

Extending the logistic regression model

- Suppose that the outcome variable $Y_i$ is a nominal variable with $J + 1$ categories and let

  $$\pi_{ij} = \Pr[Y_i = j]$$

  the probability that individual $i$ has response (choice) $j$, $j = 0, \ldots, J$.

- We assume that $\sum_{j=0}^{J} \pi_{ij} = 1$

- A sample of $n$ independent observations of this form result in a *multinomial* distribution with $n$ trials and cell probabilities $\pi_{i0}, \ldots, \pi_{iJ}$.
Extending the logistic regression model

- A natural extension to the logistic regression model is to consider the (log) odds associated with a particular (arbitrary?) reference category. For example,

\[
\log \left( \frac{\pi_{ij}}{\pi_{i0}} \right), \quad j = 1, \ldots, J
\]
Regression Models for Nominal Response Data

Extending the logistic regression model

Then an extension to the logistic regression model (with baseline category 0) is given by

\[ \eta_{ij} = \log \left( \frac{\pi_{ij}}{\pi_{i0}} \right) \]

\[ = \log \left( \frac{\Pr[Y_i = j|x_i]}{\Pr[Y_i = 0|x_i]} \right) \]

\[ = \beta_{0j} + \beta_{1j}x_{i1} + \cdots + \beta_{pj}x_{ip} \]

Notice that we are allowing all parameter coefficients to vary across each comparison...
Regression Models for Nominal Response Data

Extending the logistic regression model

- Under this model formulation,

\[
\pi_{i0} = \Pr[Y_i = 0|x_i] = \frac{1}{1 + \sum_{k=1}^{J} e^{\eta_{ik}}}
\]

and

\[
\pi_{ij} = \Pr[Y_i = j|x_i] = \frac{e^{\eta_{ij}}}{1 + \sum_{k=1}^{J} e^{\eta_{ik}}}
\]
Parameter estimation

- In order to efficiently estimate the parameters for the model we want to estimate them simultaneously via maximum likelihood

- Noting that $\pi_{i0} = 1 - (\pi_{i1} + \cdots + \pi_{iJ})$ and $y_{i0} = 1 - (y_{i1} + \cdots + y_{iJ})$, the contribution to the (log) likelihood for individual $i$ is given by (see handout)

$$
\log \left( \prod_{j=0}^{J} \pi_{ij}^{y_{ij}} \right) = \sum_{j=0}^{J} y_{ij} \log \pi_{ij}
$$

$$
= \sum_{j=1}^{J} y_{ij} \log \left( \frac{\pi_{ij}}{1 - \sum_{j=1}^{J} \pi_{ij}} \right) + \log \left( 1 - \sum_{j=1}^{J} \pi_{ij} \right)
$$
Regression Models for Nominal Response Data

Parameter estimation cont’d

- Now, note that

\[
\frac{\pi_{ij}}{1 - \sum_{j=1}^{J} \pi_{ij}} = \frac{e^{\eta_{ij}}}{1 + \sum_{k=1}^{J} e^{\eta_{ik}}} = e^{\eta_{ij}}
\]

- Thus the log-likelihood based upon \( n \) independent observations is given by (see handout)

\[
\log(L(\hat{\beta})) = \log \prod_{i=1}^{n} \left( \prod_{j=0}^{J} \pi_{ij}^{y_{ij}} \right)
= \sum_{i=1}^{n} \left\{ \sum_{j=1}^{J} y_{ij} \eta_{ij} - \log \left( 1 + \sum_{j=1}^{J} e^{\eta_{ij}} \right) \right\}
\]
Regression Models for Nominal Response Data

Parameter estimation cont’d

- To maximize the log-likelihood with respect to $\hat{\beta}$, we have

$$\frac{\partial \log(L(\hat{\beta}))}{\partial \beta_{jk}} = \sum_{i=1}^{n} y_{ij}x_{ik} - \frac{x_{ik} e^{\eta_{ij}}}{1 + \sum_{k=1}^{J} e^{\eta_{ik}}}$$

$$= \sum_{i=1}^{n} x_{ik} \left[ y_{ij} - \frac{e^{\eta_{ij}}}{1 + \sum_{k=1}^{J} e^{\eta_{ik}}} \right]$$

$$= \sum_{i=1}^{n} x_{ik} \left[ y_{ij} - \pi_{ij} \right]$$

$$= 0$$
Regression Models for Nominal Response Data

Parameter interpretation

- Consider the simple multinomial regression model

\[
\log \left( \frac{\pi_j}{\pi_0} \right) = \log \left( \frac{\Pr[Y = j|x]}{\Pr[Y = 0|x]} \right) = \beta_{0j} + \beta_{1j} x_j
\]

- Define the *relative risk ratio (RRR)* for outcome \(j\) versus the referent outcome 0 as

\[
RRR_j(x', x) = \frac{\Pr[Y = j|x']/ \Pr[Y = 0|x']}{\Pr[Y = j|x]/ \Pr[Y = 0|x]}
\]
Regression Models for Nominal Response Data

Parameter interpretation

➤ Then,

\[
\log[R_{j|x+1}(x+1, x)] = \log \left( \frac{\Pr[Y = j|x + 1]/\Pr[Y = 0|x + 1]}{\Pr[Y = j|x]/\Pr[Y = 0|x]} \right)
\]
Regression Models for Nominal Response Data

Parameter interpretation

- Interpretation of $\beta_{1j}$:
Regression Models for Nominal Response Data

Parameter interpretation

Question: Under this model specification, what is the effect of a 1-unit increase in $x$ on the $RRR$ comparing outcome 2 to outcome 1?
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

- Greene (1995) reported a study concerned with travelers’ decisions regarding travel mode between Sydney and Melbourne, Australia.

- The modes of travel (choices) that were given to travelers were air, train, bus, and car (coded as mode = 0, 1, 2, 3, respectively).

- Our primary concern is to quantify the association between household income hinc and travel choice, and the number of individuals traveling in a party psize and travel choice.

- Other unmeasured confounders???
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

▸ First let’s get a look at the dataset....

```r
> ##
> # Plot Pearson residuals vs. fitted values
> nhat <- fitted( fit )
> plot( nhat, presid^2, xlab="Fitted mean response",
> ylab="Squared Pearson residuals" )
> abline( h=1, col="red", lwd=2 )
> sfit <- loess( presid^2 ~ nhat )
> lines( sort(sfit$x), sfit$fitted[order(sfit$x)], col="blue", lwd=2 )
> abline(h=phihat, lty=2, col="red", lwd=2)
```

10 20 30 40
0 1 2 3 4 5 6

Fitted mean response
Squared Pearson residuals

• Again, it looks as though the smoother is consistently above the y=1 line, indicating overdispersion

Introduction
Multinomial Logistic Regression Model Development
Parameter estimation
Parameter interpretation

General Comments

Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

▸ First let’s get a look at the dataset....

```r
> source( "http://www.ics.uci.edu/~dgillen/Stat211/Handouts/Stat211Functions.q" )
>
> travel$travel <- rep( 0:3, 210 )
> travel <- travel[ travel$mode==1, ]
> travel <- travel[,c("travel", "hinc", "psize") ]
> travel[1:10,]

        travel  hinc  psize
      4     3      35     1
      8     3      30     2
     12     3      40     1
     16     3      70     3
     20     3      45     2
     22     1      20     1
     25     0      45     1
     32     3      12     1
     36     3      40     1
     40     3      70     2

> dim( travel )
[1] 210   3
```
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

- Some descriptive statistics

```r
> Some descriptive statistics
> cbind( table( travel$travel ), 
  table( travel$travel ) / sum( table( travel$travel ) ) )

[,1]     [,2]
 0   58 0.2761905
 1   63 0.3000000
 2   30 0.1428571
 3   59 0.2809524
```

- Description: There were 210 respondents to the survey (no missing data). The most frequently chosen modes of travel were air (n=58), train (n=63), and car (n=59). The least frequently chosen mode of transportation was bus (n=30).
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

▶ What about the distribution of travel mode by party size and household income by travel mode?

```r
> table( travel$travel, travel$psize )

     1 2 3 4 5 6
0  34 18 3 3 0 0
1  35 18 6 4 0 0
2  23  4 3 0 0 0
3  22 18 8 8 2 1

> lapply( split( travel$hinc, travel$travel ), FUN=summary )

"0"
  Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
  4.00  26.75  42.50  41.72  60.00   70.00

"1"
  Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
  4.00  10.00  18.00  23.06  35.00   72.00

"2"
  Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
  2.00  20.00  30.00  29.70  44.25   60.00

"3"
  Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
  4.00  30.00  40.00  42.22  55.00   70.00
```
Example - Travel Choices (Greene, 1995)

- Description: As the number of passengers increases, the probability of travel by air decreases.

- Further, the mean household income among travelers choosing air or car appears to be higher than those traveling by train or bus (means range from $23k to $42k).
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

- As a learning experience, let's first fit separate logistic regression models to the data using air travel as the reference group
- First, fit a logistic model for train vs. air

```r
> #
> ##### Fit separate logistic regression models for each choice with air as referent
> 
> Model 1: Train vs. Air
> 
> fit1 <- glm( (travel==1) ~ hinc + psize, data=travel,
>              family=binomial, subset=(travel==0 | travel==1) )

> expfit( fit1 )

   Exp  Est     ci95.lo    ci95.hi    z value   Pr(>|z|)
(Intercept)  4.9621  1.6367  15.0444  2.8305  0.0046
hinc        0.9484  0.9273  0.9699 -4.6228  0.0000
psize       1.1075  0.6973  1.7589  0.4324  0.6654
```
### Example - Travel Choices (Greene, 1995)

- **Interpretations:**

```r
# Plot Pearson residuals vs. fitted values
nhat <- fitted( fit )
plot( nhat, presid^2, xlab="Fitted mean response", ylab="Squared Pearson residuals" )
abline( h=1, col="red", lwd=2 )
sfit <- loess( presid^2 ~ nhat )
lines( sort(sfit$x), sfit$fitted[order(sfit$x)], col="blue", lwd=2 )
abline(h=phihat, lty=2, col="red", lwd=2)
```

- Again, it looks as though the smoother is consistently above the \( y=1 \) line, indicating overdispersion.

---

**Introduction**

**Multinomial Logistic Regression Model**

**Development**

- Parameter estimation
- Parameter interpretation

**Example - Greene (1995)**

**Example - Travel Choices (Greene, 1995)**

- **Interpretations:**
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

> Now, fit a logistic model for bus vs. air

```r
> #
> ### Model 2: Bus vs. Air
> #
> fit2 <- glm( (travel==2) ~ hinc + psize, data=travel,
>            family=binomial, subset=(travel==0 | travel==2) )

> expfit( fit2 )

exp( Est )  ci95.lo  ci95.hi  z  value  Pr(>|z|)
(Intercept) 4.1831  0.9644  18.1437  1.9116  0.0559
hinc        0.9636  0.9387   0.9891 -2.7792  0.0055
psize      0.5881  0.2920   1.1846 -1.4858  0.1373
```
Accounting for Overdispersion

### Plot Pearson residuals vs. fitted values

```r
nhat <- fitted( fit )
plot( nhat, presid^2, xlab="Fitted mean response", ylab="Squared Pearson residuals" )
abline( h=1, col="red", lwd=2 )
sfit <- loess( presid^2 ~ nhat )
lines( sort(sfit$x), sfit$fitted[order(sfit$x)], col="blue", lwd=2 )
abline(h=phihat, lty=2, col="red", lwd=2)
```

10 20 30 40
0 1 2 3 4 5 6

Fitted mean response
Squared Pearson residuals

Again, it looks as though the smoother is consistently above the y=1 line, indicating overdispersion.

Example - Greene (1995)

### Example - Travel Choices (Greene, 1995)

- **Interpretations:**

General Comments
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

> Now, fit a logistic model for car vs. air

```r
> ##
> ##### Model 3: Car vs. Air
> ##
> fit3 <- glm( (travel==3) ~ hinc + psize, data=travel,
> family=binomial, subset=(travel==0 | travel==3) )

> expfit( fit3 )

exp( Est )  ci95.lo  ci95.hi  z value  Pr(>|z|)
(Intercept) 0.3916  0.1299  1.1811  -1.6644  0.0960
hinc        0.9966  0.9761  1.0175  -0.3210  0.7482
psize       1.8121  1.2208  2.6899   2.9499  0.0032
```
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

- Interpretations:
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

- The separate logistic regression models are inefficient and do not allow us to make comparisons between other modes of transportation

- Let’s fit the multinomial logistic regression model
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

- Need to load in the \texttt{nnet} library first!
- \textbf{Note}: R will automatically choose the lowest outcome category to be the referent group.

```r
# library( nnet )
> mfit <- multinom( travel ~ hinc + psize, data=travel )
# weights: 16 (9 variable)
initial value 291.121816
iter  10 value 254.876765
final value 253.340849
converged
> mfit
Call:
  multinom(formula = travel ~ hinc + psize, data = travel)

Coefficients:
    (Intercept)    hinc    psize
1 1.5504122 -0.060854096  0.2907456
2 1.0344339 -0.033872504 -0.3397481
3 -0.9435028 -0.003542724  0.6005566

Residual Deviance: 506.6817
AIC: 524.6817
```
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

► Not that helpful...How about summary?

```r
> summary( mfit )
Call:
multinom(formula = travel ~ hinc + psize, data = travel)

Coefficients:
   (Intercept)     hinc    psize
1  1.5504122 -0.060854096  0.2907456
2  1.0344339 -0.033872504 -0.3397481
3 -0.9435028 -0.003542724  0.6005566

Std. Errors:
   (Intercept)     hinc    psize
1  0.5197195  0.01184137  0.2257456
2  0.6512401  0.01293842  0.3367476
3  0.5498495  0.01030475  0.1992012

Residual Deviance: 506.6817
AIC: 524.6817
```

Still not a very good summary! What would we like to see...
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

▶ If you want something done, you’ve got to do it yourself...

```r
> ##
> ##### Function to summarize the multinomial fit
> ##
> summ.mfit <- function( model ){
+ s <- summary( model )
+ for( i in 1:length(model$coef) ){
+     cat( "Level ", model$lev[i+1], " vs. Level ", model$lev[1], )
+     coef <- s$coefficients[i,]
+     rrr <- exp( coef )
+     se <- s$standard.errors[i,]
+     zStat <- coef / se
+     pVal <- 2*pnorm( abs(zStat), lower.tail=FALSE )
+     ci95.lo <- exp( coef - qnorm(.975)*se )
+     ci95.hi <- exp( coef + qnorm(.975)*se )
+     rslt <- cbind( rrr, se, zStat, pVal, ci95.lo, ci95.hi )
+     print( round( rslt, 3 ) )
+ }
+ }
```
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

► Now let’s summarize the fitted model

```
> summ.mfit( mfit )
Level 1 vs. Level 0
  rrr  se  zStat   pVal ci95.lo ci95.hi
(Intercept)  4.713  0.520  2.983  0.003  1.702  13.053
hinc         0.941  0.012 -5.139  0.000  0.919  0.963
psize       1.337  0.226  1.288  0.198  0.859  2.082

Level 2 vs. Level 0
  rrr  se  zStat   pVal ci95.lo ci95.hi
(Intercept)  2.814  0.651  1.588  0.112  0.785  10.083
hinc         0.967  0.013 -2.618  0.009  0.942  0.992
psize       0.712  0.337 -1.009  0.313  0.368  1.377

Level 3 vs. Level 0
  rrr  se  zStat   pVal ci95.lo ci95.hi
(Intercept)  0.389  0.550 -1.716  0.086  0.132  1.144
hinc         0.996  0.010 -0.344  0.731  0.977  1.017
psize       1.823  0.199  3.015  0.003  1.234  2.694
```
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

- Interpretation of $\beta_{\text{psize,train}}$: 

- Interpretation of $\beta_{\text{psize,bus}}$: 

- Interpretation of $\beta_{\text{psize,car}}$: 

Example - Greene (1995)

- Question: What is the effect of household income on the RRR comparing car to train?

- From page 8, the effect of a $1k increase in household income on the log-RRR comparing car to train is given by:
Accounting for Overdispersion

> ##
> ##### Plot Pearson residuals vs. fitted values
> ##
> nhat <- fitted( fit )
> plot( nhat, presid^2, xlab="Fitted mean response",
> ylab="Squared Pearson residuals" )
> abline( h=1, col="red", lwd=2 )
> sfit <- loess( presid^2 ~ nhat )
> lines( sort(sfit$x), sfit$fitted[order(sfit$x)], col="blue", lwd=2 )
> abline(h=phihat, lty=2, col="red", lwd=2)

• Again, it looks as though the smoother is consistently above the y=1 line, indicating overdispersion

Introduction
Multinomial Logistic Regression Model Development
Parameter estimation
Parameter interpretation

Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

> ##
> ##### Function to test linear contrasts
> ##
> LinContr.mfit <- function( contr.names, contr.coef, model ){
> + beta.hat <- as.vector( t( summary( model )$coefficients ) )
> + se <- as.vector( t( summary( model )$standard.errors ) )
> + corr <- summary( model )$correlation
> + cov.beta <- diag( se ) %*% corr %*% diag( se )
> + contr.index <- is.element( dimnames( corr )[[1]], contr.names )
> + beta.hat <- beta.hat[ contr.index ]
> + cov.beta <- cov.beta[ contr.index,contr.index ]
> + est <- contr.coef %*% beta.hat
> + rrr.est <- exp( est )
> + se.est <- sqrt( contr.coef %*% cov.beta %*% contr.coef )
> + zStat <- est / se.est
> + pVal <- 2*pnorm( abs(zStat), lower.tail=FALSE )
> + ci95.lo <- exp( est - qnorm(.975)*se.est )
> + ci95.hi <- exp( est + qnorm(.975)*se.est )
> + cat( "Test of H_0: " )
> + for( i in 1:(length( contr.names )-1) ){
> + cat( contr.coef[i], " * ", contr.names[i], "+ ")
> + }
> + cat( contr.coef[i+1], " * ", contr.names[i+1], "= 0 :" )
> + rslt <- data.frame( rrr.est, se.est, zStat, pVal, ci95.lo, ci95.hi )
> + round( rslt, 3 )
> + }
Example - Greene (1995)

- Now let’s consider the estimated effect of household income on the RRR comparing car to train

```r
> LinContr.mfit( contr.names=c("1:hinc", "3:hinc"),
  contr.coef=c(-1,1), mfit )
```

Test of H_0: -1 * 1:hinc + 1 * 3:hinc = 0 :

```
   rrr.est  se.est  zStat  pVal ci95.lo ci95.hi
1   1.059  0.012  4.84   0  1.035  1.084
```

- Interpretation:
Example - Greene (1995)

Because estimation is done via ML, we have all of the usual tools (including likelihood ratio tests) at our disposal.

Example - Travel Choices (Greene, 1995)

Let's test the interaction between hinc and psize with a likelihood ratio test.

```r
> mfit.int <- multinom( travel ~ hinc * psize, data=travel )
> summ.mfit( mfit.int )

Level 1 vs. Level 0

<table>
<thead>
<tr>
<th></th>
<th>rrr</th>
<th>se</th>
<th>zStat</th>
<th>pVal</th>
<th>ci95.lo</th>
<th>ci95.hi</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>21.875</td>
<td>0.930</td>
<td>3.316</td>
<td>0.001</td>
<td>3.531</td>
<td>135.505</td>
</tr>
<tr>
<td>hinc</td>
<td>0.905</td>
<td>0.026</td>
<td>-3.781</td>
<td>0.000</td>
<td>0.859</td>
<td>0.953</td>
</tr>
<tr>
<td>psize</td>
<td>0.525</td>
<td>0.475</td>
<td>-1.356</td>
<td>0.175</td>
<td>0.207</td>
<td>1.333</td>
</tr>
<tr>
<td>hinc:psize</td>
<td>1.024</td>
<td>0.014</td>
<td>1.748</td>
<td>0.081</td>
<td>0.997</td>
<td>1.052</td>
</tr>
</tbody>
</table>

Level 2 vs. Level 0

<table>
<thead>
<tr>
<th></th>
<th>rrr</th>
<th>se</th>
<th>zStat</th>
<th>pVal</th>
<th>ci95.lo</th>
<th>ci95.hi</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>13.862</td>
<td>1.200</td>
<td>2.191</td>
<td>0.028</td>
<td>1.319</td>
<td>145.683</td>
</tr>
<tr>
<td>hinc</td>
<td>0.928</td>
<td>0.032</td>
<td>-2.342</td>
<td>0.019</td>
<td>0.872</td>
<td>0.988</td>
</tr>
<tr>
<td>psize</td>
<td>0.263</td>
<td>0.757</td>
<td>-1.762</td>
<td>0.078</td>
<td>0.060</td>
<td>1.162</td>
</tr>
<tr>
<td>hinc:psize</td>
<td>1.026</td>
<td>0.020</td>
<td>1.273</td>
<td>0.203</td>
<td>0.986</td>
<td>1.066</td>
</tr>
</tbody>
</table>

Level 3 vs. Level 0

<table>
<thead>
<tr>
<th></th>
<th>rrr</th>
<th>se</th>
<th>zStat</th>
<th>pVal</th>
<th>ci95.lo</th>
<th>ci95.hi</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.019</td>
<td>1.064</td>
<td>1.832</td>
<td>0.067</td>
<td>0.873</td>
<td>56.449</td>
</tr>
<tr>
<td>hinc</td>
<td>0.926</td>
<td>0.026</td>
<td>-2.942</td>
<td>0.003</td>
<td>0.880</td>
<td>0.975</td>
</tr>
<tr>
<td>psize</td>
<td>0.332</td>
<td>0.574</td>
<td>-1.924</td>
<td>0.054</td>
<td>0.108</td>
<td>1.021</td>
</tr>
<tr>
<td>hinc:psize</td>
<td>1.043</td>
<td>0.014</td>
<td>3.016</td>
<td>0.003</td>
<td>1.015</td>
<td>1.072</td>
</tr>
</tbody>
</table>
```
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

- Note that a test of the interaction is a test of the following hypothesis:
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

Let's carry out the LRT using the `anova` command:

```r
> anova( mfit, mfit.int )
Likelihood ratio tests of Multinomial Models
Response: travel

Model Resid. df Resid. Dev Test Df LR stat. Pr(Chi)
1 hinc + psize 621 506.6817
2 hinc * psize 618 495.6589 1 vs 2 3 11.02283 0.01160309
```

Conclusion:
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

▶ Question: Based upon the interaction model, what is the effect of party size on the RRR comparing car to air among those making $20k per year? Among those making $50k per year?

```r
> ##
> ##### Compare the effect of party size for incomes of 20k and 50k
> ##
> LinContr.mfit( contr.names=c("3:psize", "3:hinc:psize"),
contr.coef=c(1,20), mfit.int )
Test of H_0: 1 * 3:psize + 20 * 3:hinc:psize = 0 :

rrr.est se.est zStat pVal ci95.lo ci95.hi
1 0.772 0.334 -0.774 0.439 0.402 1.486

> LinContr.mfit( contr.names=c("3:psize", "3:hinc:psize"),
contr.coef=c(1,50), mfit.int )
Test of H_0: 1 * 3:psize + 50 * 3:hinc:psize = 0 :

rrr.est se.est zStat pVal ci95.lo ci95.hi
1 2.743 0.282 3.583 0.0 1.579 4.764
```
## Plot Pearson residuals vs. fitted values

```r
nhat <- fitted( fit )
plot( nhat, presid^2, xlab="Fitted mean response", ylab="Squared Pearson residuals")
abline( h=1, col="red", lwd=2 )
sfit <- loess( presid^2 ~ nhat )
lines( sort(sfit$x), sfit$fitted[order(sfit$x)], col="blue", lwd=2 )
abline(h=phihat, lty=2, col="red", lwd=2)
```

### Fitted mean response vs. Squared Pearson residuals

- Again, it looks as though the smoother is consistently above the **y=1** line, indicating overdispersion.
Accounting for Overdispersion

### Plot Pearson residuals vs. fitted values

```r
nhat <- fitted( fit )
plot( nhat, presid^2, xlab="Fitted mean response",
     ylab="Squared Pearson residuals"
)
abline( h=1, col="red", lwd=2 )
sfit <- loess( presid^2 ~ nhat )
lines( sort(sfit$x), sfit$fitted[order(sfit$x)], col="blue", lwd=2 )
abline(h=phihat, lty=2, col="red", lwd=2)
```

10 20 30 40
0 1 2 3 4 5 6

**Fitted mean response**

**Squared Pearson residuals**

• Again, it looks as though the smoother is consistently above the y=1 line, indicating overdispersion

### Example - Greene (1995)

**Example - Travel Choices (Greene, 1995)**

▶ **Question**: Based upon the interaction model, what is the effect of income on the RRR comparing car to air among those with a party size of 1? Among those with a party size of 3?

```r
> ##
> ##### Compare the effect of income for party sizes of 1 and 3
> ##
> LinContr.mfit( contr.names=c("3:hinc", "3:hinc:psize"),
> contr.coef=c(1,1), mfit.int )

Test of H_0: 1 * 3:hinc + 1 * 3:hinc:psize = 0 :

<table>
<thead>
<tr>
<th>rrr.est</th>
<th>se.est</th>
<th>zStat</th>
<th>pVal 195.lo</th>
<th>ci95.hi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.966</td>
<td>-2.36</td>
<td>0.018</td>
<td>0.939</td>
</tr>
</tbody>
</table>

> LinContr.mfit( contr.names=c("3:hinc", "3:hinc:psize"),
> contr.coef=c(1,3), mfit.int )

Test of H_0: 1 * 3:hinc + 3 * 3:hinc:psize = 0 :

<table>
<thead>
<tr>
<th>rrr.est</th>
<th>se.est</th>
<th>zStat</th>
<th>pVal 195.lo</th>
<th>ci95.hi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.051</td>
<td>2.342</td>
<td>0.019</td>
<td>1.008</td>
</tr>
</tbody>
</table>
Example - Greene (1995)

Example - Travel Choices (Greene, 1995)

- Interpretations:
General Comments

- Multinomial logistic regression for a response with \( J \) categories is analogous to *simultaneously* fitting \( J - 1 \) logistic regressions with the same reference group.

- Although the choice of reference is arbitrary (models with different reference groups are simply reparameterizations of one another), it is usually easiest to choose the most common outcome as the referent.

- In general, simultaneous estimation will lead to a more efficient model than the separate logistic regressions.

- Simultaneous estimations also easily allows one to compare other outcome categories without refitting the model.