STAT 226 Lecture 20
Section 6.1 Baseline-Category Logit Models

Yibi Huang
Response $Y$ has $J > 2$ categories.

Extensions of logistic regression for nominal and ordinal $Y$ assumes a multinomial distribution for $Y$.

- 6.1 Baseline-Categorical Logit Models for Nominal Responses
- 6.2 Cumulative Logit Models for Ordinal Responses
Review of Multinomial Distributions

If $n$ trials are performed:

- in each trial there are $J > 2$ possible outcomes (categories)
- $\pi_j = P(\text{category } j)$, for each trial, $\sum_{j=1}^{J} \pi_j = 1$
- trials are independent
- $Y_j = \text{number of trials fall in category } j \text{ out of } n \text{ trials}$

then the joint distribution of $(Y_1, Y_2, \ldots, Y_J)$ is said to have a **multinomial distribution**, with $n$ trials and category probabilities $(\pi_1, \pi_2, \ldots, \pi_J)$, denoted as

$$(Y_1, Y_2, \ldots, Y_J) \sim \text{Multinom}(n; \pi_1, \pi_2, \ldots, \pi_J),$$

with probability function

$$P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_J = y_J) = \frac{n!}{y_1! y_2! \cdots y_J!} \pi_1^{y_1} \pi_2^{y_2} \cdots \pi_J^{y_J}$$

where $0 \leq y_j \leq n$ for all $j$ and $\sum_j y_j = n$.  

For a binary response variable, there is only one kind of odds that we may consider

\[ \frac{\pi}{1 - \pi}. \]

For a multi-category response variable with \( J > 2 \) categories and category probabilities \( (\pi_1, \pi_2, \ldots, \pi_J) \), we may consider various kinds of odds, though some of them are more meaningful than others.

- odds between two categories: \( \pi_i/\pi_j \).
- odds between a group of categories vs another group of categories, e.g.,

\[ \frac{\pi_1 + \pi_3}{\pi_2 + \pi_4 + \pi_5}. \]

Note the two groups of categories should be non-overlapping.
E.g., if $Y =$ source of meat (in a broad sense) with 5 categories 
beef, pork, chicken, turkey, fish

We may consider the odds of

- beef vs. chicken: $\pi_{\text{beef}} / \pi_{\text{chicken}}$
- red meat vs. white meat:
  \[
  \frac{\pi_{\text{beef}} + \pi_{\text{pork}}}{\pi_{\text{chicken}} + \pi_{\text{turkey}} + \pi_{\text{fish}}}
  \]
- red meat vs. poultry:
  \[
  \frac{\pi_{\text{beef}} + \pi_{\text{pork}}}{\pi_{\text{chicken}} + \pi_{\text{turkey}}}
  \]
Odds for Ordinal Variables

If $Y$ is ordinal with ordered categories:

$$1 < 2 < \ldots < J$$

we may consider the odds of $Y \leq j$

$$\frac{P(Y \leq j)}{P(Y > j)} = \frac{\pi_1 + \pi_2 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}$$

e.g., $Y =$ political ideology, with 5 levels

very liberal < slightly liberal < moderate

< slightly conservative < very conservative

we may consider the odds

$$\frac{P(\text{very or slightly liberal})}{P(\text{moderate or conservative})} = \frac{\pi_{\text{vlib}} + \pi_{\text{slib}}}{\pi_{\text{mod}} + \pi_{\text{scon}} + \pi_{\text{vcon}}}$$
For any sensible odds between two (groups of) categories of $Y$ can be compared across two levels of $X$.

E.g., if $Y =$ source of meat, $X =$ ethnicity of host (Italian, Chinese) we may consider

**OR between $Y$ (beef vs. chicken) and $X =$ IT or CH**

$$OR = \frac{P(Y = \text{beef}|X = \text{IT})/P(Y = \text{chicken}|X = \text{IT})}{P(Y = \text{beef}|X = \text{CH})/P(Y = \text{chicken}|X = \text{CH})}$$

**OR between $Y$ (red meat vs. poultry) and $X =$ IT or CH**

$$OR = \frac{\text{odds of red meat vs. poultry when } X = \text{IT}}{\text{odds of red meat vs. poultry when } X = \text{CH}}$$

$$= \frac{P(Y = \text{beef or pork}|X = \text{IT})/P(Y = \text{chicken or turkey}|X = \text{IT})}{P(Y = \text{beef or pork}|X = \text{CH})/P(Y = \text{chicken or turkey}|X = \text{CH})}$$

• Again, ORs can be estimated from both prospective and retrospective studies.
6.1 Baseline-Category Logit Models for Nominal Responses
Let $\pi_j = P(Y = j)$, $j = 1, 2, \ldots, J$.

Baseline-category logits are

$$\log \left( \frac{\pi_j}{\pi_J} \right), \quad j = 1, 2, \ldots, J - 1.$$  

Baseline-category logit model has form

$$\log \left( \frac{\pi_j}{\pi_J} \right) = \alpha_j + \beta_j x, \quad j = 1, 2, \ldots, J - 1.$$  

or equivalently,

$$\pi_j = \pi_J \exp(\alpha_j + \beta_j x) \quad j = 1, 2, \ldots, J - 1.$$  

- Separate set of parameters $(\alpha_j, \beta_j)$ for each logit.
- Equation for $\pi_J$ is not needed since $\log(\pi_J/\pi_J) = 0$.
Choice of Baseline-Category Is Arbitrary

Equation for other pair of categories, say, categories $a$ and $b$ can then be determined as

$$\log\left(\frac{\pi_a}{\pi_b}\right) = \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right) = \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right)$$

$$= (\alpha_a + \beta_a x) - (\alpha_b + \beta_b x)$$

$$= (\alpha_a - \alpha_b) + (\beta_a - \beta_b)x$$
Choice of Baseline-Category Is Arbitrary

Equation for other pair of categories, say, categories $a$ and $b$ can then be determined as

$$\log\left( \frac{\pi_a}{\pi_b} \right) = \log\left( \frac{\pi_a}{\pi_b} \right) = \log\left( \frac{\pi_a}{\pi_J} \right) - \log\left( \frac{\pi_b}{\pi_J} \right)$$

$$= (\alpha_a + \beta_a x) - (\alpha_b + \beta_b x)$$

$$= (\alpha_a - \alpha_b) + (\beta_a - \beta_b) x$$

Any of the categories can be chosen to be the baseline

- The model will fit equally well, achieving the same likelihood and producing the same fitted values.
- The coefficients $\alpha_j$, $\beta_j$’s will change, but their differences

$$\alpha_a - \alpha_b \text{ and } \beta_a - \beta_b$$

between any two categories $a$ and $b$ will stay the same.
Could also use this model with ordinal response variables, but this would ignore ordinal information.
### Example (Job Satisfaction and Income)

Data from General Social Survey (1991)

<table>
<thead>
<tr>
<th>Income (x)</th>
<th>Dissat</th>
<th>Little</th>
<th>Moderate</th>
<th>Very</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5K</td>
<td>2</td>
<td>4</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>5-15K</td>
<td>2</td>
<td>6</td>
<td>22</td>
<td>4</td>
</tr>
<tr>
<td>15-25K</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>&gt;25K</td>
<td>0</td>
<td>3</td>
<td>13</td>
<td>8</td>
</tr>
</tbody>
</table>

**Goal**: to know if one's job satisfaction changes with income

From the table above, there seems to be higher percentages of people in the more satisfied categories in the higher income groups.

How to we test if the tendency is significant?
<table>
<thead>
<tr>
<th>Income ($x$)</th>
<th>Job Satisfaction ($Y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dissat</td>
</tr>
<tr>
<td>0-5K</td>
<td>2</td>
</tr>
<tr>
<td>5-15K</td>
<td>2</td>
</tr>
<tr>
<td>15-25K</td>
<td>0</td>
</tr>
<tr>
<td>&gt;25K</td>
<td>0</td>
</tr>
</tbody>
</table>

Note $X =$ Income is ordinal w/ 4 categories.

To utilize the ordinal info of $X$, instead of creating dummy variables for the categories of $X$ as if $X$ is nominal, we convert the categories to

$$X = \text{income scores (3K, 10K, 20K, 35K)},$$

and fit the baseline-category logit model

$$\log \left( \frac{\pi_j}{\pi_J} \right) = \alpha_j + \beta_j x, \quad j = 1, 2, 3.$$  

for $J = 4$ job satisfaction categories.
ML estimates for coefficients \((\alpha_j, \beta_j)\) in logit model can be found via the R function \texttt{vglm()} in the package \texttt{VGAM} w/ multinomial family.

You will have to install the \texttt{VGAM} library first, by the following command. You only need to install ONCE!

```
install.packages("VGAM")  # JUST RUN THIS ONCE!
```

Once installed, must load \texttt{VGAM} at every R session before it can be used.

```
library(VGAM)
```
Recall we use $X = \text{income score (3K, 10K, 20K, 35K)}$ as the predictor.

\[
\begin{array}{cccc}
\text{Income (X)} & \text{Dissat} & \text{Little} & \text{Moderate} \\
0-5K & 2 & 4 & 13 & 3 \\
5-15K & 2 & 6 & 22 & 4 \\
15-25K & 0 & 1 & 15 & 8 \\
>25K & 0 & 3 & 13 & 8 \\
\end{array}
\]

\[
\text{Income} = c(3,10,20,35) \\
\text{Diss} = c(2,2,0,0) \\
\text{Little} = c(4,6,1,3) \\
\text{Mod} = c(13,22,15,13) \\
\text{Very} = c(3,4,8,8) \\
\text{jobsat.fit1} = \text{vglm(cbind(Diss,Little,Mod,Very) ~ Income, family=multinomial)}
\]
The fitted model is

\[
\begin{align*}
\log \left( \frac{\hat{\pi}_1}{\hat{\pi}_4} \right) &= \hat{\alpha}_1 + \hat{\beta}_1 x = 0.430 - 0.185x \\
\log \left( \frac{\hat{\pi}_2}{\hat{\pi}_4} \right) &= \hat{\alpha}_2 + \hat{\beta}_2 x = 0.456 - 0.054x \\
\log \left( \frac{\hat{\pi}_3}{\hat{\pi}_4} \right) &= \hat{\alpha}_3 + \hat{\beta}_3 x = 1.704 - 0.037x
\end{align*}
\]

As \( \hat{\beta}_j < 0 \) for \( j = 1, 2, 3 \), for each logit, estimated odds of being in less satisfied category (instead of very satisfied) decrease as \( x = \) income increases.
Interpretation of Coefficients

**Interpretation** of $\beta_i$ in the model $\log(\pi_i / \pi_J) = \alpha_i + \beta_i x$:

For every 1-unit increase in $x$, the odds of $Y$ being in category $i$ rather than category $J$ become $e^{\beta_i}$ times as large.

**Example** (Job Satisfaction)

\[
\log(\pi_1 / \pi_4) = 0.430 - 0.185x \quad \text{(Dissat. v.s. Very Sat.)}
\]
\[
\log(\pi_2 / \pi_4) = 0.456 - 0.054x \quad \text{(Little v.s. Very Sat.)}
\]
\[
\log(\pi_3 / \pi_4) = 1.704 - 0.037x \quad \text{(Moderate v.s. Very Sat.)}
\]

Estimated odds of being

- "dissatisfied" $e^{-0.185} \approx 0.83$
- "little satisfied" rather than "very satisfied" become $e^{-0.054} \approx 0.95$
- "moderately satisfied" $e^{-0.037} \approx 0.96$

times as large for each 1K increase in income.
**Example** (Job Satisfaction)

\[
\log(\frac{\widehat{\pi}_1}{\widehat{\pi}_4}) = 0.430 - 0.185x \quad \text{(Dissat. v.s. Very Sat.)}
\]

\[
\log(\frac{\widehat{\pi}_2}{\widehat{\pi}_4}) = 0.456 - 0.054x \quad \text{(Little v.s. Very Sat.)}
\]

\[
\log(\frac{\widehat{\pi}_3}{\widehat{\pi}_4}) = 1.704 - 0.037x \quad \text{(Moderate v.s. Very Sat.)}
\]

The estimated odds of being “little satisfied” rather than “dissatisfied” (neither is the baseline category) become

\[
e^{-0.054-(0.185)} \approx 1.14
\]

times as large for each 1K increase in income.
Baseline-Category Logit Model:

\[ \log \left( \frac{\pi_j}{\pi_J} \right) = \alpha_j + \beta_j x \iff \pi_j = \pi_J e^{\alpha_j+\beta_j x} \quad j = 1, 2, \ldots, J - 1. \]

The probability \( \pi_J \) for the baseline category can be determined from \( \sum_{j=1}^{J} \pi_j = 1 \) as follows:

\[ 1 = \sum_{j=1}^{J} \pi_j = \pi_J + \sum_{j=1}^{J-1} \pi_J e^{\alpha_j+\beta_j x} = \pi_J \left( 1 + \sum_{j=1}^{J-1} e^{\alpha_j+\beta_j x} \right) \]

So \( \pi_J = \frac{1}{1 + \sum_{k=1}^{J-1} e^{\alpha_k+\beta_k x}} \). The probabilities \( \pi_j \) for other categories can be obtained from \( \pi_j = \pi_J e^{\alpha_j+\beta_j x} \) to be

\[ \pi_j = \frac{e^{\alpha_j+\beta_j x}}{1 + \sum_{k=1}^{J-1} e^{\alpha_k+\beta_k x}} \], \ for \ j = 1, 2, \ldots, J - 1 \]
Probabilities of Categories (Job Satisfaction)

\[ \hat{\pi}_1 = \frac{e^{0.430-0.185x}}{1 + e^{0.430-0.185x} + e^{0.456-0.054x} + e^{1.704-0.037x}} \]

\[ \hat{\pi}_2 = \frac{e^{0.456-0.054x}}{1 + e^{0.430-0.185x} + e^{0.456-0.054x} + e^{1.704-0.037x}} \]

\[ \hat{\pi}_3 = \frac{e^{1.704-0.037x}}{1 + e^{0.430-0.185x} + e^{0.456-0.054x} + e^{1.704-0.037x}} \]

\[ \hat{\pi}_4 = \frac{1}{1 + e^{0.430-0.185x} + e^{0.456-0.054x} + e^{1.704-0.037x}} \]

E.g., at \( x = 20 \) (K), estimated prob. of being “dissatisfied” and “very satisfied” are respectively,

\[ \hat{\pi}_1 = \frac{e^{0.430-0.185(20)}}{1 + e^{0.430-0.185(20)} + e^{0.456-0.054(20)} + e^{1.704-0.037(20)}} \approx 0.009 \]

\[ \hat{\pi}_4 = \frac{1}{1 + e^{0.430-0.185(20)} + e^{0.456-0.054(20)} + e^{1.704-0.037(20)}} \approx 0.240 \]
In R, we can obtain the prob. of being “diss”, “little”, “moderate”, or “very satisfied” when income is 20K using `predict()`.

```
predict(jobsat.fit1, data.frame(Income=20), type="response")
```

<table>
<thead>
<tr>
<th></th>
<th>Diss</th>
<th>Little</th>
<th>Mod</th>
<th>Very</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.009043</td>
<td>0.1274</td>
<td>0.6238</td>
<td>0.2397</td>
</tr>
</tbody>
</table>

We see $\hat{\pi}_1 \approx 0.009$, $\hat{\pi}_2 \approx 0.127$, $\hat{\pi}_3 \approx 0.624$, and $\pi_4 \approx 0.240$.

Observe that $\hat{\pi}_1 + \hat{\pi}_2 + \hat{\pi}_3 + \hat{\pi}_4 = 1$.

**Caution**: without specifying `type="response"`, `predict()` would return the values of the logits $\log(\hat{\pi}_i/\hat{\pi}_J) = \alpha_i + \beta_i x$, not the probabilities $\hat{\pi}_i$.

```
predict(jobsat.fit1, data.frame(Income=20))
```

<table>
<thead>
<tr>
<th></th>
<th>log($\mu[:,1]/\mu[:,4]$)</th>
<th>log($\mu[:,2]/\mu[:,4]$)</th>
<th>log($\mu[:,3]/\mu[:,4]$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.278</td>
<td>-0.632</td>
<td>0.9562</td>
</tr>
</tbody>
</table>
Observe that though $\pi_j/\pi_J$ is a monotone function of $x$, $\pi_j$ may NOT be monotone in $x$. 
Deviance and Goodness of Fit

For grouped multinomial response data,

<table>
<thead>
<tr>
<th>conditions of trial (explanatory variables)</th>
<th>number of trials</th>
<th>multinomial counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1 (x_{11}, x_{12}, \ldots, x_{1p})</td>
<td>(n_1)</td>
<td>(y_{11}, y_{12}, \ldots, y_{1J})</td>
</tr>
<tr>
<td>Condition 2 (x_{21}, x_{22}, \ldots, x_{2p})</td>
<td>(n_2)</td>
<td>(y_{21}, y_{22}, \ldots, y_{2J})</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>Condition N (x_{N1}, x_{N2}, \ldots, x_{Np})</td>
<td>(n_N)</td>
<td>(y_{N1}, y_{N2}, \ldots, y_{NJ})</td>
</tr>
</tbody>
</table>

(Residual) Deviance for a Model \(M\) is defined as

\[
\text{Deviance} = -2(L_M - L_S) = 2 \sum_{ij} y_{ij} \log \left( \frac{y_{ij}}{n_i \hat{\pi}_j(x_i)} \right) = 2 \sum_{ij} (\text{observed}) \log \left( \frac{\text{observed}}{\text{fitted}} \right)
\]

where \(\hat{\pi}_j(x_i)\) = estimated prob. based on Model \(M\)

\(L_M = \text{max. log-likelihood for Model } M\)

\(L_S = \text{max. log-likelihood for the saturated model}\)
DF of Deviance

df for deviance of Model $M$ is

$$N(J - 1) - \text{(number of parameters in the model)}.$$ 

where $N = \# \text{ of rows in the data}$, $J = \# \text{ of levels of the response}$

If the model has $p$ explanatory variables,

$$\log \left( \frac{\pi_j}{\pi_J} \right) = \alpha_j + \beta_{1j}x_1 + \cdots + \beta_{pj}x_p, \quad j = 1, 2, \ldots, J - 1.$$ 

there are $p + 1$ coefficients per equation, hence $(J - 1)(p + 1)$ coefficients in total.

$$\text{df for deviance} = N(J - 1) - (J - 1)(p + 1) = (J - 1)(N - p - 1).$$

deviance(jobsat.fit1)
[1] 4.658

df.residual(jobsat.fit1)
[1] 6
Goodness of Fit Test (GOF test)

If the estimated expected counts $n_i \hat{\pi}_j(x_i)$ are large enough ($\geq 5$), the deviance has a large sample chi-squared distribution with $df = df$ of deviance.

We can use deviance to conduct Goodness of Fit test

- $H_0$: Model $M$ is correct (fits the data as well as the saturated model)
- $H_A$: Saturated model is correct

When $H_0$ is rejected, it means that Model $M$ doesn’t fit as well as the saturated model.
Example (Job Satisfaction): the P-value for the GOF test is 58.8%, no evidence of lack of fit. However, this $P$-value is not reliable because most of the cell counts are small.

```r
deviance(jobsat.fit1)
[1] 4.658
df.residual(jobsat.fit1)
[1] 6
pchisq(4.657999, df=6, lower.tail=F)
[1] 0.5884
```
Wald CIs and Wald Tests for Coefficients

- Wald CI for $\beta_j$ is $\hat{\beta}_j \pm z_{\alpha/2} \text{SE}(\hat{\beta}_j)$.
- Wald test of $H_0: \beta_j = 0$ uses $z = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)} \sim N(0, 1)$

Example (Job Satisfaction):

```r
# the 4 to 6th coefficients, the 1st to 3rd are intercepts
corf(summary(jobsat.fit1))[4:6,]
```

|               | Estimate | Std. Error | z value | Pr(>|z|) |
|---------------|----------|------------|---------|----------|
| Income:1      | -0.18537 | 0.10251    | -1.808  | 0.07057  |
| Income:2      | -0.05441 | 0.03112    | -1.748  | 0.08038  |
| Income:3      | -0.03739 | 0.02088    | -1.790  | 0.07340  |

- 95% for $\beta_1$: $-0.185 \pm 1.96 \times 0.1025 \approx (-0.386, 0.016)$
- 95% for $e^{\beta_1}$: $(e^{-0.386}, e^{0.016}) \approx (0.680, 1.016)$

Interpretation: Estimated odds of being “dissatisfied” rather than “very satisfied” become 0.680 to 1.016 times as large for each 1K increase in income w/ 95% confidence.
Likelihood Ratio Tests

**Example** (Job Satisfaction): Overall test of income effect

\[ H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \]

is equivalent of the comparison of the two models

\[ H_0 : \log \left( \frac{\pi_j}{\pi_4} \right) = \alpha_j, \quad j = 1, 2, 3 \]
\[ H_1 : \log \left( \frac{\pi_j}{\pi_4} \right) = \alpha_j + \beta_j x, \quad j = 1, 2, 3. \]

\[ \text{LRT} = -2(L_0 - L_1) = -2(-21.358 - (-16.954)) = 8.808 \]
\[ = \text{diff in deviances} = 13.467 - 4.658 = 8.809 \]
\[ Df = \text{diff. in number of parameters} = 6 - 3 = 3 \]
\[ = \text{diff. in residual df} = 9 - 6 = 3 \]
\[ \text{pchisq}(8.809, df=3, \text{lower.tail}=\text{FALSE}) \]
\[ P\text{-value} = P(\chi^2_3 > 8.809) \approx 0.03194. \]
Need to load the VGAM library.

```
lrtest(jobsat.fit1)
Likelihood ratio test
lrtest(jobsat.fit1, "Income")
```

Model 1: `cbind(Diss, Little, Mod, Very) ~ Income`
Model 2: `cbind(Diss, Little, Mod, Very) ~ 1`

```
#   Df LogLik Df Chisq Pr(>Chisq)
1 6  -16.9
2 9  -21.4 3  8.81  0.032
```

```
jobsat.fit2 = vglm(cbind(Diss,Little,Mod,Very) ~ 1, 
       family=multinomial)
```

```
logLik(jobsat.fit2)
[1]  -21.36
logLik(jobsat.fit1)
[1]  -16.95
```

Note that $H_0$ implies job satisfaction is independent of income. We got some evidence ($P$-value = 0.032) of dependence between job satisfaction and income.
Note we get a different conclusion if we conduct Pearson’s Chi-square test of independence:

\[ X^2 = 11.5, \quad df = (4 - 1)(4 - 1) = 9, \quad P\text{-value} = 0.2415 \]

```
jobsat = matrix(c(2,2,0,0,4,6,1,3,13,22,15,13,3,4,8,8), nrow=4)
chisq.test(jobsat)
```

Warning in chisq.test(jobsat): Chi-squared approximation may be incorrect

Pearson’s Chi-squared test

```
data:  jobsat
X-squared = 12, df = 9, p-value = 0.2
```

LR test of independence gives similar conclusion \((G^2 = 13.47, \quad df = 9, \quad P\text{-value} = 0.1426)\)

Why the Baseline Category Logit model give different conclusion from Pearson’s test of independence?