Job satisfaction:

<table>
<thead>
<tr>
<th>Income (1000s)</th>
<th>Very dissatisfied</th>
<th>Little dissatisfied</th>
<th>Moderately satisfied</th>
<th>Very satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 5</td>
<td>2</td>
<td>4</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>5 – 15</td>
<td>2</td>
<td>6</td>
<td>22</td>
<td>4</td>
</tr>
<tr>
<td>15 – 25</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>&gt; 25</td>
<td>0</td>
<td>3</td>
<td>13</td>
<td>8</td>
</tr>
</tbody>
</table>

Use income scores $x = 3, 10, 20, 30$ and let $\pi_j$ denote the probability of job satisfaction in category $j$.

**Multinomial Baseline-Logit**

The model has form

$$\log \left( \frac{\pi_j}{\pi_4} \right) = \alpha_j + \beta_j x \quad j = 1, 2, 3$$

```r
> library(VGAM)
> income=c(3,10,20,30)
> VD=c(2,2,0,0)
> LD=c(4,6,1,3)
> MS=c(13,22,15,13)
> VS=c(3,4,8,8)
> dat=data.frame(income,VD,LD,MS,VS)
> fit.blogit=vglm(cbind(VD,LD,MS,VS)~income,family=multinomial,data=dat)
> summary(fit.blogit)
```

Call:

`vglm(formula = cbind(VD, LD, MS, VS) ~ income, family = multinomial, data = dat)`

Pearson Residuals:

<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>-0.37286</td>
<td>-0.034836</td>
<td>-0.2220866</td>
</tr>
<tr>
<td>2</td>
<td>0.71461</td>
<td>0.514133</td>
<td>0.3144595</td>
</tr>
<tr>
<td>3</td>
<td>-0.54303</td>
<td>-1.284401</td>
<td>-0.1783847</td>
</tr>
<tr>
<td>4</td>
<td>-0.18431</td>
<td>0.747081</td>
<td>0.0082767</td>
</tr>
</tbody>
</table>

Coefficients:

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept):1</td>
<td>0.563824</td>
<td>0.960138</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>0.645091</td>
<td>0.668771</td>
</tr>
<tr>
<td>(Intercept):3</td>
<td>1.818698</td>
<td>0.528955</td>
</tr>
<tr>
<td>income:1</td>
<td>-0.198773</td>
<td>0.102096</td>
</tr>
<tr>
<td>income:2</td>
<td>-0.070502</td>
<td>0.036954</td>
</tr>
<tr>
<td>income:3</td>
<td>-0.046918</td>
<td>0.025519</td>
</tr>
</tbody>
</table>

Number of linear predictors: 3

Names of linear predictors:
log(mu[,1]/mu[,4]), log(mu[,2]/mu[,4]), log(mu[,3]/mu[,4])

Dispersion Parameter for multinomial family: 1
Residual deviance: 4.17662 on 6 degrees of freedom
Log-likelihood: -16.71316 on 6 degrees of freedom
Number of iterations: 5

The prediction equations are

\[
\log\left(\frac{\hat{\pi}_1}{\hat{\pi}_4}\right) = 0.564 - 0.199x
\]

\[
\log\left(\frac{\hat{\pi}_2}{\hat{\pi}_4}\right) = 0.645 - 0.071x
\]

\[
\log\left(\frac{\hat{\pi}_3}{\hat{\pi}_4}\right) = 1.819 - 0.047x
\]

For each logit, the odds of being in a less satisfied category (instead of “very satisfied”) decreases as income increases.

ML estimates determine the effects for all pairs of categories. For comparing group 1 and 2

\[
\log\left(\frac{\hat{\pi}_1}{\hat{\pi}_2}\right) = \log\left(\frac{\hat{\pi}_1}{\hat{\pi}_4}\right) - \log\left(\frac{\hat{\pi}_2}{\hat{\pi}_4}\right) = -0.081 - 0.128x
\]

In addition,

\[
\hat{\pi}_4(x) = \frac{1}{1 + \exp(0.564 - 0.199x) + \exp(0.645 - 0.071x) + \exp(1.819 - 0.047x)}
\]

and \(\hat{\pi}_4(30) = 0.36\). Hence, the 24 subjects at \(x = 30\) have an estimated expected frequency of 24(0.36) = 8.75 in category 4. The observed frequency is 8. This leads us to the goodness of fit test.

Test of fit gives a deviance of 4.17662 and Pearson \(X^2\) of

\[> \text{sum(resid(fit.blogit,type="pearson")}^2)\]

\[[1] 3.632002\]

on 6 degrees of freedom (12 logits minus 6 parameters). The p-values of 0.6528 and 0.7263 respectively indicate a that the model is a good fit.

**Remark:** If data is entered in ungrouped format, i.e. row for each observation, the deviance would be 205 on 306 df (as the saturated model is different) and could not perform a goodness of fit test.

A global test of income effect is \(H_0 : \beta_1 = \beta_2 = \beta_3 = 0\). The deviance is 4.17662

\[> \text{vglm(cbind(VD,LD,MS,VS)} ~ 1, \text{family=multinomial, data=dat})\]

Call:
\[\text{vglm(formula = cbind(\text{VD, LD, MS, VS})} ~ 1, \text{family = multinomial, data = dat})\]
Coefficients:
(Intercept):1  -1.7491999
(Intercept):2  -0.4964369
(Intercept):3   1.0076405

Degrees of Freedom: 12 Total; 9 Residual
Residual deviance: 13.4673
Log-likelihood: -21.35851

and the null is 13.4673, so the LR statistic is 9.29068 on 3 df and a corresponding p-value of 0.0257 (Score and Wald statistics included in SAS handout).

Remark: If income was treated as qualitative, i.e. categorical, indicated by 3 dummy variables, this would lead to a saturated model with 12 logits and 12 parameters.

\[
\log \left( \frac{\pi_j}{\pi_d} \right) = \alpha_j + \beta_{1j} x_1 + \beta_{2j} x_2 + \beta_{3j} x_3 \quad j = 1, 2, 3
\]

Multinomial Cumulative-Logit
Parallel lines
The model has form
\[
\logit [P(Y \leq j | x)] = \alpha_j + \beta x \quad j = 1, 2, 3
\]

> fit.clogit1=vglm(cbind(VD,LD,MS,VS)~income,family=cumulative(parallel=TRUE),data=dat)
> summary(fit.clogit1)

Call:
vglm(formula = cbind(VD, LD, MS, VS) ~ income, family = cumulative(parallel = TRUE),
data = dat)

Pearson Residuals:

<table>
<thead>
<tr>
<th></th>
<th>logit(P[Y&lt;=1])</th>
<th>logit(P[Y&lt;=2])</th>
<th>logit(P[Y&lt;=3])</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.50515</td>
<td>-0.17146</td>
<td>-0.32616</td>
</tr>
<tr>
<td>2</td>
<td>0.30838</td>
<td>0.26279</td>
<td>0.65528</td>
</tr>
<tr>
<td>3</td>
<td>-0.64867</td>
<td>-1.06902</td>
<td>-0.74889</td>
</tr>
<tr>
<td>4</td>
<td>-0.77837</td>
<td>1.12906</td>
<td>0.29597</td>
</tr>
</tbody>
</table>

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept):1</td>
<td>-2.473156</td>
<td>0.568376</td>
<td>-4.3513</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>-0.781728</td>
<td>0.373724</td>
<td>-2.0917</td>
</tr>
<tr>
<td>(Intercept):3</td>
<td>2.211091</td>
<td>0.445123</td>
<td>4.9674</td>
</tr>
<tr>
<td>income</td>
<td>-0.056347</td>
<td>0.020871</td>
<td>-2.6998</td>
</tr>
</tbody>
</table>

Number of linear predictors: 3

Names of linear predictors: logit(P[Y<=1]), logit(P[Y<=2]), logit(P[Y<=3])

Dispersion Parameter for cumulative family: 1

Residual deviance: 5.9527 on 8 degrees of freedom
The fitted model is
\[
\logit \left( \hat{P}(Y \leq j|x) \right) = \hat{\alpha}_j - 0.056x \quad j = 1, 2, 3.
\]
Hence the odds of response at low end of job satisfaction scale decrease as \(x\) increases, i.e. \(\exp(-0.056) = 0.95\). Estimated odds of job satisfaction below any given level (instead of above it) multiply by 0.95 for a 1-unit increase in \(x\) (1-unit=$1000). For a $10,000 increase in income, the estimated odds multiply by \(\exp(10(-0.056)) = 0.57\). (If we were to reverse the order of the responses, then \(\hat{\beta} = +0.056\).

Odds ratio is the same between same two categories of \(x\) irrespective of cutoff region for response categories (to make response binary) as shown in the diagrams in the class notes.

A goodness of fit test yields a p-value of
\[
1 - \text{pchisq}(\text{deviance(.fit.clogit1)}, \text{df.residual(.fit.clogit1)})
\]
\[0.6525305\]
so we conclude that the model is a good fit.

A test of \(H_0: \text{job satisfaction independent of income (i.e. } \beta = 0 \text{ in cumulative logit model)}\) yields
- A Wald z-stat of -2.6998 (or \(\chi^2\) of 7.17) and a p-value of 0.007.
- A LR statistic of 13.4673 – 5.9527 = 7.5146 on 1 df and a p-value of 0.006.

Hence there is strong evidence of association.
Nonparallel lines

A model with nonparallel lines, i.e. different $\beta_j$ for $j = 1, 2, 3$ instead of one common slope, if fit but it does not differ from the parallel lines model (conclusion via LR test stat).

```r
> fit.clogit2=vglm(cbind(VD,LD,MS,VS)~income,family=cumulative(parallel=FALSE),data=dat)
> summary(fit.clogit2)

Call:
vglm(formula = cbind(VD, LD, MS, VS) ~ income, family = cumulative(parallel = FALSE),
data = dat)

Pearson Residuals:

<table>
<thead>
<tr>
<th></th>
<th>logit(P[Y&lt;=1])</th>
<th>logit(P[Y&lt;=2])</th>
<th>logit(P[Y&lt;=3])</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.19039</td>
<td>0.11922</td>
<td>-0.31038</td>
</tr>
<tr>
<td>2</td>
<td>0.52559</td>
<td>0.24274</td>
<td>0.65217</td>
</tr>
<tr>
<td>3</td>
<td>-0.41013</td>
<td>-1.16777</td>
<td>-0.75413</td>
</tr>
<tr>
<td>4</td>
<td>-0.25546</td>
<td>0.81644</td>
<td>0.28282</td>
</tr>
</tbody>
</table>

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept):1</td>
<td>-1.74105</td>
<td>0.816828</td>
<td>-2.1315</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>-0.82432</td>
<td>0.449753</td>
<td>-1.8328</td>
</tr>
<tr>
<td>(Intercept):3</td>
<td>2.20524</td>
<td>0.515114</td>
<td>4.2811</td>
</tr>
<tr>
<td>income:1</td>
<td>-0.14443</td>
<td>0.091070</td>
<td>-1.5860</td>
</tr>
<tr>
<td>income:2</td>
<td>-0.05356</td>
<td>0.029750</td>
<td>-1.8003</td>
</tr>
<tr>
<td>income:3</td>
<td>-0.05603</td>
<td>0.024771</td>
<td>-2.2619</td>
</tr>
</tbody>
</table>

Number of linear predictors: 3

Names of linear predictors: logit(P[Y<=1]), logit(P[Y<=2]), logit(P[Y<=3])

Dispersion Parameter for cumulative family: 1

Residual deviance: 4.37717 on 6 degrees of freedom

Log-likelihood: -16.81344 on 6 degrees of freedom

Number of iterations: 5