Coursework 7: STAT 570

Philip Pham

November 25, 2018

1. Create a binary variable Z_i, with $Z_i = 0$ corresponding to $Y_i \in \{0, 1\}$ and $Z_i = 1$ corresponding to $Y_i \in \{2, 3\}$. Let $q(x_i) = \mathbb{P}(Z_i = 1 \mid x_i)$, with $\mathbf{x}_i = \begin{pmatrix} 1 & x_{1i} & x_{2i} \end{pmatrix}^\intercal$, represent the probability of mental impairment being *Moderate* or *Impaired*, given covariates $\mathbf{x}_i$, $i = 1, \ldots, n = 40$. Provide a single plot that shows the association between $q(x_i)$ and x_{1i} and x_{2i}, on a response scale you feel is appropriate. Comment on the plot.

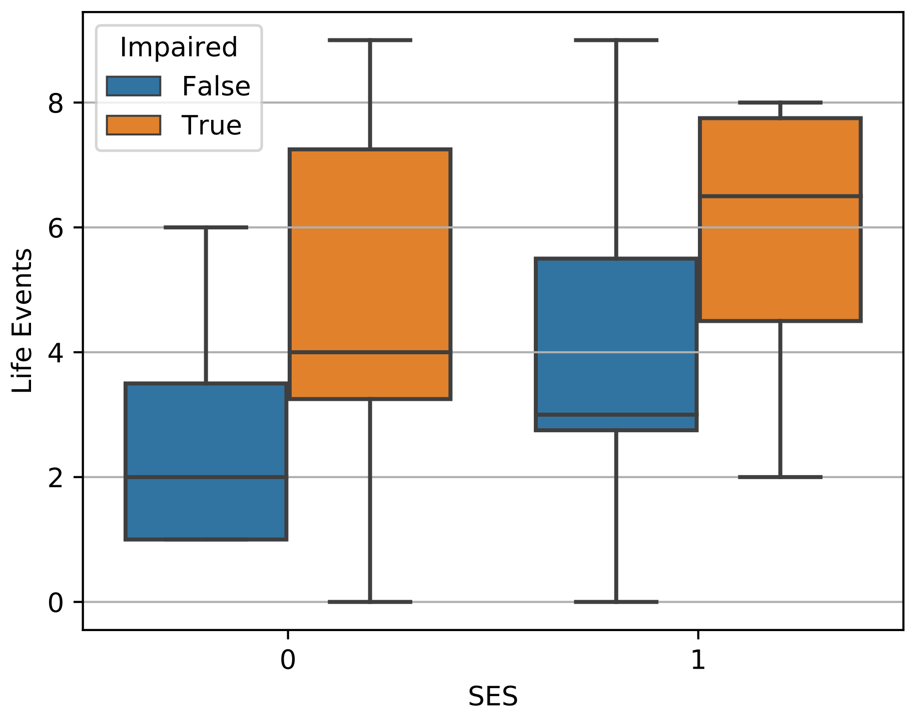


Figure 1: Orange denotes $Z_i = 1$ and blue denotes $Z_i = 0$.

Solution: See Figure 1. Conditioned on SES, those that are impaired ($Z_i = 1$) have a greater number of life events on average.

2. Suppose $Z_i \mid q_i \sim \text{Binomial}(1, q_i)$ independently for $i = 1, \ldots, n = 40$, where $q_i = q(x_i)$. Consider the logistic regression model,

$$q(x_i) = \log\left(\frac{q(\mathbf{x}_i)}{1 - q(\mathbf{x}_i)}\right) = \mathbf{x}_i^\intercal \boldsymbol{\gamma} = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i}, \tag{1}$$

where $\boldsymbol{\gamma} = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 \end{pmatrix}^\intercal$. Write down the log-likelihood $l(\boldsymbol{\gamma})$ for the sample z_i, $i = 1, \ldots, n$.

Solution: Solving for $q(\mathbf{x}_i)$ in Equation 1, we find

$$q(\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i^\intercal \boldsymbol{\gamma})}{1 + \exp(\mathbf{x}_i^\intercal \boldsymbol{\gamma})} = \frac{1}{1 + \exp(-\mathbf{x}_i^\intercal \boldsymbol{\gamma})}. \tag{2}$$

The likelihood function is $L(\boldsymbol{\gamma}) = \prod_{i=1}^n (q(\mathbf{x}_i))^{z_i} (1 - q(\mathbf{x}_i))^{1-z_i}$, so the log-likelihood function becomes

$$\begin{aligned} l(\boldsymbol{\gamma}) = \log L(\boldsymbol{\gamma}) &= \sum_{i=1}^n (z_i \log q(\mathbf{x}_i) + (1 - z_i) \log (1 - q(\mathbf{x}_i))) \qquad (3) \\ &= \sum_{i=1}^n \left(z_i \log \frac{q(\mathbf{x}_i)}{1 - q(\mathbf{x}_i)} + \log (1 - q(\mathbf{x}_i)) \right) \\ &= \sum_{i=1}^n \left(z_i \mathbf{x}_i^\intercal \boldsymbol{\gamma} + \log \frac{1}{1 + \exp(\mathbf{x}_i^\intercal \boldsymbol{\gamma})} \right) = \sum_{i=1}^n - \log (1 + \exp((1 - 2z_i) \mathbf{x}_i^\intercal \boldsymbol{\gamma})). \end{aligned}$$

3. Fit the model described in the previous part, and give confidence intervals for the odds ratios.

   Carefully interpret these odds ratios.

	Estimate	Standard error	95% CI lower bound	95% CI upper bound
γ_0	-0.925065	0.723346	-2.342797	0.492666
γ_1	-1.629731	0.780849	-3.160167	-0.099296
γ_2	0.309899	0.147920	0.019980	0.599818

Table 1: Estimates and confidence intervals for $\hat{\boldsymbol{\gamma}}$ using maximum likelihood estimation.

Solution: Taking the derivative of Equation 3, we have the score function:

$$\begin{aligned} S(\boldsymbol{\gamma}) = \nabla^\intercal l(\boldsymbol{\gamma}) &= \sum_{i=1}^n \frac{2z_i - 1}{1 + \exp((1 - 2z_i) \mathbf{x}_i^\intercal \boldsymbol{\gamma})} \exp((1 - 2z_i) \mathbf{x}_i^\intercal \boldsymbol{\gamma}) \mathbf{x}_i. \\ &= \sum_{i=1}^n \frac{2z_i - 1}{1 + \exp((2z_i - 1) \mathbf{x}_i^\intercal \boldsymbol{\gamma})} \mathbf{x}_i = X^\intercal (\mathbf{z} - \mathbf{q}(X)), \qquad (4) \end{aligned}$$

where $\mathbf{z} = \begin{pmatrix} z_1 & z_2 & \cdots & z_n \end{pmatrix}^\intercal$ and $\mathbf{q}(X) = \begin{pmatrix} q_1 & q_2 & \cdots & q_n \end{pmatrix}^\intercal$.
From Equation 4, we have the Fisher information matrix:

$$\begin{aligned} I_n(\boldsymbol{\gamma}) &= \operatorname{var}(S(\boldsymbol{\gamma}) \mid \boldsymbol{\gamma}) = \mathbb{E}[S(\boldsymbol{\gamma}) S(\boldsymbol{\gamma})^\intercal \mid \boldsymbol{\gamma}] \\ &= \mathbb{E}[X^\intercal (\mathbf{z} - \mathbf{q}(X)) (\mathbf{z} - \mathbf{q}(X))^\intercal X \mid \boldsymbol{\gamma}] \\ &= X^\intercal \mathbb{E}[(\mathbf{z} - \mathbf{q}(X)) (\mathbf{z} - \mathbf{q}(X))^\intercal \mid \boldsymbol{\gamma}] X \\ &= \sum_{i=1}^n q(\mathbf{x}_i) (1 - q(\mathbf{x}_i)) \mathbf{x}_i \mathbf{x}_i^\intercal = \sum_{i=1}^n \frac{1}{2 + \exp(-\mathbf{x}_i^\intercal \boldsymbol{\gamma}) + \exp(\mathbf{x}_i^\intercal \boldsymbol{\gamma})} \mathbf{x}_i \mathbf{x}_i^\intercal, \qquad (5) \end{aligned}$$

where we have used independence of the observations and variance of the binomial distribution to get the last line.

We solve Equation 4, $S(\hat{\boldsymbol{\gamma}}) = \mathbf{0}$, to get an estimate for $\boldsymbol{\gamma}$. Using Equation 5, we have that

$$\hat{\boldsymbol{\gamma}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\gamma, I_n^{-1}(\hat{\gamma})\right), \tag{6}$$

that is, $\hat{\boldsymbol{\gamma}}$ is asymptotically normal.

Using Equation 6, we obtain the estimates and intervals in Table 1.

The predicted log odds ratio given some $\mathbf{x}_i$ is

$$\hat{\theta}_i = \mathbf{x}_i^\intercal \hat{\boldsymbol{\gamma}}, \tag{7}$$

which will have variance

$$\operatorname{var}\left(\hat{\theta}_i\right) = \mathbf{x}_i^\intercal \operatorname{var}(\hat{\boldsymbol{\gamma}})\, \mathbf{x}_i \approx \mathbf{x}_i^\intercal I_n^{-1}(\hat{\boldsymbol{\gamma}})\, \mathbf{x}_i, \tag{8}$$

using Equation 6.

From Equation 8, we can compute confidence intervals for the log odds ratio and exponentiate to get confidence intervals for the odds ratio since log is a monotonic transformation. Doing so results in the estimates in Table 2.

The odds ratio is how much more likely one is to have *Moderate* or *Impaired* mental impairment. Exponentiating Equation 7, we have

$$\exp(\theta_i) = \exp(\gamma_0)\exp(\gamma_1 x_{1i})\exp(\gamma_2 x_{2i}). \tag{9}$$

$\exp(\gamma_0)$ is the expected odds ratio for a subject with 0 SES and no life events. $\exp(\gamma_1)$ is the expected odds ratio between a subject with SES 1 and SES 0. $\exp(\hat{\gamma}_1) < 1$, so SES is associated with less mental impairment. $\exp(\gamma_2)$ is the expected odds ratio for a subject with an additional life event. $\exp(\hat{\gamma}_2) > 1$, so life events are associated with more severe mental impairment.

4. We will now consider analyses that do not coarsen the data. We begin by defining notation in a generic situation. Suppose the random variable, Y_i, for individual i, $i = 1, \ldots, n$, can take values $0, 1, 2, \ldots, J-1$ (so that that there are J levels). Assume that for individual i, the data follow a multinomial distribution, $Y_i \mid p_i \sim \text{Multinomial}(1, \mathbf{p}_i)$ independently, where $p_i = \begin{pmatrix} p_{i0} & \cdots & p_{i,J-1} \end{pmatrix}^\intercal$, and p_{ij} represents the probability

$$p_{ij} = \mathbb{P}(Y_i = j \mid \mathbf{x}_i), \text{ for } j = 0, 1, \ldots, J-1, \tag{10}$$

where $\mathbf{x}_i = \begin{pmatrix} 1 & x_{1i} & x_{2i} & \cdots & x_{ki} \end{pmatrix}^\intercal$ for $i = 1, \ldots, n$.

Suppose the response categories re nominal, that is, have no ordering. In this case, we may consider the *generalized logit model*:

$$p_{ij} = \frac{\exp(\mathbf{x}_i^\intercal \boldsymbol{\beta}_j)}{\sum_{l=0}^{J-1} \exp(\mathbf{x}_i^\intercal \boldsymbol{\beta}_l)}, \text{ for } j = 0, \ldots, J-1, \tag{11}$$

where $\boldsymbol{\beta}_j = \begin{pmatrix} \beta_{j0} & \beta_{j1} & \cdots & \beta_{jk} \end{pmatrix}^\intercal$.

SES	Life Events	Count	Estimate	95% CI lower bound	95% CI upper bound
0	0	1	0.396506	0.096059	1.636675
	1	3	0.540551	0.158334	1.845440
	2	2	0.736926	0.249432	2.177188
	3	3	1.004642	0.368208	2.741129
	4	3	1.369616	0.501460	3.740769
	5	2	1.867180	0.630203	5.532120
	6	1	2.545502	0.742501	8.726699
	8	1	4.730948	0.915136	24.457420
	9	2	6.449640	0.982321	42.346488
1	0	1	0.077708	0.011740	0.514377
	1	2	0.105938	0.020316	0.552412
	2	2	0.144424	0.034495	0.604687
	3	5	0.196892	0.056880	0.681549
	4	2	0.268420	0.089704	0.803195
	5	2	0.365934	0.132703	1.009076
	6	1	0.498873	0.181299	1.372728
	7	2	0.680107	0.228639	2.023045
	8	3	0.927182	0.270204	3.181546
	9	2	1.264015	0.305120	5.236406

Table 2: Estimates for the odds ratios given $\mathbf{x}_i$ with $\hat{\boldsymbol{\gamma}}$.

Identifiability may be enforced by taking $\boldsymbol{\beta}_{J-1} = \mathbf{0}$, to give

$$\log \frac{p_{ij}}{p_{i,J-1}} = \mathbf{x}_i^\intercal \boldsymbol{\beta}_j, \text{ for } j = 0, \ldots, J-2, \tag{12}$$

with $p_{i,J-1} = 1 - \sum_{j=0}^{J-2} p_{ij}$. Consider the case of $j = 3$ levels and a single binary covariate x so that that $\mathbf{x}_i = \begin{pmatrix} 1 & x_i \end{pmatrix}^\intercal$. Give a 3×2 table containing the probabilities of $\mathbb{P}(Y = j \mid x)$ in terms the β_{jx} coefficients for rows $j = 0, 1, 2$ and columns $x = 0, 1$. Hence, give interpretations of $\exp(\beta_{jx})$ for $j = 0, 1, 2$ and $x = 0, 1$.

Is the generalized logit model suitable for ordinal data?

j	$x = 0$	$x = 1$
0	$\frac{\exp(\beta_{00})}{1 + \exp(\beta_{00}) + \exp(\beta_{10})}$	$\frac{\exp(\beta_{00} + \beta_{01})}{1 + \exp(\beta_{00} + \beta_{01}) + \exp(\beta_{10} + \beta_{11})}$
1	$\frac{\exp(\beta_{10})}{1 + \exp(\beta_{00}) + \exp(\beta_{10})}$	$\frac{\exp(\beta_{10} + \beta_{11})}{1 + \exp(\beta_{00} + \beta_{01}) + \exp(\beta_{10} + \beta_{11})}$
2	$\frac{1}{1 + \exp(\beta_{00}) + \exp(\beta_{10})}$	$\frac{1}{1 + \exp(\beta_{00} + \beta_{01}) + \exp(\beta_{10} + \beta_{11})}$

Table 3: Multinomial probabilities for various j and x.

Solution: See Table 3 for the table of probabilities.

Equation 12 provides a way to interpret the β_{jx}. Let p_{0j} and p_{1j} denote the probabilities when $x = 0$ and $x = 1$, respectively. In this case, we have the odds ratios:

$$\frac{p_{0j}}{p_{02}} = \exp(\beta_{j0})$$
$$\frac{p_{1j}}{p_{12}} = \exp(\beta_{j0}) \exp(\beta_{j1}).$$

Thus, the coefficients β_{j0} are the expected log odds ratio for level j relative to level $J-1=2$ when the $x=0$. β_{j1} is the expected increase in this log odds ratio when $x=1$. In this sense, we can consider the level $J-1$ the default case, and the $\exp\left(\beta_{jx}\right)$ express how much more likely we are to observe level j.

This model isn't suitable for ordinal data, for it is agnostic to the order of the data. From the above interpretation, it's more similar to fitting $J-1$ individual logisitic regression models. For an ordinal model, we might want behavior like the most probable level varies monotonically with some covariate. There's no way to model such behavior with the *generalized logit model* since each class has separate paramters.

5. Let
$$\pi_{ij} = \mathbb{P}\left(Y_i \leq j \mid \mathbf{x}_i\right), \tag{13}$$
for $j=0,\ldots,J-2$ and with $\mathbf{x}_i = \begin{pmatrix}1 & x_{1i} & x_{2i} & \cdots & x_{ki}\end{pmatrix}^\intercal$. Consider the proportional odds model
$$\log\frac{\pi_{ij}}{1-\pi_{ij}} = \alpha_j - \mathbf{x}_i^\intercal\boldsymbol{\beta}, \tag{14}$$
for $j=0,1,J-2$, and where $\boldsymbol{\beta} = \begin{pmatrix}\beta_0 & \beta_1 & \cdots & \beta_k\end{pmatrix}^\intercal$. Write down, in as simplified a form as possible, the log-likelihood $l\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right)$ where $\boldsymbol{\alpha} = \begin{pmatrix}\alpha_0 & \alpha_1 & \cdots & \alpha_{J-2}\end{pmatrix}^\intercal$, for the sample y_i, $i=1,\ldots,n$.

Solution: Let $\alpha_{J-1}=\infty$ and $\alpha_{-1}=-\infty$. In this case, we have that
$$\begin{aligned}p_{ij} = \pi_{i,j}-\pi_{i,j-1} &= \frac{1}{1+\exp\left(\mathbf{x}_i^\intercal\boldsymbol{\beta}-\alpha_j\right)} - \frac{1}{1+\exp\left(\mathbf{x}_i^\intercal\boldsymbol{\beta}-\alpha_{j-1}\right)} \\ &= \begin{cases}\frac{\exp(\alpha_j)-\exp(\alpha_{j-1})}{\exp\left(-\mathbf{x}_i^\intercal\boldsymbol{\beta}+\alpha_{j-1}+\alpha_j\right)+\exp(\alpha_{j-1})+\exp(\alpha_j)+\exp\left(\mathbf{x}_i^\intercal\boldsymbol{\beta}\right)}, & j=0,1,\ldots,J-2; \\ \frac{1}{1+\exp\left(\alpha_{J-2}-\mathbf{x}_i^\intercal\boldsymbol{\beta}\right)}, & j=J-1.\end{cases}\end{aligned} \tag{15}$$

Note that we must have $\alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_{J-2}$ for each class to have nonnegative probability.

Then, likelihood function is
$$L\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right) = \prod_{i=1}^{n}\prod_{j=0}^{J-1} p_{ij}^{\mathbf{1}_{\{j\}}(y_i)}, \tag{16}$$
where
$$\mathbf{1}_A\left(x\right) = \begin{cases}1, & x\in A; \\ 0, & \text{otherwise.}\end{cases} \tag{17}$$

Taking the log of Equation 16, we have the log-likelihood function
$$l\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right) = \sum_{i=1}^{n}\sum_{j=0}^{J-1}\mathbf{1}_{\{j\}}\left(y_i\right)\log p_{ij}. \tag{18}$$

6. For the data in Table 6, provide a single plot that shows the association between $\mathbf{p}\left(\mathbf{x}_i\right)$ and x_{1i} and x_{2i}, on a scale you feel is appropriate.

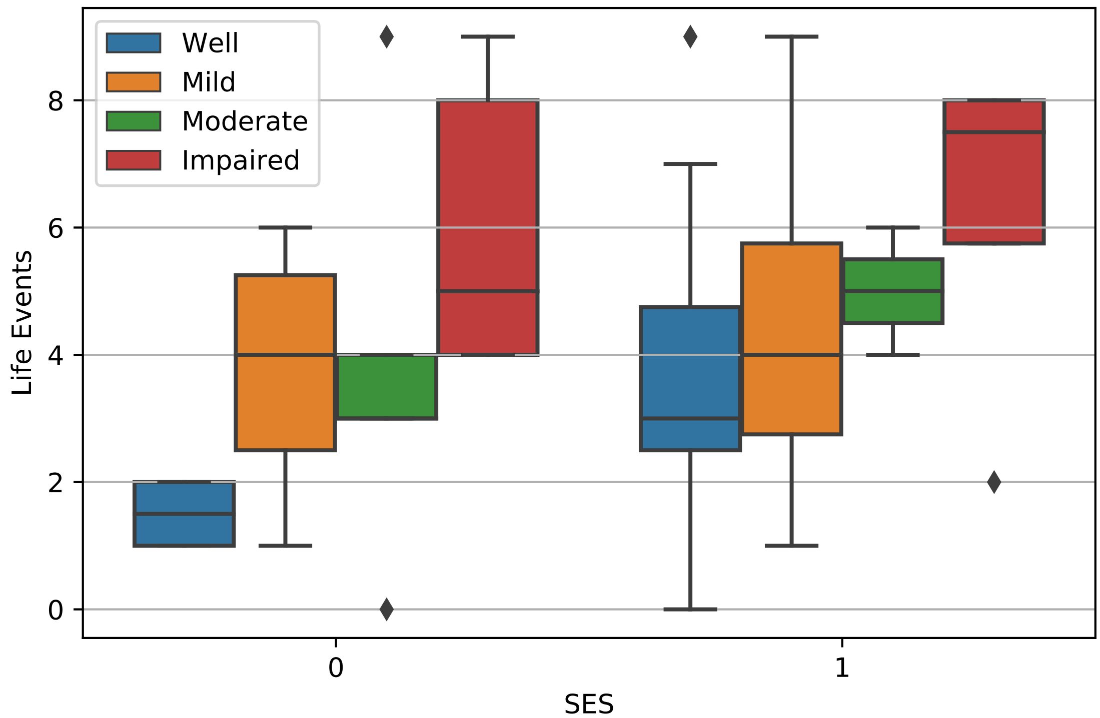


Figure 2: Boxplots showing the relationship between mental impairment, SES and life events.

Solution: See Figure 2. Conditioned on SES, the expected probability of having a more severe form of mental impairment increases with life events.

The effect of SES on observed mental impairment is more ambiguous. When looking at the *Well* and *Moderate* levels, it seems that SES may have a slight protective effect against mental impairment, for we observe many subjects with a high number of life events but no mental impairment or less severe impairment. There doesn't seem to be much evidence of this phenomenon in the *Mild* level. Ultimately, there's probably not enough data to come to any conclusion about SES.

7. Fit the proportional odds models:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \alpha_j^{(0)} \tag{19}$$

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \alpha_j^{(1)} - x_{1i}\beta_1^{(1)} \tag{20}$$

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \alpha_j^{(2)} - x_{2i}\beta_2^{(2)} \tag{21}$$

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \alpha_j^{(12)} - x_{1i}\beta_1^{(12)} - x_{2i}\beta_2^{(12)}. \tag{22}$$

Compare models using likelihood ratio statistics, and summarize the association between mental impairment, SES and life events, using your favored model.

Solution: The results of fitting the models can be seen in Table 4. The coefficient estimates agree with what we saw in Figure 2 and the discussion in the solution of

	α_0	α_1	α_2	β_1	β_2	Log-likelihood
Equation 19	-0.847	0.405	1.237			-54.521
Equation 20	-1.364	-0.042	0.831	-0.855		-53.437
Equation 21	0.261	1.656	2.588		0.288	-51.264
Equation 22	-0.282	1.213	2.209	-1.111	0.319	-49.549

Table 4: Results of fitting various models corresponding to each equation.

Part 6: positive β_2 indicate that observed severity of mental impairment increases with life events, and negative β_1 indicate that SES lessens the observed severity of mental impairment.

Wilks' theorem tells us how to compute a test statistic for a log-likelihood ratio test:

$$D = 2\left(l\left(\boldsymbol{\alpha}^{(p)}, \boldsymbol{\beta}^{(p)}\right) - l\left(\boldsymbol{\alpha}^{(q)}, \boldsymbol{\beta}^{(q)}\right)\right) \xrightarrow{\mathcal{D}} \chi^2_{p-q}, \tag{23}$$

where $p > q$ and $\boldsymbol{\beta}^{(p)}$ and $\boldsymbol{\beta}^{(q)}$ each have dimensionality p and q, respectively. That is, the deviance converges asymptotically to a chi-squared distribution with $p - q$ degrees of freedom.

The test statistic in Equation 23 can be computed for various pairs of models with the last column of Table 4 when the two models are nested.

Results of the tests on pairs of nested models can be seen in Table 5. Since we are testing multiple hypotheses, we might consider apply the Bonferroni correction and rejecting at significance level α/m, where m is the number of tests and α is the desired familywise error rate (FWER). In this case, for $\alpha = 0.05$, we would reject the null hypothesis in tests with a p-value of less than 0.01.

Alternate Model	Null Model	Degrees of freedom	Deviance	p-value
Equation 20	Equation 19	1	2.16770	0.14094
Equation 21	Equation 19	1	6.51498	0.01070
Equation 22	Equation 19	2	9.94416	0.00693
Equation 22	Equation 20	1	7.77646	0.00529
Equation 22	Equation 21	1	3.42918	0.06405

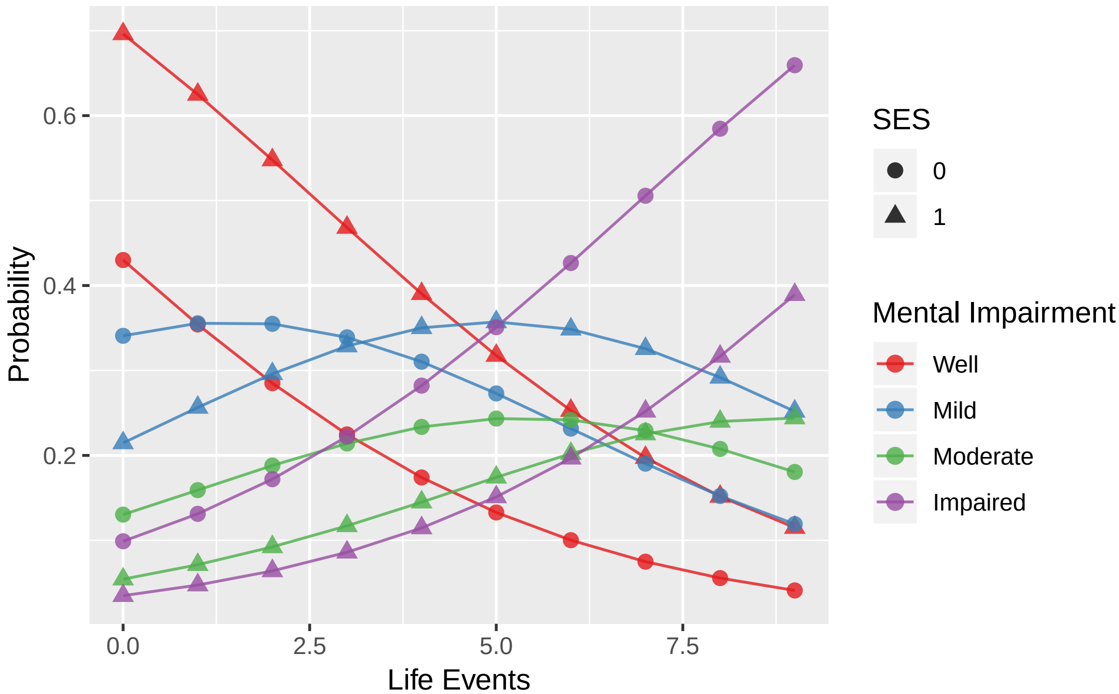
Table 5: Likelihood ratio tests comparing the models in Table 4.

From the first test that compares the model in Equation 20 against the null model in Equation 19, SES alone does not improve the fit very much. From the second test that tests the model in Equation 21 against the null model, we see that life events improves fit more: the p-value of 0.01070 is the on the border of statistical significance. The improved fit from the model in Equation 22 that incorporates both SES and life events is found to be statistically significant when compared against the null model and the model only containing SES. However, when tested against the model containing only life events, it doesn't quite meet statistical significance.

Despite the last test not rejecting the model in Equation 21, I'd still favor the model in Equation 22 that includes both SES and life events with the caveat that an additional experiment should be done. If possible, an additional experiment should be done with higher power (more subjects) that tests the null model in Equation 21 against the alternate model in Equation 22.

8. Provide a plot of fitted probabilities under the model in Equation 22, as a function of x_1 and x_2.

Figure 3: Fitted probabilities of mental impairment levels using the model described in Equation 22.



Solution: See Figure 3. As described in the solution of Part 7, life events increases the expected observed level of mental impairment. SES lessens the severity so curves of the same color are shifted downward.

Appendix

Code

Code for the logistic regression model and boxplots can be found in `mental_impairment.ipynb`. Code for the proportional odds models and Figure 3 can be found in `proportional_odds.ipynb`.

Data

The raw mental impairment data is in Table 6.

Subject	Mental Impairment	SES	Life Events
1	Well	1	1
2	Well	1	9
3	Well	1	4
4	Well	1	3
5	Well	0	2
6	Well	1	0
7	Well	0	1
8	Well	1	3
9	Well	1	3
10	Well	1	7
11	Well	0	1
12	Well	0	2
13	Mild	1	5
14	Mild	0	6
15	Mild	1	3
16	Mild	0	1
17	Mild	1	8
18	Mild	1	2
19	Mild	0	5
20	Mild	1	5
21	Mild	1	9
22	Mild	0	3
23	Mild	1	3
24	Mild	1	1
25	Moderate	0	0
26	Moderate	1	4
27	Moderate	0	3
28	Moderate	0	9
29	Moderate	1	6
30	Moderate	0	4
31	Moderate	0	3
32	Impaired	1	8
33	Impaired	1	2
34	Impaired	1	7
35	Impaired	0	5
36	Impaired	0	4
37	Impaired	0	4
38	Impaired	1	8
39	Impaired	0	8
40	Impaired	0	9

Table 6: Data on mental impairment, socioeconomic status (SES) and life events, for 40 subjects.