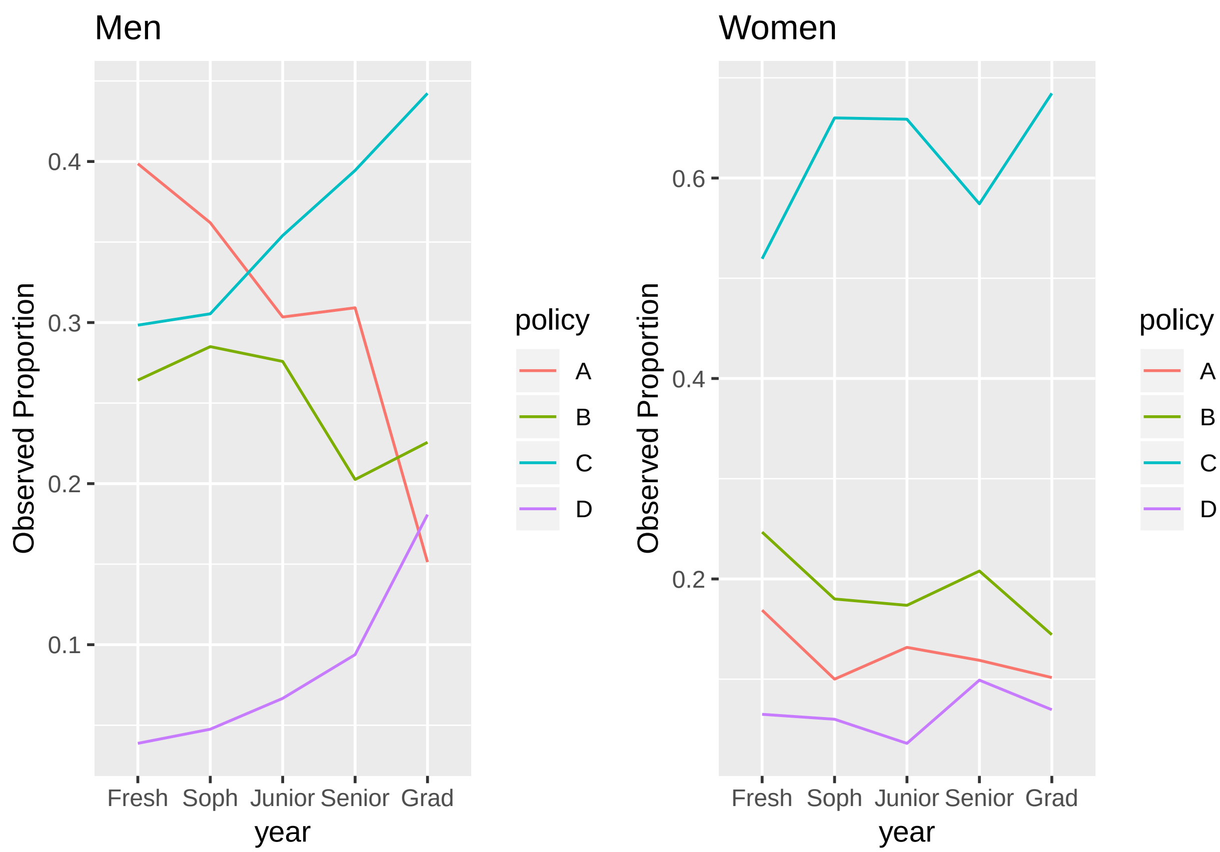
HW 4 Question 1

(a)



(b)

We first arrange the policy choices into an ordinal outcome A < B < C < D. We then fit the proportional odds model

$$\log \frac{\gamma_j(x_i)}{1-\gamma_j(x_i)} = \theta_j - \beta_1 \mathbf{1}_{\{\text{Male}\}} - \beta_2^T \text{year} - \beta_3^T \text{sex:year},$$

where $\gamma_j(x_i) = P(Y_i \le j | x_i)$. Results of the fit are shown in the table below.

	Value	Std. Error	t value
sexMale	-0.98	0.22	-4.37
yearSoph	0.39	0.33	1.18
yearJunior	0.24	0.25	0.98
yearSenior	0.34	0.27	1.24
yearGrad	0.48	0.24	1.98
sexMale:yearSoph	-0.27	0.35	-0.76
sexMale:yearJunior	0.16	0.28	0.57
sexMale:yearSenior	0.24	0.30	0.80

	Value	Std. Error	t value
sexMale:yearGrad	0.86	0.27	3.19
A|B	-1.39	0.21	-6.71
B|C	-0.29	0.21	-1.41
C|D	2.17	0.21	10.22

Our plots in part (a) reveal that year has a different effect on policy choice based on whether a respondent is male or female. Therefore, it is sensible to include an interaction term in the model.

(c)

Next we fit the main effects proportional odds model

$$\log \frac{\gamma_j(x_i)}{1-\gamma_j(x_i)} = \theta_j - \beta_1 \mathbf{1}_{\{\text{Male}\}} - \beta_2^T \text{year},$$

where $\gamma_j(x_i) = P(Y_i \leq j|x_i)$. Results of the fit are shown in the table below.

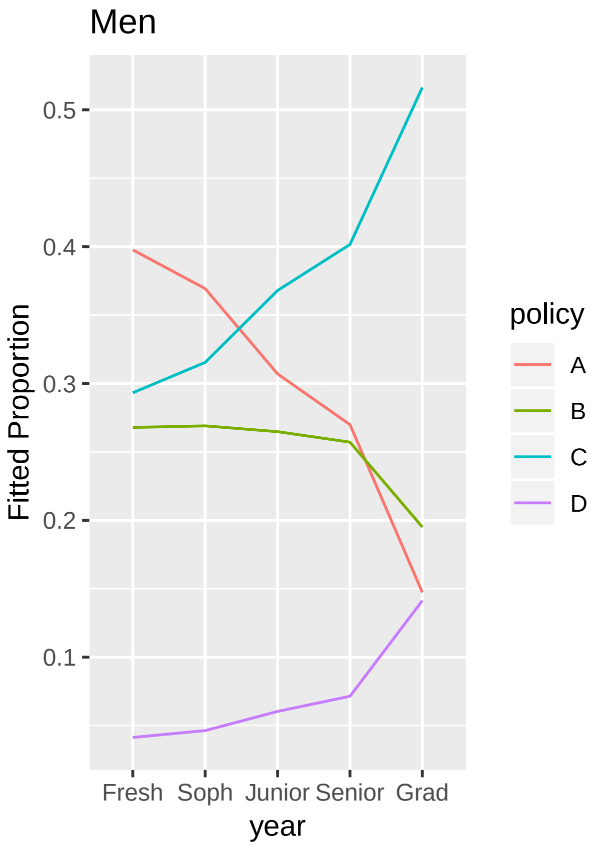
	Value	Std. Error	t value
sexMale	-0.65	0.08	-7.61
yearSoph	0.13	0.11	1.15
yearJunior	0.40	0.11	3.61
yearSenior	0.54	0.11	4.84
yearGrad	1.18	0.10	11.51
A|B	-1.11	0.11	-10.02
B|C	-0.01	0.11	-0.12
C|D	2.44	0.12	20.45

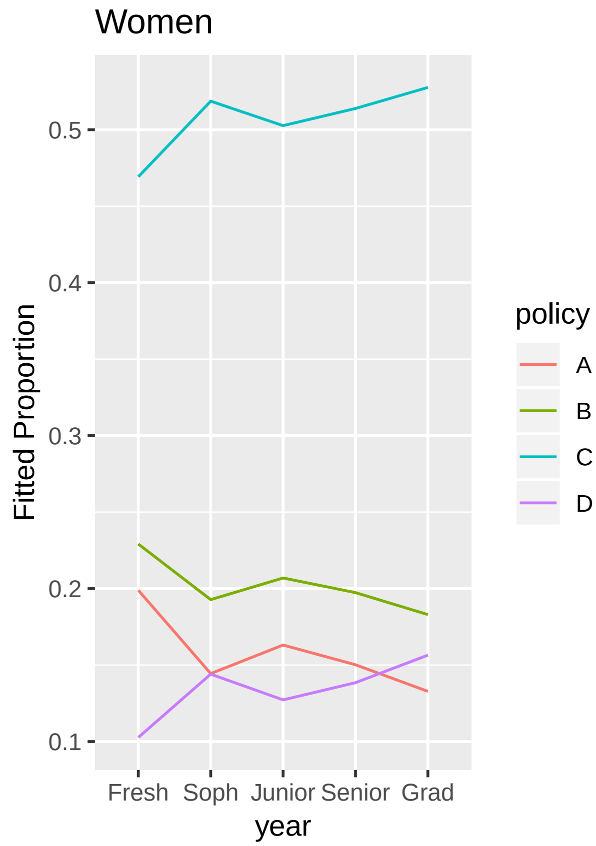
We can compare the two models using a likelihood ratio test. The difference in deviance between the main effects model and the interactions model is 7757-7736 = 21.

Under the null hypothesis that the main effects model is sufficient, we expect the difference in deviance to be distributed according to a χ_4^2 distribution (note that there are 4 different sex:year interaction terms). This yields a p-value of 3×10^{-4}.

We therefore reject the null, and conclude that the model with interactions is prefered.

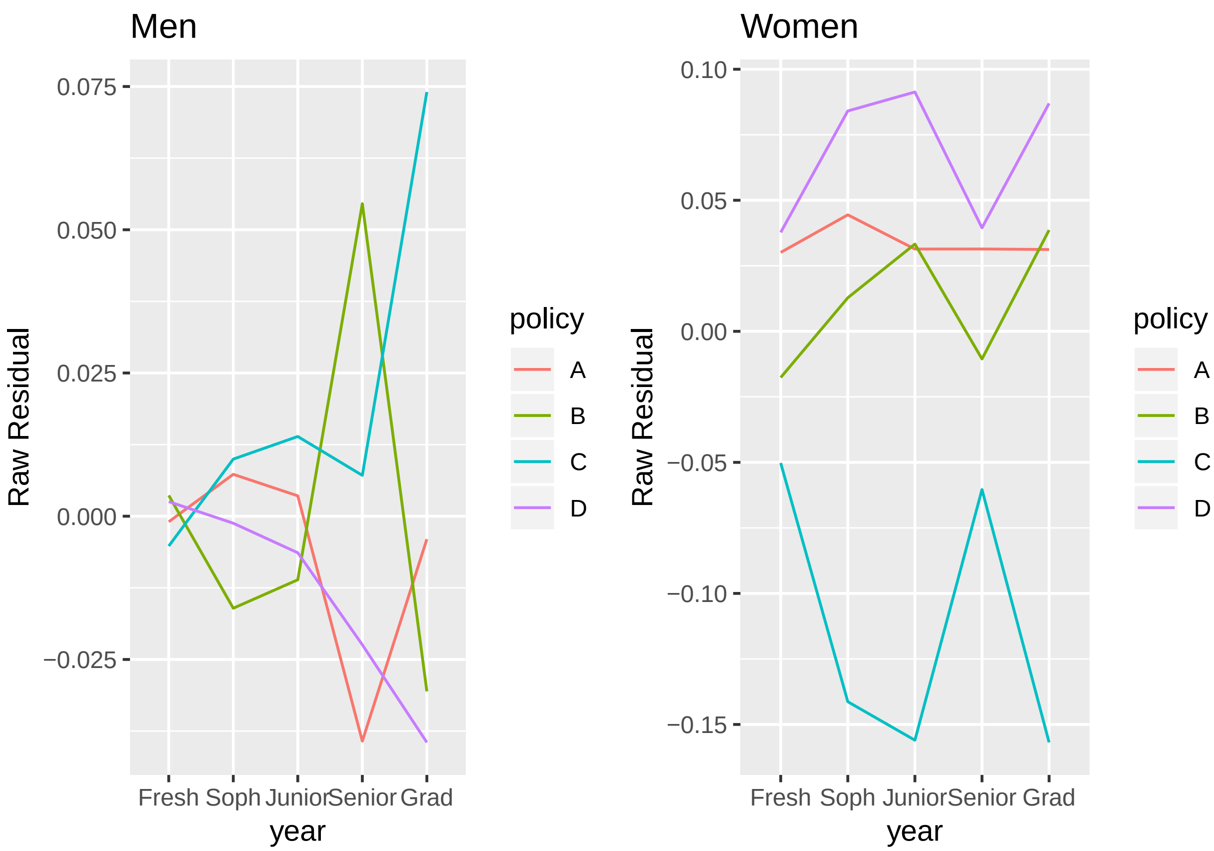
(d) and (e)





Comparing to the observed proportions plotted in part (a), we can see that the fitted proportions for men recapitulate the observed proportions well. The key trends are also preserved for women, but overall the fit doesn't look as good (see also the raw residual plot in (f)). Given that there are many more male respondents than female respondents (2565 vs 582), a better fit to the male data is expected.

(f)



We can see that men tend to have much smaller residuals than women, and that the fitted values for women tend to be entirely larger than or smaller than the observed proportions for each policy (e.g. the fitted proportions for policy D are larger than the observed proportions across all years).

We might improve the residuals by normalizing in some way. If the data do indeed come from this model, we don't expect the variances of the observed proportions to all be the same. Therefore, making claims based on the relative sizes of the raw residuals could be error prone.

(g)

For convenience, the table from part (b) is repeated below.

	Value	Std. Error	t value
sexMale	-0.98	0.22	-4.37
yearSoph	0.39	0.33	1.18
yearJunior	0.24	0.25	0.98
yearSenior	0.34	0.27	1.24
yearGrad	0.48	0.24	1.98
sexMale:yearSoph	-0.27	0.35	-0.76
sexMale:yearJunior	0.16	0.28	0.57
sexMale:yearSenior	0.24	0.30	0.80
sexMale:yearGrad	0.86	0.27	3.19
A|B	-1.39	0.21	-6.71

	Value	Std. Error	t value
B|C	-0.29	0.21	-1.41
C|D	2.17	0.21	10.22

First, let's look at the year terms. Note that because we included year*sex interactions in the model, and women are the reference sex, these coefficients correspond to the effect of year on women. These coefficients are either not significant or (at best) very marginally significant, which means that womens' opinions don't vary much across the years. This is not too surprising given our initial plots, which showed that the trends for women were flatter than those for men.

(h)

Now we fit the multinomial main effects model

$$\log \frac{p_{ij}}{p_{i1}} = \beta_{0j} + \beta_{1j}\mathbf{1}_{\{\text{Male}\}} + \beta_{2j}^T \text{year},$$

for $j = B, C, D$ (i.e. a separate β is fit for each outcome), and where $p_{ij} = P(Y_i = j|x_i)$. Code is shown below only to demonstrate proper use of the gof function-normally this would not be included in a write-up.

```
library(nnet)
mod_multi_main <- multinom(policy~sex+year, uncviet, weights = y)
ldata <- t(matrix(uncviet$y, 4,10))
res <- gof(mod_multi_main, ldata)
```

The deviance is given by

```
res[[3]]
```

```
## [1] 19.19418
```

The model has 12 degrees of freedom (30-3(6)), compared to 22 for the model in part (c). Requiring so many estimated parameters just for the main effects makes building on this model difficult, as we will see below in part (i) when we try to add an interaction term.

(i)

Next we add sex:year interactions to the model in part (h), that is, we fit the multinomial model

$$\log \frac{p_{ij}}{p_{i1}} = \beta_{0j} + \beta_{1j}\mathbf{1}_{\{\text{Male}\}} + \beta_{2j}^T \text{year} + \beta_{3j}^T \text{sex:year},$$

for $j = B, C, D$ (i.e. a separate β is fit for each outcome), and where $p_{ij} = P(Y_i = j|x_i)$.

Next we build on the multinomial model by adding sex:year interactions in addition to the main effects. The deviance of this model (again obtained using the gof function) is zero. This may seem like a surprising result, but notice that this model has 30-3(10) = 0 degrees of freedom. In other words, it is saturated! So using the multinomial model does not allow for a meaningful fit of the full sex:year interaction term. On the other hand, the proportional odds model does allow us to include the interaction term, because doing so does not use up all the degrees of freedom. The assumption of proportional odds reduces the number of estimated parameters in the model for a given set of covariates-instead of estimating three separate beta vectors, we only need to estimate a single beta and three intercept terms.