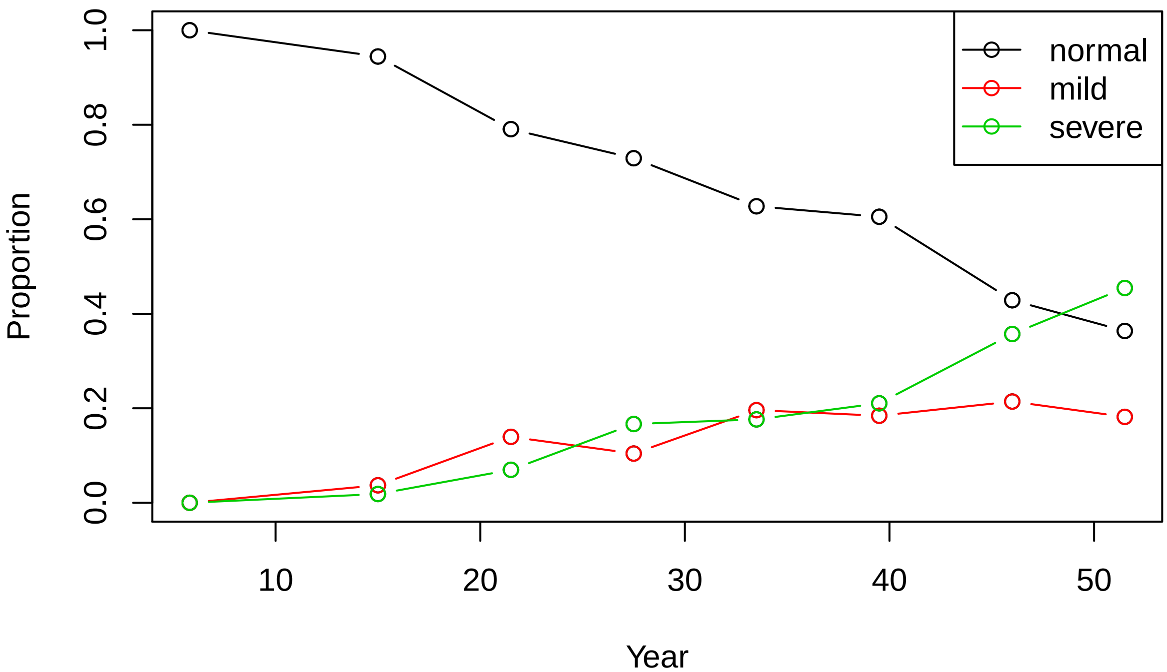
HW3 Q1 solution

(a)



From the plot we can see that the proportion of "normal" significantly decreases as year goes by, while the proportion of "mild" and "severe" increase slowly with year, and more specifically, the increasing trend of "severe" is more significant than that of "mild".

(b)

```
## # weights:  9 (4 variable)
## initial  value 407.585159
## iter  10 value 208.809599
## final  value 208.724782
## converged

## [1] "coefficients:"

##        (Intercept)   year
## mild         -4.29 0.0836
## severe       -5.06 0.1093

## [1] "T values for year:"

##   mild severe
##   5.47   6.64

## [1] "Residual deviance:"

## [1] 417
```

The coefficients for predictor “year”" are positive and significant, which suggests that the odds of moving from normal to mild/severe disease will increase over time. Specifically, the slope of “severe”" category is greater than that of “mild”" category, and hence the probability of getting severe disease should grow faster over time than that of getting mild disease.

(c)

```
## # weights:  9 (4 variable)
## initial  value 407.585159
## iter  10 value 204.490734
## final  value 204.434441
## converged
```

```
## [1] "coefficients:"
```

```
##        (Intercept) log(year)
## mild         -8.94      2.17
## severe      -11.98      3.07
```

```
## [1] "T values for year:"
```

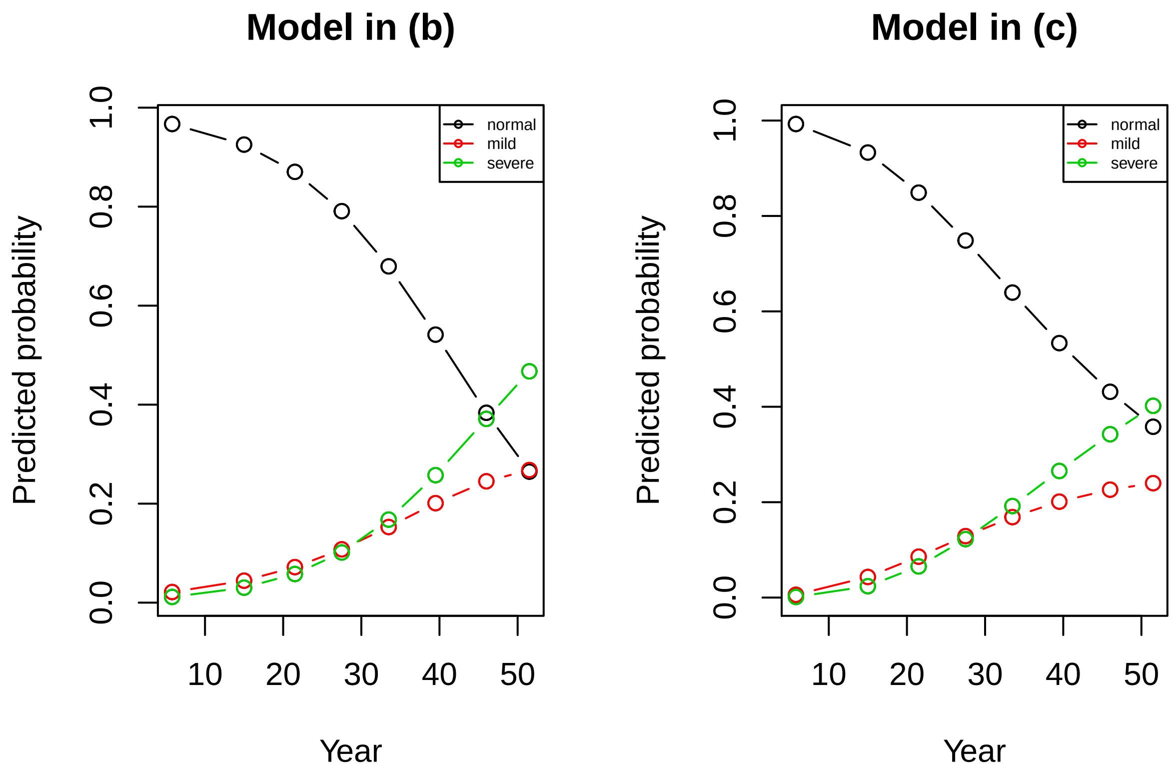
```
##   mild severe
##   4.73   5.43
```

```
## [1] "Residual deviance:"
```

```
## [1] 409
```

We can see that after changing predictor from “year” to “log(year)”, the fitting result gets better at least in terms of residual deviance. We have used the multinom function. The difference of deviances can be used to compare likelihoods.

(d)



(e)

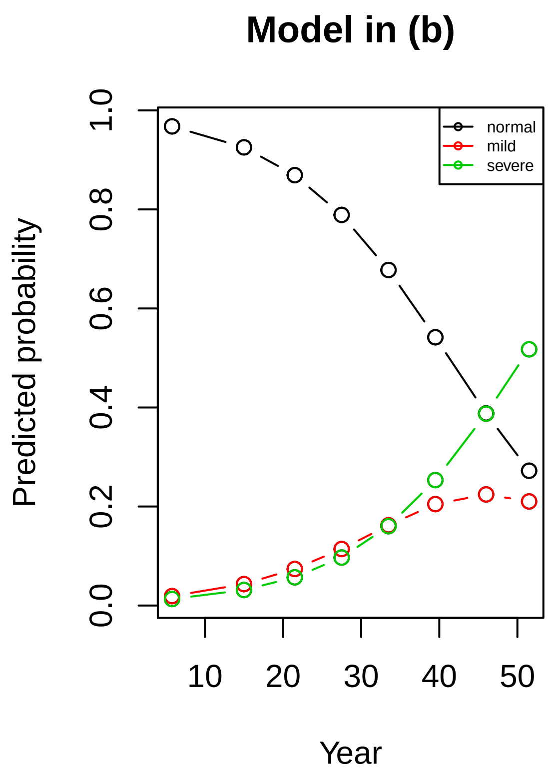
```
## Call:
## polr(formula = status ~ year, data = pneumo, weights = pneumo$Freq)
##
## Coefficients:
##       Value Std. Error t value
## year 0.0959     0.0119    8.03
##
## Intercepts:
##              Value  Std. Error t value
## normal|mild  3.956  0.410      9.656
## mild|severe  4.869  0.441     11.038
##
## Residual Deviance: 416.92
## AIC: 422.92
```

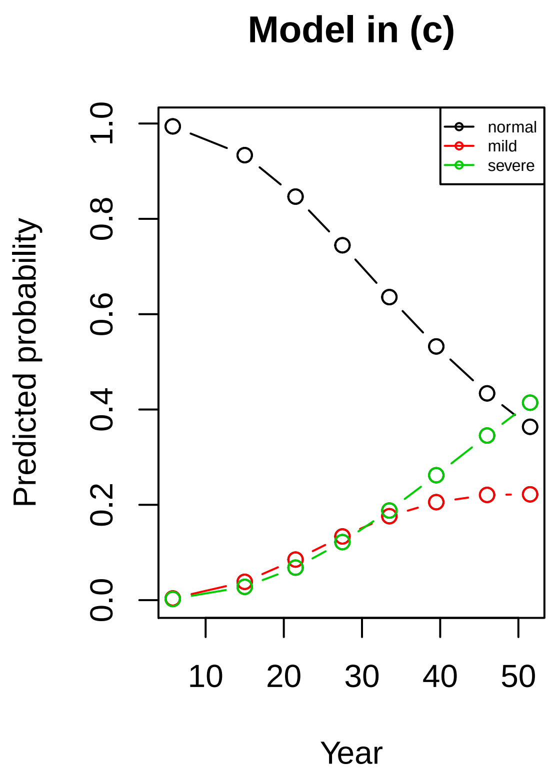
From the result we can see that the estimated value of θ_1 is 3.956, which is the difference between logit of the probability of "normal" and "mild", for any fixed year.

(f)

We'll repeat part (c) and (d) for the proportional odds model) If we use "log(year)" as predictor and fit a proportional odds model, the result is the following. Again

```
## Call:
## polr(formula = status ~ log(year), data = pneumo, weights = pneumo$Freq)
##
## Coefficients:
##            Value Std. Error t value
## log(year)    2.6      0.381    6.82
##
## Intercepts:
##              Value  Std. Error t value
## normal|mild   9.676  1.323      7.312
## mild|severe 10.582  1.344      7.875
##
## Residual Deviance: 408.55
## AIC: 414.55
```





(g)

It can be seen that the fitting results are largely the same.

(h)

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.9664     0.4189   -9.47 2.86e-21
## year          0.0963     0.0124    7.79 6.90e-15
```

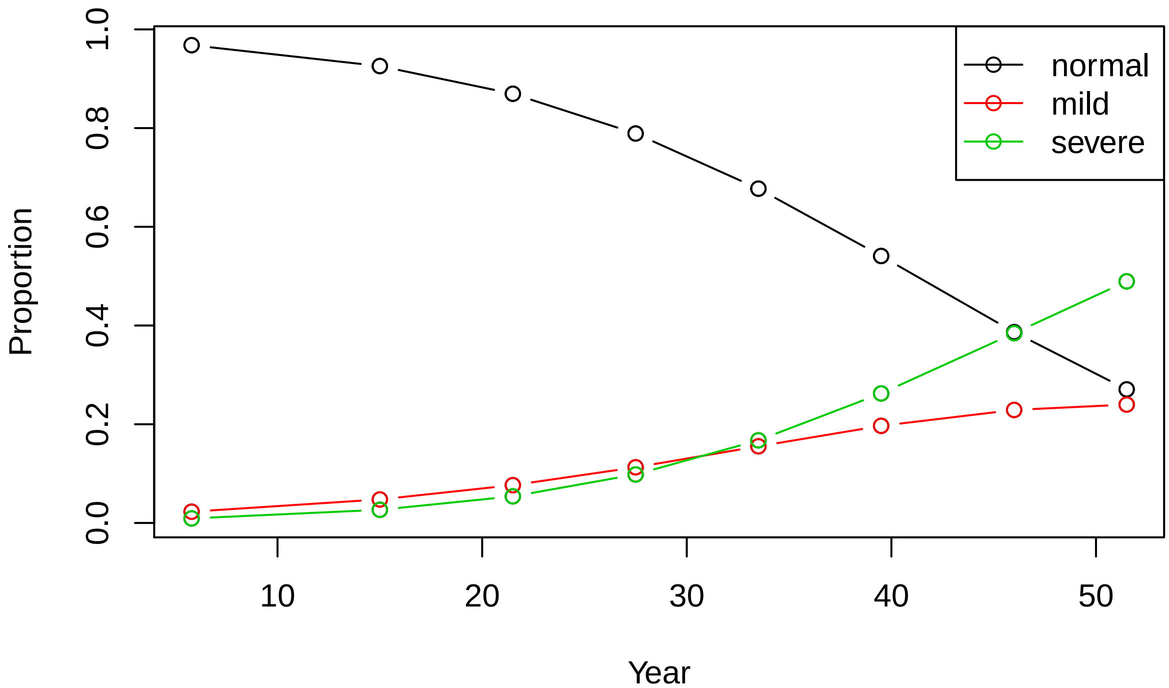
It can be seen that "year" has a significant positive efect on the odds of getting lung disease.

(i)

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.1134     0.8625   -1.29    0.197
## year          0.0355     0.0235    1.51    0.131
```

From the result we can see that the "year" effct is not significant in this model. And the probability of a mild disease within the diseased group under temperature t is

$$\frac{1}{1 + \exp(-1.1134 + 0.0355t)}$$



(j)

The fitting result is pretty similar to that of multinomial model and proportional odds model.