Assignment 1: Survival Analysis

Faizan Khalid Mohsin

January 23, 2019

Table of Contents

[Question 1 1](#_Toc536610076)

[Question 1 d: Fitting the Model and Interpreting the Results 1](#_Toc536610077)

[Question 1 e: Assessing the Overall Model Fit. 2](#_Toc536610078)

[Question 1 f: Model Fit for Model with Interaction Terms. 4](#_Toc536610079)

[Question 4 6](#_Toc536610080)

[Question 4 a: CHD Mortality Analyisis 6](#_Toc536610081)

[Question 4 b: Non-CHD Mortality Analysis 10](#_Toc536610082)

[Question 4 c: CHD Mortality Log-Rate Analysis 12](#_Toc536610083)

# 

# Question 1

## Question 1 d: Fitting the Model and Interpreting the Results

age = rep(22:29, 2)  
death = c(433, 412, 337, 331, 287, 242, 215, 192,   
 24, 36, 66, 102, 138, 171, 185, 200)  
personyears = c(91444, 86835, 75892, 63241, 52023, 42123, 36915, 32215,  
 8556, 12708, 23203, 35415, 46207, 55675, 60470, 64770)   
z = c(rep(0, 8), rep(1, 8))  
data = data.frame(z, age, personyears)  
model1 = glm(death ~ z + factor(age) + offset(log(personyears)),   
 family = poisson(link="log") )  
summary(model1)

##   
## Call:  
## glm(formula = death ~ z + factor(age) + offset(log(personyears)),   
## family = poisson(link = "log"))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.44325 -0.14770 0.00831 0.30377 1.05482   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.348768 0.046822 -114.235 < 2e-16 \*\*\*  
## z -0.602158 0.041853 -14.388 < 2e-16 \*\*\*  
## factor(age)23 0.004702 0.066494 0.071 0.94362   
## factor(age)24 -0.044167 0.068449 -0.645 0.51876   
## factor(age)25 0.097314 0.067515 1.441 0.14948   
## factor(age)26 0.145055 0.068388 2.121 0.03392 \*   
## factor(age)27 0.179327 0.069648 2.575 0.01003 \*   
## factor(age)28 0.183552 0.070734 2.595 0.00946 \*\*   
## factor(age)29 0.197412 0.071658 2.755 0.00587 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 226.5601 on 15 degrees of freedom  
## Residual deviance: 2.2334 on 7 degrees of freedom  
## AIC: 131.32  
##   
## Number of Fisher Scoring iterations: 3

The coefficient of the marital status is -0.602. Therefore, its exponential is 0.5477151. Which is the rate ratio between being single and married. Hence, the mortality rate of people who are single is approximately twice that of those who are single (after adjusting for age).

## Question 1 e: Assessing the Overall Model Fit.

Will use the predict function to estimate the expected events in each category. After that we need to exponentiate the estimated values as they are the log expected number of deaths.

The below code produces the expected number of events in each age/marital status group.

# Will use the predict function to estimate the expected results.  
  
predicted = predict(model1, newdata = data)  
  
events = exp(predicted) # predicted values.   
expected\_events = data.frame(Age = age[1:8], Single = events[1:8], Married = events[9:16])

We get the following expected number of events:

kable(expected\_events, caption = "Expected number of events for each age/marital status group.", digits = 2)

Expected number of events for each age/marital status group.

|  |  |  |
| --- | --- | --- |
| Age | Single | Married |
| 22 | 434.73 | 22.27 |
| 23 | 414.76 | 33.24 |
| 24 | 345.20 | 57.80 |
| 25 | 331.38 | 101.62 |
| 26 | 285.92 | 139.08 |
| 27 | 239.59 | 173.41 |
| 28 | 210.85 | 189.15 |
| 29 | 186.57 | 205.43 |

Just from looking at the values they look quite similar to the observed number of events.

Compare the expected numbers to observed event counts to assess the overall model ﬁt. What statistical test can you use for this?

We will use the chi-square goodness of fit test to compare the expected to observed event counts to assess the overall model fit.

chisq.test(death, events)

## Warning in chisq.test(death, events): Chi-squared approximation may be  
## incorrect

##   
## Pearson's Chi-squared test  
##   
## data: death and events  
## X-squared = 240, df = 225, p-value = 0.2348

The chi-square test with p-value of 0.2348 suggests that the observed event count does follow the distribution of the fitted model and the overall model fit is reasonable.

To be clear, we are using Pearson’s Chi-square goodness of fit test.

We also perform a likelihood ratio test to test the overall goodness of fit of the model, which we know follows a chi-square distribution with the degrees of freedom being the difference in the number of parameters estimated. Here, the null model (nested in the fitted model), is compared to our fitted model. The null model only has the intercept paramter that is being estimated and our fitted model has 9 parameters including the intercept term being estimated. Hence, the dgrees of freedom for the likelihood ratio test is 8. We do this test both manually and using the “lmtest” package.

require(lmtest)

## Loading required package: lmtest

## Warning: package 'lmtest' was built under R version 3.4.4

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

deviance1 = model1$null.deviance - model1$deviance  
1-pchisq(deviance1, 8)

## [1] 0

lrtest(model1)

## Likelihood ratio test  
##   
## Model 1: death ~ z + factor(age) + offset(log(personyears))  
## Model 2: death ~ 1  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 9 -56.66   
## 2 1 -665.46 -8 1217.6 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Using both methods we found the p-value to be practically zero. Hence, the overall model is statistically significant, the same as we found using the chi-square test. Hence, the model fits the data well.

## Question 1 f: Model Fit for Model with Interaction Terms.

model2 = glm(death ~ z + factor(age) + z:factor(age) +   
 offset(log(personyears)), family = poisson(link="log") )  
summary(model2)

##   
## Call:  
## glm(formula = death ~ z + factor(age) + z:factor(age) + offset(log(personyears)),   
## family = poisson(link = "log"))  
##   
## Deviance Residuals:   
## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.352744 0.048057 -111.383 < 2e-16 \*\*\*  
## z -0.523590 0.209705 -2.497 0.01253 \*   
## factor(age)23 0.002003 0.068823 0.029 0.97679   
## factor(age)24 -0.064239 0.072642 -0.884 0.37652   
## factor(age)25 0.100155 0.073011 1.372 0.17013   
## factor(age)26 0.152785 0.076117 2.007 0.04472 \*   
## factor(age)27 0.193333 0.080260 2.409 0.01600 \*   
## factor(age)28 0.207009 0.083430 2.481 0.01309 \*   
## factor(age)29 0.230052 0.086705 2.653 0.00797 \*\*   
## z:factor(age)23 0.007864 0.272362 0.029 0.97697   
## z:factor(age)24 0.078191 0.249189 0.314 0.75369   
## z:factor(age)25 -0.073738 0.238330 -0.309 0.75702   
## z:factor(age)26 -0.090084 0.233895 -0.385 0.70013   
## z:factor(age)27 -0.102622 0.232285 -0.442 0.65864   
## z:factor(age)28 -0.120222 0.232449 -0.517 0.60502   
## z:factor(age)29 -0.133998 0.232776 -0.576 0.56485   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 2.2656e+02 on 15 degrees of freedom  
## Residual deviance: 1.4699e-13 on 0 degrees of freedom  
## AIC: 143.08  
##   
## Number of Fisher Scoring iterations: 3

First, we see that none of the interaction terms are statistically insignificant. This shows that the new model with the additional interaction terms are not improving the model fit. Secondly, the significance of the age factors 26, 27, 28 and 29, slightly decreased in significance though still remained significant. Further, the AIC increased when we add the interaction terms. It went from 131.32 to 142.08. As lower AIC values are better, this also shows that the model fit did not improve.

Lastly, we also do a likelihood ratio test, which we know follows a chi-square distribution with the degrees of freedom being the difference in the number of parameters estimated. We do the test using R below, using the package lmtest and manually as well.

aa = lrtest(model1, model2)  
aa

## Likelihood ratio test  
##   
## Model 1: death ~ z + factor(age) + offset(log(personyears))  
## Model 2: death ~ z + factor(age) + z:factor(age) + offset(log(personyears))  
## #Df LogLik Df Chisq Pr(>Chisq)  
## 1 9 -56.659   
## 2 16 -55.542 7 2.2334 0.9458

deviance = model1$deviance - model2$deviance  
pchisq(deviance, df = 7, lower.tail = FALSE)

## [1] 0.9457985

We find the p-value to be the same using both methods. The p-value is 0.946. Hence, the model fit does not improve after adding interaction terms.

# Question 4

## Question 4 a: CHD Mortality Analyisis

The data is obtained from the internet link provided by the professor, and is cleaning using code. We do not show this code as this was already provided by him. The one thing we will mention is that after the data is cleaned it is called “frdeaths”.

Now, it can be seen below, the age group ‘<35’ (the first row) does not have any events, as the has 44 for the 0 column, and then zero for all the rest of the columns. The columns are indicating the number of events. Hence, group ‘<35’ did not experience any events. We need all groups to have non-zero number of events. Therefore, we will combine this group with the group ‘35-45’ and have a new group called ‘<45’. Below is the code for this. The final dataset will be called frdeaths2.

# Shows that age group '<35' does not have any events.   
table(frdeaths$ageg, frdeaths$chd)

##   
## 0 1 2 3 4 5 6 8  
## 1 44 0 0 0 0 0 0 0  
## 2 80 4 0 0 0 0 0 0  
## 3 83 19 10 3 1 0 0 0  
## 4 48 25 16 9 6 7 4 1  
## 5 51 24 21 6 6 3 1 0  
## 6 28 18 19 15 5 5 1 1  
## 7 13 18 13 11 6 7 4 0  
## 8 12 8 13 6 3 5 4 1  
## 9 6 9 5 1 6 2 1 2

# Creating group '<45'  
frdeaths2 = frdeaths  
  
frdeaths2$ageg[frdeaths2$ageg == 1] = 2  
frdeaths2$ageg = frdeaths2$ageg - 1  
  
table(frdeaths2$ageg, frdeaths2$chd)

##   
## 0 1 2 3 4 5 6 8  
## 1 124 4 0 0 0 0 0 0  
## 2 83 19 10 3 1 0 0 0  
## 3 48 25 16 9 6 7 4 1  
## 4 51 24 21 6 6 3 1 0  
## 5 28 18 19 15 5 5 1 1  
## 6 13 18 13 11 6 7 4 0  
## 7 12 8 13 6 3 5 4 1  
## 8 6 9 5 1 6 2 1 2

From this table we can see that now all age groups have non-zero number of events, as the first row now has 4 under the column 1 (the second column).

We will now fit the model.

# Poisson regression for CHD mortality:  
model\_chd <- glm(chd ~ as.factor(yearg) + as.factor(ageg) + sexg + area, offset=log(followupyears), data=frdeaths2, family=poisson(link='log'))  
summary(model\_chd)

##   
## Call:  
## glm(formula = chd ~ as.factor(yearg) + as.factor(ageg) + sexg +   
## area, family = poisson(link = "log"), data = frdeaths2, offset = log(followupyears))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4231 -0.8383 -0.2598 0.3385 2.5968   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -8.642723 0.558096 -15.486 < 2e-16 \*\*\*  
## as.factor(yearg)2 -0.086425 0.365981 -0.236 0.813319   
## as.factor(yearg)3 -0.406372 0.385421 -1.054 0.291719   
## as.factor(yearg)4 -0.515168 0.385998 -1.335 0.181994   
## as.factor(yearg)5 -0.182582 0.350229 -0.521 0.602144   
## as.factor(yearg)6 -0.553775 0.374010 -1.481 0.138702   
## as.factor(yearg)7 -0.121814 0.337201 -0.361 0.717913   
## as.factor(yearg)8 0.008774 0.324435 0.027 0.978425   
## as.factor(yearg)9 -0.283330 0.333905 -0.849 0.396141   
## as.factor(yearg)10 -0.251526 0.327491 -0.768 0.442463   
## as.factor(yearg)11 -0.601970 0.342283 -1.759 0.078630 .   
## as.factor(yearg)12 -0.652896 0.340459 -1.918 0.055150 .   
## as.factor(yearg)13 -0.627740 0.332670 -1.887 0.059164 .   
## as.factor(yearg)14 -0.742263 0.333222 -2.228 0.025912 \*   
## as.factor(yearg)15 -0.719336 0.327588 -2.196 0.028102 \*   
## as.factor(yearg)16 -0.958997 0.336262 -2.852 0.004345 \*\*   
## as.factor(yearg)17 -0.783202 0.324739 -2.412 0.015874 \*   
## as.factor(yearg)18 -0.778580 0.320792 -2.427 0.015222 \*   
## as.factor(yearg)19 -1.040235 0.328061 -3.171 0.001520 \*\*   
## as.factor(yearg)20 -1.257356 0.334550 -3.758 0.000171 \*\*\*  
## as.factor(yearg)21 -0.913888 0.317969 -2.874 0.004051 \*\*   
## as.factor(yearg)22 -1.150742 0.324868 -3.542 0.000397 \*\*\*  
## as.factor(yearg)23 -1.202447 0.323372 -3.718 0.000200 \*\*\*  
## as.factor(yearg)24 -1.165199 0.319064 -3.652 0.000260 \*\*\*  
## as.factor(yearg)25 -1.451612 0.327191 -4.437 9.14e-06 \*\*\*  
## as.factor(yearg)26 -1.415661 0.324168 -4.367 1.26e-05 \*\*\*  
## as.factor(yearg)27 -1.736494 0.336232 -5.165 2.41e-07 \*\*\*  
## as.factor(yearg)28 -1.762550 0.335611 -5.252 1.51e-07 \*\*\*  
## as.factor(yearg)29 -1.364591 0.319896 -4.266 1.99e-05 \*\*\*  
## as.factor(ageg)2 2.689137 0.519989 5.172 2.32e-07 \*\*\*  
## as.factor(ageg)3 3.872122 0.506795 7.640 2.17e-14 \*\*\*  
## as.factor(ageg)4 4.595187 0.511198 8.989 < 2e-16 \*\*\*  
## as.factor(ageg)5 5.337886 0.512749 10.410 < 2e-16 \*\*\*  
## as.factor(ageg)6 6.000115 0.515377 11.642 < 2e-16 \*\*\*  
## as.factor(ageg)7 6.608856 0.519595 12.719 < 2e-16 \*\*\*  
## as.factor(ageg)8 7.472555 0.527613 14.163 < 2e-16 \*\*\*  
## sexg -0.948191 0.070821 -13.389 < 2e-16 \*\*\*  
## area -0.160390 0.070951 -2.261 0.023786 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 2121.50 on 719 degrees of freedom  
## Residual deviance: 609.07 on 682 degrees of freedom  
## AIC: 1624.4  
##   
## Number of Fisher Scoring iterations: 6

The calender year coefficient estimates from group year 14 to 29 are statistically significant.

Now we will present the estimated calender year log-rate ratios time trend below.

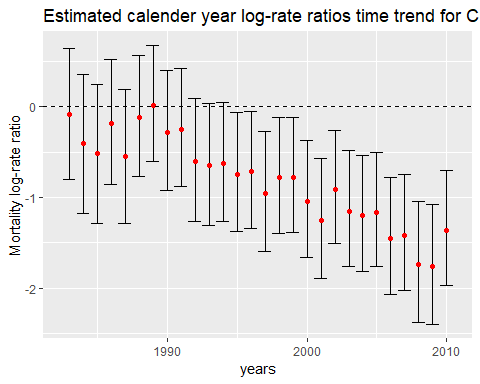
However, first we will create the function that will do this for us, plotting the coefficients with their CI’s. We will call it plot\_year\_trend().

# Defining function that will plot the estimated   
# calendar time (log rate-ratios) with their CI's.  
  
plot\_year\_trend = function(model, title, ylab){  
   
 # Extract variables from model  
 model\_year\_coef = coef(model)[2:29]  
 confints\_all = as.data.frame(confint(model))  
 confints\_calender\_years = confints\_all[2: 29,]  
 u\_ci = confints\_calender\_years[, 2]  
 l\_ci = confints\_calender\_years[, 1]  
   
 # Now we will create a dataframe for ggplot  
 df = data.frame(years = seq(1983, length.out =28),   
 model\_year\_coef,   
 l\_ci, u\_ci)  
   
 # Now we plot the coefficients with their CI's  
 ggplot(df, aes(x = years, y = model\_year\_coef)) +  
 geom\_errorbar(aes(ymax = u\_ci, ymin = l\_ci)) +  
 geom\_point(col = 2) +   
 geom\_hline(yintercept=0, linetype="dashed") +  
 ggtitle(title) + labs( y = ylab)  
}  
  
# Example  
# heading ="Calender year estimated time trend (log-rate ratios)"  
# plot\_year\_trend(model\_chd, heading)

Now we plot it.

heading ="Estimated calender year log-rate ratios time trend for CHD deaths"  
ylab = "Mortality log-rate ratio"  
plot\_year\_trend(model\_chd, heading, ylab)

## Waiting for profiling to be done...



We see that there is a downward trend of the log-rate ratios of the mortality rate over the calender year. Further we can see that the last 16 CI’s do not overlap with zero and are progressively found further away from the zero line over the calend years. Also, as a side note we see that perhaps there is some seasonality as well, with the wavey sort of pattern.

This downward trend also corroborates our findings when we fit the model with several of the log-rate ratios of the later calender years (year 14 and onward) being statistically significant. This means that the difference between the log-rate ratio of the mortality rate of the first calender year (the reference year/group) is statistically different from those of the later years.

## Question 4 b: Non-CHD Mortality Analysis

Now to model the non-CHD deaths, we first create the non-CHD death counts and then check if there are non-zero events in each age category.

frdeaths2$nonchd = frdeaths2$events - frdeaths2$chd  
table(frdeaths2$ageg, frdeaths2$nonchd)

##   
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18  
## 1 83 31 9 4 0 1 0 0 0 0 0 0 0 0 0 0 0  
## 2 39 28 25 14 5 2 2 0 1 0 0 0 0 0 0 0 0  
## 3 8 13 19 26 17 11 8 1 4 5 1 3 0 0 0 0 0  
## 4 18 20 25 18 15 8 4 2 1 0 1 0 0 0 0 0 0  
## 5 8 15 16 16 9 10 7 6 3 0 1 0 1 0 0 0 0  
## 6 9 5 6 6 8 6 6 9 6 4 4 0 0 3 0 0 0  
## 7 2 6 6 6 6 7 4 5 2 0 2 4 0 0 0 2 0  
## 8 4 3 3 4 3 3 3 2 2 1 0 0 1 0 1 0 2

Indeed there are stricly positive number of events per age category. We can now proceed with the analysis.

Below we model the non-CHD deaths similarly as we did for the CHD deaths.

model\_nonchd <- glm(nonchd~ as.factor(yearg) + as.factor(ageg) + sexg + area, offset=log(followupyears), data=frdeaths2, family=poisson(link='log'))  
summary(model\_nonchd)

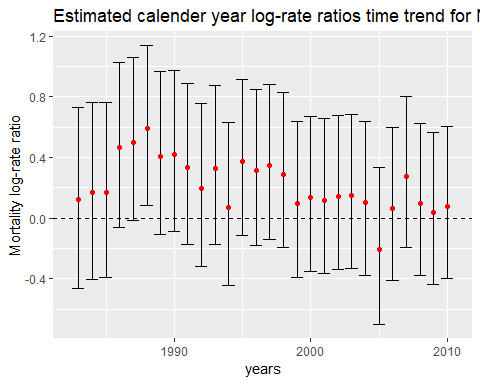
##   
## Call:  
## glm(formula = nonchd ~ as.factor(yearg) + as.factor(ageg) + sexg +   
## area, family = poisson(link = "log"), data = frdeaths2, offset = log(followupyears))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.0034 -0.8637 -0.1965 0.5821 2.4264   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.61461 0.25308 -26.137 < 2e-16 \*\*\*  
## as.factor(yearg)2 0.12188 0.30183 0.404 0.6864   
## as.factor(yearg)3 0.16744 0.29524 0.567 0.5706   
## as.factor(yearg)4 0.17167 0.29163 0.589 0.5561   
## as.factor(yearg)5 0.46366 0.27508 1.686 0.0919 .   
## as.factor(yearg)6 0.49898 0.27192 1.835 0.0665 .   
## as.factor(yearg)7 0.58842 0.26728 2.202 0.0277 \*   
## as.factor(yearg)8 0.40919 0.27159 1.507 0.1319   
## as.factor(yearg)9 0.41802 0.26899 1.554 0.1202   
## as.factor(yearg)10 0.33609 0.26997 1.245 0.2132   
## as.factor(yearg)11 0.19706 0.27324 0.721 0.4708   
## as.factor(yearg)12 0.32751 0.26688 1.227 0.2198   
## as.factor(yearg)13 0.07222 0.27298 0.265 0.7913   
## as.factor(yearg)14 0.37528 0.26103 1.438 0.1505   
## as.factor(yearg)15 0.31159 0.26099 1.194 0.2325   
## as.factor(yearg)16 0.34554 0.25896 1.334 0.1821   
## as.factor(yearg)17 0.28883 0.25921 1.114 0.2652   
## as.factor(yearg)18 0.09806 0.26243 0.374 0.7086   
## as.factor(yearg)19 0.13555 0.25991 0.522 0.6020   
## as.factor(yearg)20 0.11800 0.25897 0.456 0.6486   
## as.factor(yearg)21 0.14169 0.25761 0.550 0.5823   
## as.factor(yearg)22 0.14994 0.25666 0.584 0.5591   
## as.factor(yearg)23 0.10362 0.25642 0.404 0.6861   
## as.factor(yearg)24 -0.20934 0.26228 -0.798 0.4248   
## as.factor(yearg)25 0.06580 0.25525 0.258 0.7966   
## as.factor(yearg)26 0.27533 0.25137 1.095 0.2734   
## as.factor(yearg)27 0.09355 0.25407 0.368 0.7127   
## as.factor(yearg)28 0.03805 0.25469 0.149 0.8812   
## as.factor(yearg)29 0.07436 0.25394 0.293 0.7696   
## as.factor(ageg)2 0.84798 0.14735 5.755 8.68e-09 \*\*\*  
## as.factor(ageg)3 1.68696 0.13508 12.489 < 2e-16 \*\*\*  
## as.factor(ageg)4 2.19207 0.14158 15.483 < 2e-16 \*\*\*  
## as.factor(ageg)5 2.67939 0.14425 18.575 < 2e-16 \*\*\*  
## as.factor(ageg)6 3.32023 0.14585 22.765 < 2e-16 \*\*\*  
## as.factor(ageg)7 3.75602 0.15197 24.716 < 2e-16 \*\*\*  
## as.factor(ageg)8 4.38135 0.16265 26.937 < 2e-16 \*\*\*  
## sexg -0.58154 0.04507 -12.903 < 2e-16 \*\*\*  
## area -0.05270 0.04604 -1.145 0.2523   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 3099.44 on 719 degrees of freedom  
## Residual deviance: 755.11 on 682 degrees of freedom  
## AIC: 2459.1  
##   
## Number of Fisher Scoring iterations: 5

We see that, apart from one, none of the coefficients of the calender years are statistically significant. This already suggests that there shoud not be any downward trend.

We plot the non-CHD mortality model calender year coefficient trend with confidence intervals and a horizontal line for value of the intercept (the mortality log rate-ratio of the first calender year - the reference group). We use the same plot\_year\_trend() function to do this below.

heading ="Estimated calender year log-rate ratios time trend for Non-CHD deaths"  
ylab = "Mortality log-rate ratio"  
plot\_year\_trend(model\_nonchd, heading, ylab)

## Waiting for profiling to be done...



We see in this graph that, apart from one, all the confidence intervals intercept the zero line, suggesting that none of them as statistically significantly different from zero. Hence, we do not observe any downward trend, as was in the CHD case, above and beyond some statistical noise. This also makes sense as our estimates for the log-rate ratios of the calender years were not statistically significant.

## Question 4 c: CHD Mortality Log-Rate Analysis

To plot the CHD mortality log-rate trend over calendar time in terms of the estimated baseline, instead of the log-rate ratios, we need to not have any intercept which would refer to a reference group’s rate ratio. Hence, by simply forcing the intercept to be zero in the glm, the coefficients of the model will represent the mortality log-rates of the calendar years.

model\_chd\_mortality\_rate <- glm(chd ~ as.factor(yearg) + as.factor(ageg) + sexg + area -1, offset=log(followupyears), data=frdeaths2, family=poisson(link='log'))  
summary(model\_chd\_mortality\_rate)

##   
## Call:  
## glm(formula = chd ~ as.factor(yearg) + as.factor(ageg) + sexg +   
## area - 1, family = poisson(link = "log"), data = frdeaths2,   
## offset = log(followupyears))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4231 -0.8383 -0.2598 0.3385 2.5968   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## as.factor(yearg)1 -8.64272 0.55810 -15.486 < 2e-16 \*\*\*  
## as.factor(yearg)2 -8.72915 0.55138 -15.831 < 2e-16 \*\*\*  
## as.factor(yearg)3 -9.04910 0.56578 -15.994 < 2e-16 \*\*\*  
## as.factor(yearg)4 -9.15789 0.56710 -16.149 < 2e-16 \*\*\*  
## as.factor(yearg)5 -8.82531 0.54420 -16.217 < 2e-16 \*\*\*  
## as.factor(yearg)6 -9.19650 0.56051 -16.407 < 2e-16 \*\*\*  
## as.factor(yearg)7 -8.76454 0.53724 -16.314 < 2e-16 \*\*\*  
## as.factor(yearg)8 -8.63395 0.53005 -16.289 < 2e-16 \*\*\*  
## as.factor(yearg)9 -8.92605 0.53655 -16.636 < 2e-16 \*\*\*  
## as.factor(yearg)10 -8.89425 0.53319 -16.681 < 2e-16 \*\*\*  
## as.factor(yearg)11 -9.24469 0.54290 -17.028 < 2e-16 \*\*\*  
## as.factor(yearg)12 -9.29562 0.54216 -17.145 < 2e-16 \*\*\*  
## as.factor(yearg)13 -9.27046 0.53772 -17.240 < 2e-16 \*\*\*  
## as.factor(yearg)14 -9.38499 0.53839 -17.432 < 2e-16 \*\*\*  
## as.factor(yearg)15 -9.36206 0.53521 -17.492 < 2e-16 \*\*\*  
## as.factor(yearg)16 -9.60172 0.54080 -17.755 < 2e-16 \*\*\*  
## as.factor(yearg)17 -9.42593 0.53393 -17.654 < 2e-16 \*\*\*  
## as.factor(yearg)18 -9.42130 0.53172 -17.719 < 2e-16 \*\*\*  
## as.factor(yearg)19 -9.68296 0.53624 -18.057 < 2e-16 \*\*\*  
## as.factor(yearg)20 -9.90008 0.54033 -18.322 < 2e-16 \*\*\*  
## as.factor(yearg)21 -9.55661 0.53030 -18.021 < 2e-16 \*\*\*  
## as.factor(yearg)22 -9.79347 0.53446 -18.324 < 2e-16 \*\*\*  
## as.factor(yearg)23 -9.84517 0.53350 -18.454 < 2e-16 \*\*\*  
## as.factor(yearg)24 -9.80792 0.53085 -18.476 < 2e-16 \*\*\*  
## as.factor(yearg)25 -10.09434 0.53572 -18.843 < 2e-16 \*\*\*  
## as.factor(yearg)26 -10.05838 0.53381 -18.842 < 2e-16 \*\*\*  
## as.factor(yearg)27 -10.37922 0.54120 -19.178 < 2e-16 \*\*\*  
## as.factor(yearg)28 -10.40527 0.54077 -19.241 < 2e-16 \*\*\*  
## as.factor(yearg)29 -10.00731 0.53111 -18.842 < 2e-16 \*\*\*  
## as.factor(ageg)2 2.68914 0.51999 5.172 2.32e-07 \*\*\*  
## as.factor(ageg)3 3.87212 0.50680 7.640 2.17e-14 \*\*\*  
## as.factor(ageg)4 4.59519 0.51120 8.989 < 2e-16 \*\*\*  
## as.factor(ageg)5 5.33789 0.51275 10.410 < 2e-16 \*\*\*  
## as.factor(ageg)6 6.00011 0.51538 11.642 < 2e-16 \*\*\*  
## as.factor(ageg)7 6.60886 0.51960 12.719 < 2e-16 \*\*\*  
## as.factor(ageg)8 7.47255 0.52761 14.163 < 2e-16 \*\*\*  
## sexg -0.94819 0.07082 -13.389 < 2e-16 \*\*\*  
## area -0.16039 0.07095 -2.261 0.0238 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 444789.59 on 720 degrees of freedom  
## Residual deviance: 609.07 on 682 degrees of freedom  
## AIC: 1624.4  
##   
## Number of Fisher Scoring iterations: 6

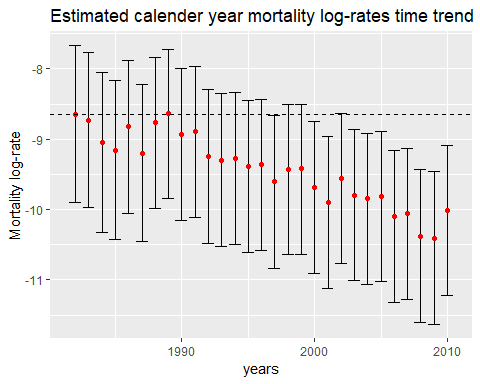
We modify the plot\_year\_trend function to plot the mortality trend over calendar time in terms of the estimated mortality log-rates with Ci’s.

# Defining function that will plot the estimated   
# calendar time (log rate-ratios) with their CI's.  
  
  
plot\_year\_trend\_rates = function(model, title, ylab){  
   
 # Extract variables from model  
 model\_year\_coef = coef(model)[1:29]  
 confints\_all = as.data.frame(confint(model))  
 confints\_calender\_years = confints\_all[1: 29,]  
 u\_ci = confints\_calender\_years[, 2]  
 l\_ci = confints\_calender\_years[, 1]  
   
 # Now we will create a dataframe for ggplot  
 df = data.frame(years = seq(1982, length.out =29),   
 model\_year\_coef,   
 l\_ci, u\_ci)  
   
 # Now we plot the coefficients with their CI's  
 ggplot(df, aes(x = years, y = model\_year\_coef)) +  
 geom\_errorbar(aes(ymax = u\_ci, ymin = l\_ci)) +  
 geom\_point(col = 2) +   
 geom\_hline(yintercept=model\_year\_coef[1], linetype="dashed") +  
 ggtitle(title) + labs(y = ylab)  
}  
  
# Example  
# heading ="Calender year estimated time trend (log-rate ratios)"  
# plot\_year\_trend(model\_chd, heading)

We now plot this using the plot\_year\_trend\_rates() function below.

heading ="Estimated calender year mortality log-rates time trend CHD."  
yaxis = "Mortality log-rate"  
plot\_year\_trend\_rates(model\_chd\_mortality\_rate, heading, ylab=yaxis)

## Waiting for profiling to be done...



From the above plot we see a similar decreasing trend as when we plotted the log-rate ratio over the calender years before. The horizontal line is drawn accross the estimted mortality log-rate of the reference year 1982.