Tutorial 6: Graduation II

Yubo Rasmussen

## Tutorial 6: Graduation II

### Question 1

Based on HW, 2006. The table gives the central exposed to risk, , the number ofdeaths, , and for a set of male assured lives, , for ages . Here is the usual maximum likelihood estimate of the force of mortality at age .

| Age |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 60 | 52250 | 314 | -5.11 | 314.8 | -0.044 |
| 61 | 47400 | 312 | -5.02 | 321.1 | -0.506 |
| 62 | 44860 | 352 | -4.85 | 341.7 | 0.559 |
| 63 | 42920 | 334 | -4.86 | 367.5 | -1.749 |
| 64 | 38920 | 405 | -4.57 | 374.7 | 1.564 |
| 65 | 24950 | 257 | -4.58 | 270.1 | -0.797 |
| 66 | 16900 | 206 | -4.26 | 205.7 | 0.087 |
| 67 | 14990 | 212 | -4.26 | 205.1 | 0.478 |
| 68 | 13550 | 235 | -4.05 | 208.5 | 1.835 |
| 69 | 12160 | 223 | -4.00 | 213.0 | 0.807 |
| 70 | 10950 | 200 | -4.00 | 213.0 | -0.807 |
| 71 | 9770 | 217 | -3.81 | 213.7 | 0.226 |
| 72 | 8770 | 195 | -3.81 | 215.7 | -1.408 |

The table is graduated by fitting a Poisson generalised linear model with a graduation of the form . The table also shows the fitted number of deaths, , and the Pearson residuals, , from this model. You are to use R to answer this question.

#### Part (a)

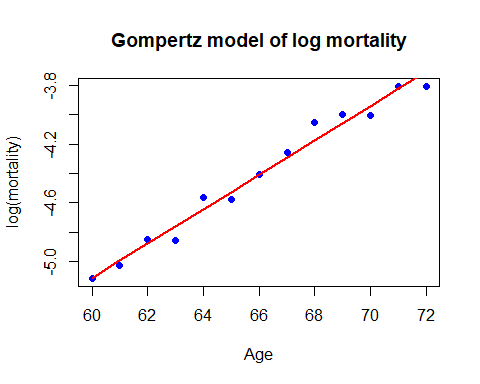
Enter the ages, deaths and central exposures into R and hence obtain a plot against age.

# Enter data and calculate log(mu\_x)  
#  
Age <- 60:72  
Dth <- c(314,312,352,334,405, 257,206,212,235,223, 200,217,195)  
Exp <- c(5225,4740,4486,4292,3892, 2495,1690,1499,1355,1216,  
 1095,977,877)\*10  
Obs <- log(Dth/Exp)

#### Part (b)

Use the glm( ) function to fit the model and add the fitted log mortality to the plot.

Fit.glm <- glm(Dth ~ offset(log(Exp)) + Age, family = "poisson")  
  
# Obtain plot  
#  
par(mfrow = c(1,1))  
plot(Age, Obs, main = "Gompertz model of log mortality",  
 ylab = "log(mortality)", col = "blue", pch = 16)  
lines(Age, Fit.glm$linear - log(Exp), col = "red", lwd = 2)



#### Part (c)

Verify that the graduated deaths, , and the Pearson residuals, , are as shown in the table. Comment on the success of the graduation.

Z.x <- (Dth - Fit.glm$fit)/sqrt(Fit.glm$fit)  
cbind(Age, Exp, Dth, round(Obs, digits = 2),  
 round(Fit.glm$fit, digits = 1), round(Z.x, digits = 3))

Age Exp Dth   
1 60 52250 314 -5.11 314.8 -0.044  
2 61 47400 312 -5.02 321.1 -0.506  
3 62 44860 352 -4.85 341.7 0.559  
4 63 42920 334 -4.86 367.5 -1.749  
5 64 38920 405 -4.57 374.7 1.564  
6 65 24950 257 -4.58 270.1 -0.797  
7 66 16900 206 -4.41 205.7 0.020  
8 67 14990 212 -4.26 205.1 0.478  
9 68 13550 235 -4.05 208.5 1.835  
10 69 12160 223 -4.00 210.4 0.870  
11 70 10950 200 -4.00 213.0 -0.891  
12 71 9770 217 -3.81 213.7 0.226  
13 72 8770 195 -3.81 215.7 -1.408

# Comment  
# Graduation seems reasonable based on the plot in (b). We can check residuals to confirm this in part (d)

#### Part (d)

Test the suitability of the graduation with (i) the -test (ii) the standardised deviations test (use four equal area cells for the test) (iii) the sign test (iv) the change of signs test (v) the runs test, and (vi) the serial correlation test. What is your conclusion?

# Tests  
#  
# Hand calculation and solution with Test\_GoF.r are shown.  
#  
source("Test\_GoF.r")  
#  
# X^2 test: 13 ages, 2 fitted parameters => 11 df  
#  
Chis2 <- sum(Z.x^2); Chis2

[1] 13.89245

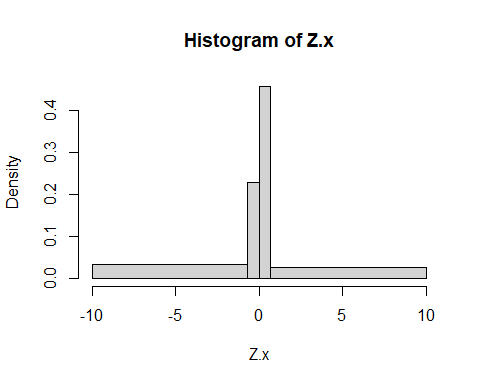
Sig.Pr <- 1 - pchisq(Chis2, 11); Sig.Pr

[1] 0.2390019

#  
Chi.Square(Z.x, 2)

$Chis2  
[1] 13.89245  
  
$DF  
[1] 11  
  
$Sig.Pr  
[1] 0.2390019

#  
# Conc: Sig Pr = 24% so no evidence against the graduation  
#  
# ========================================================  
#  
# Standardised deviations test: 4 cells => 3 df  
#  
Expected <- length(Age)/4  
Boundary <- c(-10, qnorm(c(0.25, 0.5, 0.75)), 10)  
Observed <- hist(Z.x, breaks = Boundary)$counts



cbind(Observed, Expected)

Observed Expected  
[1,] 4 3.25  
[2,] 2 3.25  
[3,] 4 3.25  
[4,] 3 3.25

Chis2 <- sum( (Observed - Expected)^2/Expected); Chis2

[1] 0.8461538

Sig.Pr <- 1 - pchisq(Chis2, 3); Sig.Pr

[1] 0.8383987

#  
Standard.Area(Z.x, 4)

$Boundary  
[1] -0.674 0.000 0.674  
  
$Obs  
[1] 4 2 4 3  
  
$Exp  
[1] 3.25  
  
$DF  
[1] 3  
  
$Chis2  
[1] 0.8461538  
  
$Sig.Pr  
[1] 0.8383987

#  
# Conc: Sig Pr = 84% so a very close fit (as is obvious from the  
# output of cbind( )  
#  
# ========================================================  
#  
# Sign test: Null dist\_n is B(13, 1/2)  
#  
# You can use logical expressions to count positives  
#  
N.positive <- sum(Z.x > 0); N.positive

[1] 7

#  
Sign(Z.x)

$N.plus  
[1] 7  
  
$N.minus  
[1] 6  
  
$Sig.Prob  
[1] 1

#  
# Conc: No formal test required since we have observed very close  
# to E(S) under the null hypothesis.  
#  
# ========================================================  
#  
# Change of sign test: Null dist\_n is B(12, 1/2)  
#  
Change <- 0  
for(i in 1:(length(Age) - 1)) {  
 if(Z.x[i] \* Z.x[i+1] < 0) Change <- Change + 1  
}  
Change

[1] 8

#  
Change.Sign(Z.x)

$N  
[1] 13  
  
$Change  
[1] 8  
  
$Sig.Pr  
[1] 0.927002

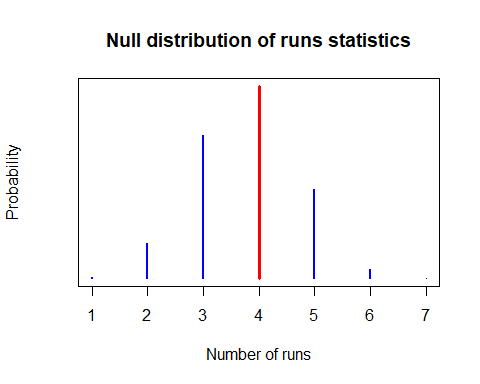
#  
# We are looking for a small number of changes and we have observed  
# more than E(C) so no formal test required. There is no evidence  
# against the graduation.  
#  
# ========================================================  
#  
# Runs test  
#  
# We use the functions discussed in lectures  
#  
#  
# By inspection we have n1 = 7, n2 = 6 and g = 4. Using the formula on  
# p82 we find  
#  
Sig.Prob <- Runs(7, 6, 4); Sig.Prob

[1] 0.7913753

#  
# Or use the function Runs.test( )  
#  
Runs.test(Z.x)

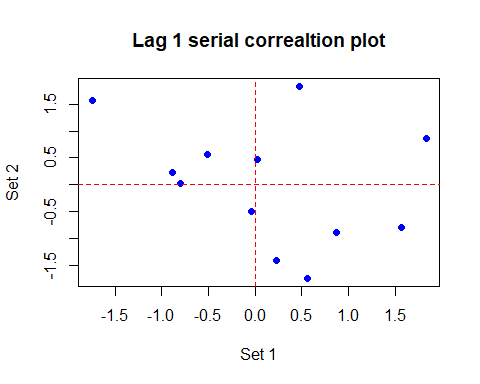
$n1  
[1] 7  
  
$n2  
[1] 6  
  
$g  
[1] 4  
  
$Sig.Prob  
[1] 0.7913753

#  
# to get the same answer. Or use the permutation form of the test  
#  
Runs.Test.Perm(Z.x, 2000)



$Runs  
[1] 4  
  
$Null.dist  
Null.dist  
 1 2 3 4 5 6 7   
0.0020 0.0740 0.3055 0.4095 0.1895 0.0190 0.0005   
  
$Sig.Pr  
[1] 0.791

#  
# We are looking for small values of g; we have found no evidence against  
# the graduation.  
#  
# ========================================================  
#  
# Serial correlation test  
#  
Set.1 <- Z.x[-length(Z.x)]; Set.2 <- Z.x[-1]  
plot(Set.1, Set.2, main = "Lag 1 serial correaltion plot",  
 xlab = "Set 1", ylab = "Set 2", col = "blue", pch = 16)  
abline(h = 0, col = "red", lty = 2)  
abline(v = 0, col = "red", lty = 2)



Corr <- cor(Set.1, Set.2); Corr

[1] -0.3196728

#  
# which we confirm with  
#  
Serial(Z.x)

$Serial  
[1] -0.3196728  
  
$Sig.Pr  
[1] 0.8659349

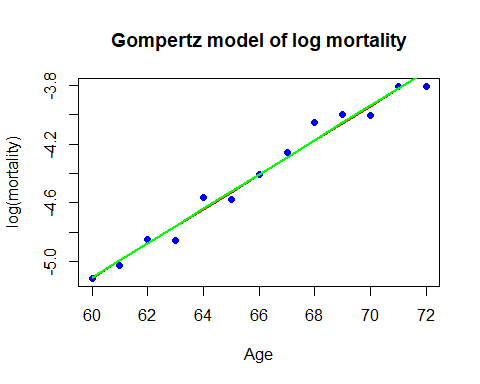
#  
# The plot suggests that the residuals are negatively correlated  
# and the test confirms this (r = -0.32). Large positive values of r  
# provide evidence against the graduation so nothing to worry about here.  
#  
# CONC: Graduation is fine!  
#

#### Part (e)

An actuary suggests graduating the table by minimising

Explain briefly the role of the weights in the above expression and use the -method to suggest suitable values for the weight function . Fit the model in R and compare the fitted model to that obtained in (b).

# (e)  
#  
# Weighted least squares  
#  
# The weights, w\_x take account of the differing reliability of the  
# observed values of log mu\_x. On p61 we saw that  
#  
# Var(log mu\_x) = 1/d\_x  
#  
# so the correct value of w\_x is d\_x. The values of log mu\_x are  
# in Obs to the code with lm( ) is  
#  
Fit.lm <- lm(Obs ~ Age, weights = Dth)  
#  
# and plot to compare both models is  
#  
plot(Age, Obs, main = "Gompertz model of log mortality",  
 ylab = "log(mortality)", col = "blue", pch = 16)  
lines(Age, Fit.glm$linear - log(Exp), col = "red", lwd = 2)  
lines(Age, Fit.lm$fit, col = "green", lwd = 2)



#  
# and the two fitted lines are indistinguishable  
#  
cbind(Fit.glm$coef, Fit.lm$coef)

[,1] [,2]  
(Intercept) -12.1449297 -12.155083  
Age 0.1172168 0.117403

#  
# with very small differences in the fitted coefficients

### Question 2

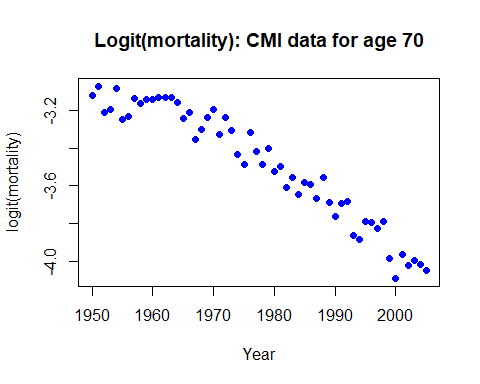
Obtain the deaths, , , and central exposures, , for age 70 from the CMI data files. Convert the central exposures to initial exposures, . Let be the maximum likelihood estimates of .

# URL for CMI\_read.r on github  
CMI\_Deaths\_url <- "https://raw.githubusercontent.com/yubae-bit/F79SU/main/CMI%20and%20HMD%20data%20sets/CMI\_Deaths.csv"  
CMI\_Exposures\_url <- "https://raw.githubusercontent.com/yubae-bit/F79SU/main/CMI%20and%20HMD%20data%20sets/CMI\_Exposures.csv"  
CMI\_Read\_url <- "https://raw.githubusercontent.com/yubae-bit/F79SU/main/CMI%20and%20HMD%20data%20sets/CMI\_read.r"  
  
# Download the necessary files  
download.file(CMI\_Deaths\_url, destfile = "CMI\_Deaths.csv", mode = "wb")  
download.file(CMI\_Exposures\_url, destfile = "CMI\_Exposures.csv", mode = "wb")  
download.file(CMI\_Read\_url, destfile = "CMI\_read.r", mode = "wb")  
  
# Now source the R script  
source("CMI\_read.r")  
  
# Alternatively to running the code above, you can simply add the files above to the same working directory as this file.  
  
# Select age 70 data and obtain Q.x and its logit  
#  
D.70 <- Dth[Age == 70, ]  
E.70 <- Exp[Age == 70, ]  
E.init <- E.70 + D.70/2  
Logit <- function(x) log(x/(1-x))  
Q.t <- D.70/E.init  
Obs <- Logit(Q.t)  
#

#### Part (a)

Plot against .

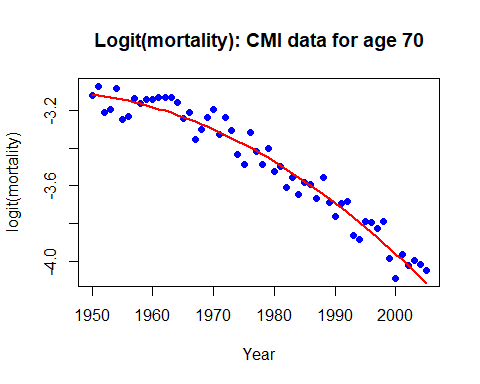
# (a) Plot the logits  
#  
plot(Year, Obs, ylab = "logit(mortality)", col = "blue", pch = 16,  
 main = "Logit(mortality): CMI data for age 70")



#### Part (b)

Fit the generalised linear model for on with quadratic predictor, logit link and binomial error. Obtain the fitted logits, , add the fitted curve to your plot in (a) and comment informally on the quality of the fit.

# (b) Fit model and calculate Pearson residuals  
#  
# First, centre the year values  
#  
Yr <- Year - mean(Year)  
Yr2 <- Yr^2  
Fit.glm <- glm(Q.t ~ Yr + Yr2, weights = E.init, family = "binomial")  
#  
# Add fitted line to plot  
#  
plot(Year, Obs, ylab = "logit(mortality)", col = "blue", pch = 16,  
 main = "Logit(mortality): CMI data for age 70")  
lines(Year, Fit.glm$lin, col = "red", lwd = 2)



#  
# Plot looks pretty good!

#### Part (c)

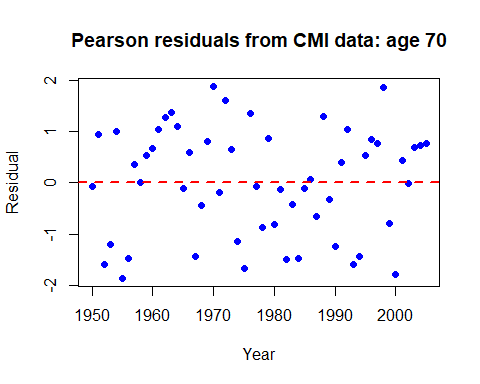
Calculate the fitted deaths, .

Q.fit <- exp(Fit.glm$lin)/(1 + exp(Fit.glm$lin))  
D.fit <- E.init \* Q.fit

#### Part (d)

Calculate the Pearson residuals, , and plot them against . Again, comment on the quality of the fit to the data.

Z.t <- (D.70 - D.fit)/sqrt(D.fit)  
#  
# Plot residuals  
#  
plot(Year, Z.t, ylab = "Residual", col = "blue", pch = 16,  
 main = "Pearson residuals from CMI data: age 70")  
abline(h = 0, col = "red", lty = 2, lwd = 2)



#  
# Again, this looks fine - no large Z.x and no pattern either.  
#  
# Check results so far  
#  
cbind(E.init, D.70, round(Q.t, dig = 3), round(Q.fit, dig = 3),  
 round(Z.t, digits = 2))

E.init D.70   
1950 10452.0 443 0.042 0.043 -0.08  
1951 10023.0 444 0.044 0.042 0.95  
1952 9586.0 372 0.039 0.042 -1.59  
1953 9388.0 370 0.039 0.042 -1.19  
1954 9132.0 400 0.044 0.042 0.99  
1955 8799.0 329 0.037 0.041 -1.86  
1956 8545.0 324 0.038 0.041 -1.47  
1957 8403.5 350 0.042 0.041 0.37  
1958 8261.0 335 0.041 0.041 0.01  
1959 8043.5 333 0.041 0.040 0.54  
1960 7931.5 328 0.041 0.040 0.67  
1961 7914.0 331 0.042 0.039 1.05  
1962 7845.0 329 0.042 0.039 1.27  
1963 7660.0 320 0.042 0.039 1.37  
1964 7458.0 304 0.041 0.038 1.09  
1965 7471.5 281 0.038 0.038 -0.11  
1966 7672.0 297 0.039 0.037 0.58  
1967 7488.0 253 0.034 0.037 -1.43  
1968 7427.5 264 0.036 0.037 -0.43  
1969 7788.0 294 0.038 0.036 0.80  
1970 8240.5 325 0.039 0.036 1.88  
1971 8364.0 290 0.035 0.035 -0.18  
1972 8527.0 322 0.038 0.035 1.61  
1973 8495.5 300 0.035 0.034 0.65  
1974 8550.0 267 0.031 0.033 -1.14  
1975 8810.1 262 0.030 0.033 -1.66  
1976 9073.5 317 0.035 0.032 1.34  
1977 9110.0 289 0.032 0.032 -0.07  
1978 9264.5 275 0.030 0.031 -0.87  
1979 9386.0 303 0.032 0.031 0.86  
1980 9336.2 268 0.029 0.030 -0.81  
1981 9236.4 271 0.029 0.030 -0.14  
1982 9251.4 244 0.026 0.029 -1.49  
1983 9915.6 275 0.028 0.028 -0.41  
1984 10116.4 257 0.025 0.028 -1.47  
1985 10745.0 291 0.027 0.027 -0.11  
1986 10599.7 284 0.027 0.027 0.07  
1987 8853.1 221 0.025 0.026 -0.66  
1988 8614.0 239 0.028 0.026 1.30  
1989 10963.4 268 0.024 0.025 -0.32  
1990 13963.5 317 0.023 0.024 -1.24  
1991 13884.2 337 0.024 0.024 0.40  
1992 13367.9 328 0.025 0.023 1.04  
1993 13301.4 273 0.021 0.023 -1.59  
1994 13194.2 266 0.020 0.022 -1.44  
1995 13292.2 294 0.022 0.021 0.53  
1996 11715.4 258 0.022 0.021 0.85  
1997 12006.2 256 0.021 0.020 0.77  
1998 11964.8 265 0.022 0.020 1.86  
1999 12450.9 227 0.018 0.019 -0.79  
2000 12158.9 200 0.016 0.019 -1.79  
2001 12826.0 239 0.019 0.018 0.43  
2002 11613.0 204 0.018 0.018 -0.02  
2003 9159.0 165 0.018 0.017 0.70  
2004 7017.0 124 0.018 0.017 0.73  
2005 7887.5 135 0.017 0.016 0.76

#### Part (e)

Test the suitability of the graduation with (i) the $χ$2-test (ii) the standardised deviations test (use ten equal area cells for the test) (iii) the sign test (iv) the change of signs test (v) the runs test, and (vi) the serial correlation test. What is your conclusion?

# (e) Tests  
#  
# Load file of tests  
#  
source("Test\_GoF.r")  
#  
# X^2 test: 56 years, 3 fitted parameters => 53 df  
#  
Chi.Square(Z.t, 3)

$Chis2  
[1] 60.75029  
  
$DF  
[1] 53  
  
$Sig.Pr  
[1] 0.2168187

#  
# Conc: Sig Pr = 22% so no evidence against the graduation  
#  
# ========================================================  
#  
# Standardised deviations test: 10 cells => 9 df  
#  
Standard.Area(Z.t, 10)

$Boundary  
[1] -1.282 -0.842 -0.524 -0.253 0.000 0.253 0.524 0.842 1.282  
  
$Obs  
 [1] 10 4 3 3 7 2 3 10 8 6  
  
$Exp  
[1] 5.6  
  
$DF  
[1] 9  
  
$Chis2  
[1] 14.71429  
  
$Sig.Pr  
[1] 0.099089

#  
# Conc: Sig Pr = 10% so not significant  
#  
# ========================================================  
#  
# Sign test: Null dist\_n is B(56, 1/2)  
#  
Sign(Z.t)

$N.plus  
[1] 29  
  
$N.minus  
[1] 27  
  
$Sig.Prob  
[1] 0.8938531

#  
# Conc: No formal test required since we have observed 29 positives,  
# very close to E(S) = 28 under the null hypothesis.  
#  
# ========================================================  
#  
# Change of sign test: Null dist\_n is B(55, 1/2)  
#  
Change.Sign(Z.t)

$N  
[1] 56  
  
$Change  
[1] 27  
  
$Sig.Pr  
[1] 0.5

#  
# We are looking for a small number of changes and we have observed  
# 27 with E(C) = 27.5 so no formal test required. There is no evidence  
# against the graduation.  
#  
# ========================================================  
#  
# Runs test  
#  
Runs.test(Z.t)

$n1  
[1] 29  
  
$n2  
[1] 27  
  
$g  
[1] 14  
  
$Sig.Prob  
[1] 0.5

#  
# We have n1 = 29, n2 = 27 and g = 14 with sig prob = 50%. Since we  
# are looking for small values of g we have found no evidence against  
# the graduation.  
#  
# ========================================================  
#  
# Serial correlation test  
#  
Serial(Z.t)

$Serial  
[1] 0.1171543  
  
$Sig.Pr  
[1] 0.1924674

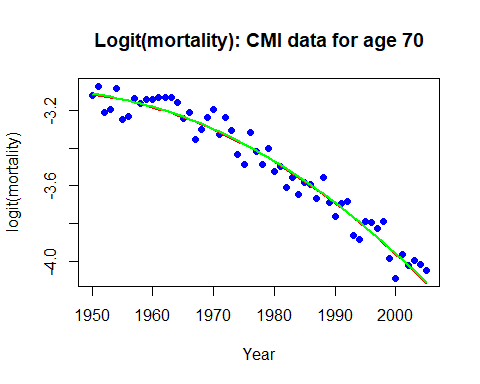
#  
# No sign of serious serial correlation here with r = 0.12. The sig  
# prob is 19% so all is well!  
#  
# CONC: Graduation is fine!

#### Part (f)

An actuary suggests graduating the table by minimising

Explain briefly the role of the weights in the above expression and use the -method to suggest suitable values for the weight function . Fit the model in R and compare the fitted model to that obtained in (b).

# (f) Weighted least squares  
#  
# The weights, w\_x take account of the differing reliability of the  
# observed values of logit q\_x. On p61 we saw that  
#  
# Var(log mu\_x) = 1/d\_x.  
#  
# Here we need Var[ logit( q\_t) ]. We can use the delta method here  
# too and we find (unsurprisingly!) that we have  
#  
# Wt = E\_t q\_t (1 - q\_t) = d\_t approx  
#  
# In other words the weights are the same as in the Poisson model  
#  
Q.t <- D.70/E.init  
Obs <- Logit(Q.t)  
Fit.lm <- lm(Obs ~ Yr + Yr2, weights = D.70)  
#  
# Plot to compare both models is  
#  
plot(Year, Obs, ylab = "logit(mortality)", col = "blue", pch = 16,  
 main = "Logit(mortality): CMI data for age 70")  
lines(Year, Fit.glm$lin, col = "red", lwd = 2)  
lines(Year, Fit.lm$fit, col = "green", lwd = 2)



#  
# and the two fitted lines are indistinguishable  
#  
cbind(Fit.glm$coef, Fit.lm$coef)

[,1] [,2]  
(Intercept) -3.4235186367 -3.4213614094  
Yr -0.0182377684 -0.0182300161  
Yr2 -0.0002538369 -0.0002542883

#  
# with very small differences in the fitted coefficients