SAHIL SHAH - 420

STATISTICAL RISK MODELLING

R CODES WITH OUTPUT, EXPLANATION AND COMMENTS

“data=read.table('Graduation(1).csv' , header=TRUE , sep=',')

data$CRUDE<-data$DEATHS/data$ETR”

Reading data using read.table function with headers as true and separator as “,” after setting the required directory.
Crude Data = deaths/ETR

```
> head(data)
  AGE   ETR DEATHS        CRUDE GRADUATED EXPECTED ZX
1  25 78500     24 0.0003057325         0        0  0
2  26 80425     24 0.0002984147         0        0  0
3  27 81975     24 0.0002927722         0        0  0
4  28 83725     24 0.0002866527         0        0  0
5  29 84875     72 0.0008483063         0        0  0
6  30 85075     48 0.0005642081         0        0  0
```

```
“ gompertz<-lm(log(data$CRUDE)~data$AGE)
gompertz
coef(gompertz)
B=exp(as.numeric(coef(gompertz)))[1]
C=exp(as.numeric(coef(gompertz)))[2]
c(B,C)
data$GRADUATED<-round(B*C^data$AGE,6)
plot(data$AGE,data$CRUDE,
xlab="Age",ylab="Mortality rate",main="Crude and
graduated rates" , pch=10)
lines(data$AGE,data$GRADUATED , col='red') ”
```

Used Gompertz law to fill the graduated column using the lm function.

Used coef function and as.numeric function to find parameters B and C.

Using round function to rounding off the graduated column to 6 decimal palces.

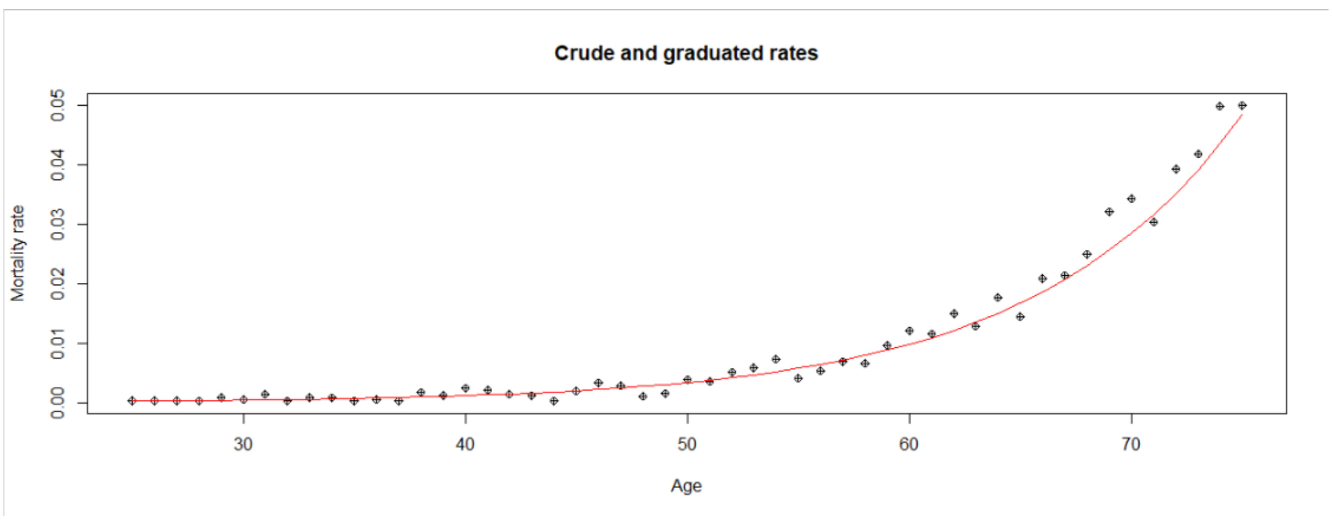
B = 1.668727e-05

C = 1.112153e+00

```
Call:
lm(formula = log(data$CRUDE) ~ data$AGE)

Coefficients:
(Intercept)     data$AGE
   -11.0009       0.1063
```

```
> head(data)
  AGE   ETR DEATHS         CRUDE GRADUATED EXPECTED ZX
1  25 78500     24 0.0003057325  0.000238        0  0
2  26 80425     24 0.0002984147  0.000265        0  0
3  27 81975     24 0.0002927722  0.000294        0  0
4  28 83725     24 0.0002866527  0.000327        0  0
5  29 84875     72 0.0008483063  0.000364        0  0
6  30 85075     48 0.0005642081  0.000405        0  0
```



“diff1<-function(v)v[-1]-v[-length(v)]
diff_crude=round(diff1(diff1(diff1(data$CRUDE)))*10^6,0)
diff_grad=round(diff1(diff1(diff1(data$GRADUATED)))*10^6,0)
plot(data$AGE,data$CRUDE, xlab="Age",ylab="Mortality rate",main="Crude and graduated rates" , pch=10)
lines(data$AGE,data$GRADUATED,col="red")
smoothness_df = data.frame("Age" = data[data$AGE <= 72,"AGE"],diff_crude,diff_grad)
smoothness_df”

Using diff function to check for smoothness by applying 3rd differences to graduated and crude rates.
The criteria used to check for smoothness is generally that the 3rd differences should be small in magnitude and progress regularly.
In the above data frame the 3rd differences of crude rates are much larger in magnitude and progress erratically.
However the magnitude of 3rd differences of the graduated rates are very small and progress regularly.
Thus graduated rates computed by fitting Gompertz Law are smooth and that is exactly what’s desired.

```
> smoothness_df
   Age diff_crude diff_grad
1   25         -2         2
2   26        568         0
3   27      -1414         0
4   28       1973         0
5   29      -3099         2
6   30       3648        -1
7   31      -2233         1
8   32         17         2
9   33       1342        -1
10  34      -1322         2
11  35       2241         1
12  36      -3676         0
13  37       3676         2
14  38      -3183         2
15  39        947         0
16  40       1004         2
17  41      -1142         3
18  42       3333         1
19  43      -3124         3
20  44      -1295         2
21  45        398         2
22  46       3528         5
23  47       -274         2
24  48      -4533         5
25  49       4447         3
26  50      -2588         6
27  51       1433         4
28  52      -5286         7
29  53       9026         7
30  54      -4183         6
31  55      -1951         9
32  56       4992         9
33  57      -3863         9
34  58      -2369        13
35  59       6962        11
36  60      -9542        15
37  61      12421        14
38  62     -14898        19
39  63      17650        17
40  64     -15583        23
41  65       9139        23
42  66        267        25
43  67      -8292        31
44  68      -1303        31
45  69      18882        37
46  70     -19024        40
47  71      11723        45
48  72     -13392        48
```

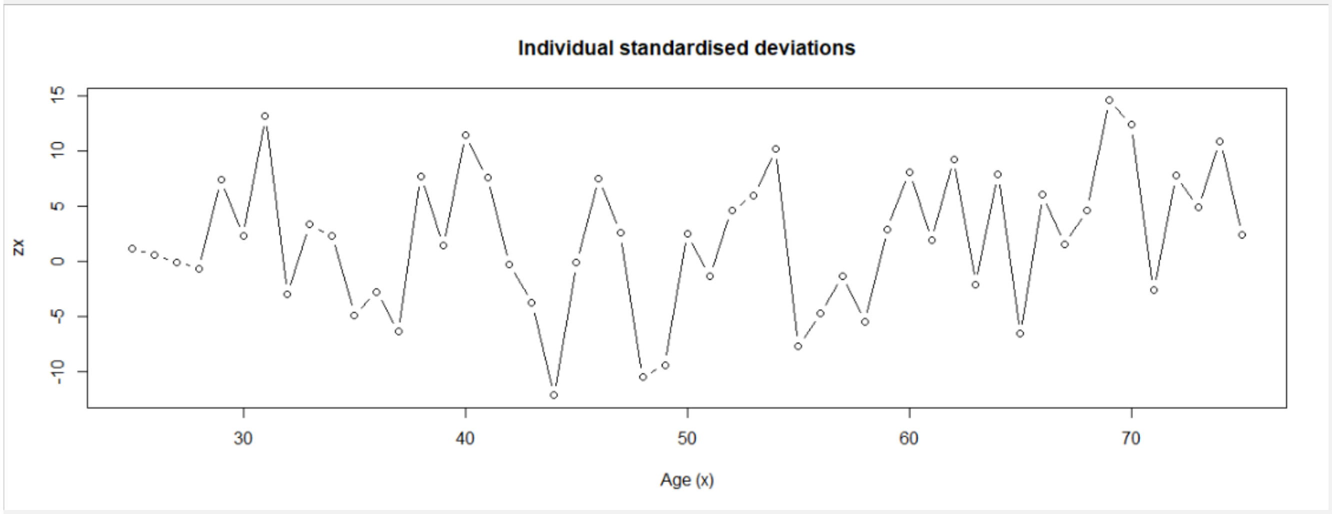
“data$EXPECTED<-
round(data$GRADUATED*data$ETR,2)
data$ZX<-round((data$DEATHS-
data$EXPECTED)/sqrt(data$EXPECTED),2)
head(data)
plot(data$AGE,data$ZX,type="b",xlab="Age (x)",
ylab="zx",main="Individual standardised deviations")
data$prob = data$EXPECTED/sum(data$EXPECTED)
head(data)
a=chisq.test(data$DEATHS,data$prob) ”

Calculated Expected rates and Zx and rounded both the numbers to 2 decimal places using round function.
Then conducted a χ^2 test using chisq.test function.
Looking at the χ^2test we can conclude that the Graduation is not OK as the p.value=0.2774 which is greater than 0.05.
H0 of a χ^2 test is that the the Graduation is OK
H1 states that the graudation is not OK.
As the p.value is greater than 0.05 we have sufficient evidence to reject H0 and hence the graduation is not OK.

```
> head(data)
  AGE   ETR DEATHS        CRUDE GRADUATED EXPECTED    ZX         prob
1  25 78500     24 0.0003057325  0.000238    18.68  1.23 0.0003031505
2  26 80425     24 0.0002984147  0.000265    21.31  0.58 0.0003458318
3  27 81975     24 0.0002927722  0.000294    24.10 -0.02 0.0003911096
4  28 83725     24 0.0002866527  0.000327    27.38 -0.65 0.0004443395
5  29 84875     72 0.0008483063  0.000364    30.89  7.40 0.0005013019
6  30 85075     48 0.0005642081  0.000405    34.46  2.31 0.0005592381
```

```
        Pearson's Chi-squared test

data:  data$DEATHS and data$prob
X-squared = 1734, df = 1700, p-value = 0.2774
```



Deficiencies of the chi-squared test

The χ^2 test will fail to detect several defects that could be of considerable financial importance.

(a) There could be a few large deviations offset by a lot of very small deviations. In other words, the χ^2 test could be satisfied although the data do not satisfy the distributional assumptions that underlie it. This is, in essence, because the χ^2 test statistic summarizes a lot of information in a single figure.

(b) The graduation might be biased above or below the data by a small amount. The χ^2 statistic can often fail to detect consistent bias if it is small, but we should still wish to avoid it.

(c) Even if the graduation is not biased as a whole, there could be significant groups of consecutive ages (called runs or clumps) over which it is biased up or down. This is still to be avoided.

```
“binning_var = c(-Inf, -3, -2, -1, 0, 1, 2, 3, Inf)
testing_table = data.frame(data$ZX, bin=cut(data$ZX, binning_var,
include.lowest=TRUE))
prop.table(table(testing_table$bin))
stand_dev_test = data.frame(Bins =
levels(unique(testing_table$bin)), Expected = c(0, 0.02, 0.14, 0.34,
0.34, 0.14, 0.02, 0), Observed =
round(as.numeric(prop.table(table(testing_table$bin))), 3))
stand_dev_test = stand_dev_test %>% mutate(Expected =
Expected * 51, Observed = Observed * 51)
chisq.test(stand_dev_test$Expected, stand_dev_test$Observerd,
correct = T)”
```

STANDARDISED DEVIATIONS TEST

We can use standardized deviations test to look for the defect (a) of χ^2test.
Created intervals of 1 from -3 to +3 under binning_var.
Then made a testing table and a probability table using prop.table function.
Then made a new dataframe with observed and expected probability in the columns.
Then conducted a χ^2test using chisq.test function.

OVERALL SHAPE - The data shows that it is negatively skewed and has more values on the tail.

ABSOLUTE DEVIATIONS – The data is overgraduated and shows existence of duplicates as the absolute deviations are too big. This is proved by the fact that there are less than 50% values lying in the (-2/3 to 2/3) bin.

OUTLIERS – Too many outliers are present as close to 20% of the data lies in the (-infinity to -3) bin and 40% of the data lies in the (3 to Infinity) bin. Showing that the data has approximately 60% outliers.

SYMMETRY – The number of positive and the number of negative deviations are not close to 50% rather there are only 20 negative and 31 positive deviations. This shows that observed mortality rates do not conform to the model with the rates assumed in the graduation. The data shows nature of discrepancy.

H0 of a χ^2 test is that the the Graduation is OK
H1 states that the graduation is not OK.
χ^2 test gives a p.value of less than 0.05 and as mentioned above as the p.value was very low we have insufficient evidence to reject H0 and therefore the Graduation is OK.

```
> stand_dev_test
       Bins Expected Observed
1 [-Inf,-3]     0.00    9.996
2   (-3,-2]     1.02    3.978
3   (-2,-1]     7.14    1.989
4    (-1,0]    17.34    3.978
5     (0,1]    17.34    1.020
6     (1,2]     7.14    3.978
7     (2,3]     1.02    6.018
8  (3, Inf]     0.00   19.992
```

“data$signs = ifelse(data$GRADUATED>data$CRUDE,1,0)
head(data)
tail(data)
sum(data$signs)
p=2*pbinom(20,51,0.5)”

<u>**SIGNS TEST**</u>

It checks for defect (b) of χ^2 test.

It is a 2 tailed test and therefore 2 is multiplied to the p.value.

H0 : Data is biased

H1 : Data is not biased

As the p.value = 0.16 which is greater than 0.05 we have insufficient evidence to reject H0 . Hence data is not biased.

```
> head(data)
  AGE   ETR DEATHS        CRUDE GRADUATED EXPECTED    ZX         prob signs
1  25 78500     24 0.0003057325  0.000238    18.68  1.23 0.0003031505     0
2  26 80425     24 0.0002984147  0.000265    21.31  0.58 0.0003458318     0
3  27 81975     24 0.0002927722  0.000294    24.10 -0.02 0.0003911096     1
4  28 83725     24 0.0002866527  0.000327    27.38 -0.65 0.0004443395     1
5  29 84875     72 0.0008483063  0.000364    30.89  7.40 0.0005013019     0
6  30 85075     48 0.0005642081  0.000405    34.46  2.31 0.0005592381     0
```

“(sum(data$DEATHS)-sum(data$EXPECTED))/(sqrt(sum(data$EXPECTED)))”

CUMULATIVE DEVIATIONS TEST

It checks cumulations of positive and negative groups.
Test statistic = 18.831
H0: Graduation rates are OK
H1: Graduation rates are too low
As test statistic is greater than 1.96 we have sufficient evidence to reject H0 and conclude that Graduation rates are too low.

“z1=data$ZX[1:length(data$ZX)-1]
z2=data$ZX[2:length(data$ZX)]
a=cor(z1,z2)
a
a*sqrt(51)”

SERIAL CORRELATIONS TEST

It detects clumping of signs of deviations
It is a 1 sided test.
As a = r_j =0.1477 we can conclude that Z_x has similar values.
H0 : No grouping of signs
H1 : Grouping of signs
As test statistic is 1.05 which is less than 1.649 we have insufficient evidence to reject H0. Therefore there is no evidence of grouping of signs .

FINAL OUTPUT OF DATA

	AGE	ETR	DEATHS	CRUDE	GRADUATED	EXPECTED	ZX	prob	signs
1	25	78500	24	0.0003057325	0.000238	18.68	1.23	0.0003031505	0
2	26	80425	24	0.0002984147	0.000265	21.31	0.58	0.0003458318	0
3	27	81975	24	0.0002927722	0.000294	24.10	-0.02	0.0003911096	1
4	28	83725	24	0.0002866527	0.000327	27.38	-0.65	0.0004443395	1
5	29	84875	72	0.0008483063	0.000364	30.89	7.40	0.0005013019	0
6	30	85075	48	0.0005642081	0.000405	34.46	2.31	0.0005592381	0