Statistical Risk Modelling – 1

Project: Semester 3

Done By: Gunavanth Mahendra Duremanthi

1. Calculated crude rates using the formula: $Cruderate = \frac{Deaths}{Exposed\ to\ Risk}$

   First 5 data points:

Age	ETR	Deaths	Crude Rate	Graduated Rate	Expected	Z_x
25	78500	24	0.0003057325	0	0	0
26	80425	24	0.0002984147	0	0	0
27	81975	24	0.0002927722	0	0	0
28	83725	24	0.0002866527	0	0	0
29	84875	72	0.0008483063	0	0	0

Table 1: Data with Crude Rates

2. We know $\mu = Bc^x$

   $\Rightarrow \log \mu = \log B + x * \log c, where\ x\ is\ age$

   Fitting a linear regression model on log μ against x, we will get the values of B and c as:

   $B = e^{intercept} = e^{-11.0008645} = 1.668727e-05$

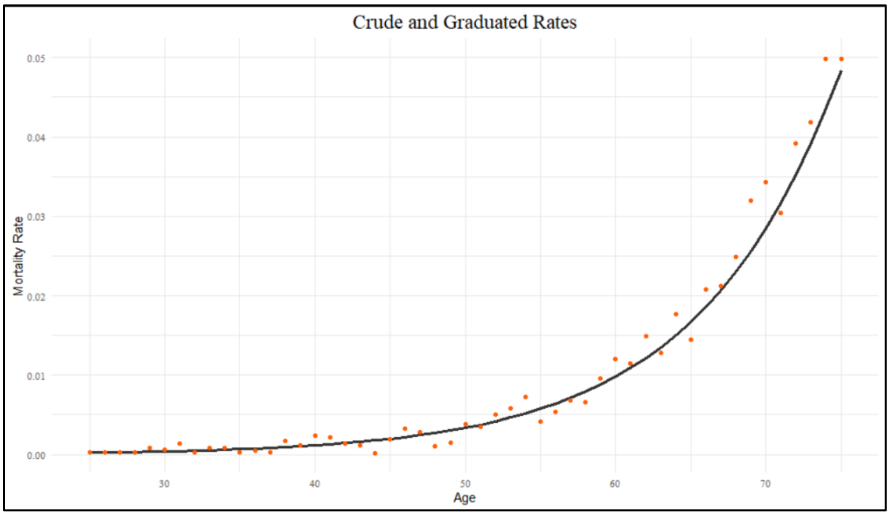
   $c = e^{slope} = e^{0.1062976} = 1.112153$

   Using the Gompertz formula, we estimated the graduated rates for all ages in the data.

   First 5 Data points:

Age	ETR	Deaths	Crude Rate	Graduated Rate	Expected	Z_x
25	78500	24	0.0003057325	0.000238	0	0
26	80425	24	0.0002984147	0.000265	0	0
27	81975	24	0.0002927722	0.000294	0	0
28	83725	24	0.0002866527	0.000327	0	0
29	84875	72	0.0008483063	0.000364	0	0

Table 2: Data with Graduated Rates



Plot 1: Crude vs Graduated Rates

3. After calculating the 3rd degree difference of the Graduated rates, we see that the differences are extremely small compared to the data and they progress regularly. Hence, we can say that the graduated rates are smooth.

   Here is a snapshot of the results:

Graduated Rate	3rd Degree difference
0.000327	2.0e-06
0.000364	0
0.000405	0
0.000450	0

Table 3: Results of test for smoothness.

4. Calculated Zx using the formula $\frac{Actual-Expected}{\sqrt{Expected}}$, and expected = $\mu_x^o * ETR$

   Here is the updated table:

Age	ETR	Deaths	Crude Rate	Graduated Rate	Expected	zx
25	78500	24	0.0003057325	0.000238	18.68300	1.230108
26	80425	24	0.0002984147	0.000265	21.31263	0.582116
27	81975	24	0.0002927722	0.000294	24.10065	-0.020502
28	83725	24	0.0002866527	0.000327	27.37807	-0.645606
29	84875	72	0.0008483063	0.000364	30.89450	7.395361

Table 4: Fully filled out table.

χ^2 test:

H0: The graduated rates are representative of the crude rates.

H1: The graduated rates are not representative of the crude rates.

Degrees of Freedom: 50

P-value: ≈ 0

Conclusion: There is sufficient evidence reject H0 at the 5% level. Hence, we conclude that the graduated rates are not representative of the crude rates.

5.

a. Individual Standardised Deviations Test:

H0: $z_x \sim N(0,1)$ i.e., no excessive deviations present.

H1: Excessive deviations present.

	(-Inf, -3]	(-3, -2]	(-2, -1]	(-1, 0]	(0, 1]	(1, 2]	(2, 3]	(3, Inf)
Expected	0	1.02	7.14	17.34	17.34	7.14	1.02	0
Observed	9.996	3.978	1.989	3.978	1.020	3.978	6.018	19.992

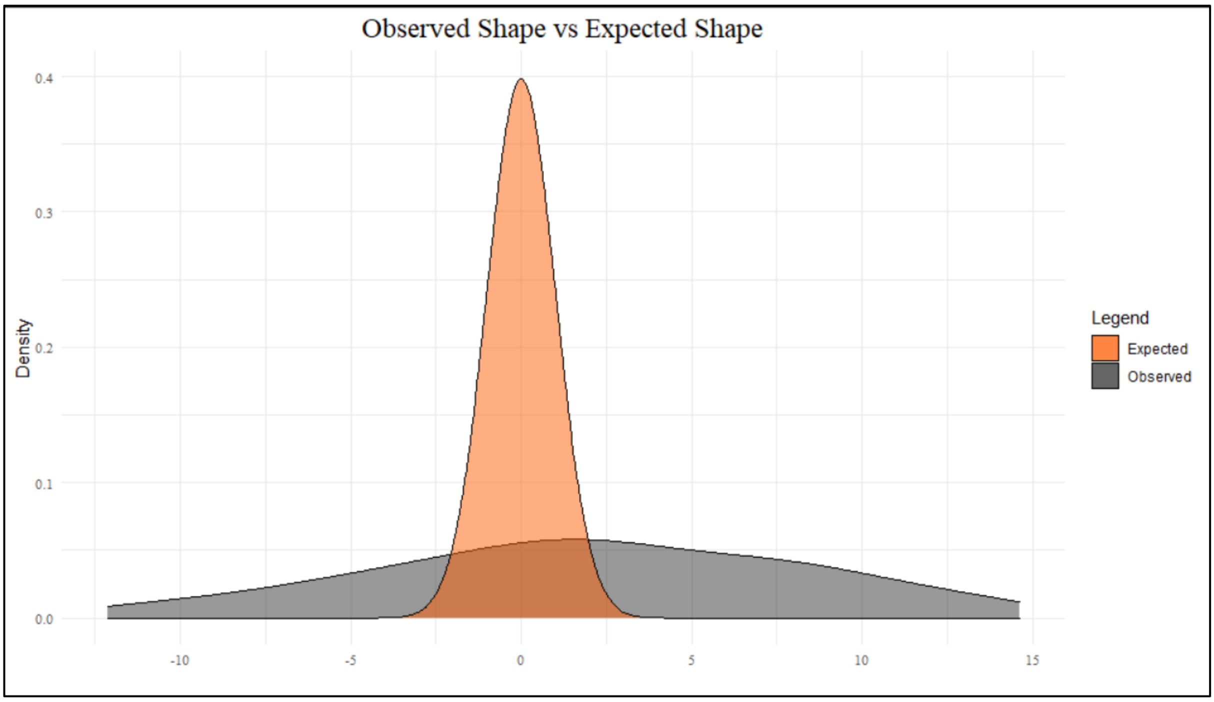
Table 5: Individual Standardised Deviations Test

Clearly, the observed data does not follow a Standard Normal distribution. We run a χ^2 test to formally test this assumption.

P-value: 1.773e-10 ≈ 0

Conclusion: There is sufficient evidence to reject H0 at the 5% level, hence we conclude that excessive deviations do exist.

i.



Plot 2: Observed Shape vs Expected Shape.

Although the observed values appear to have a fairly symmetrical distribution, they do not follow a standard normal distribution.

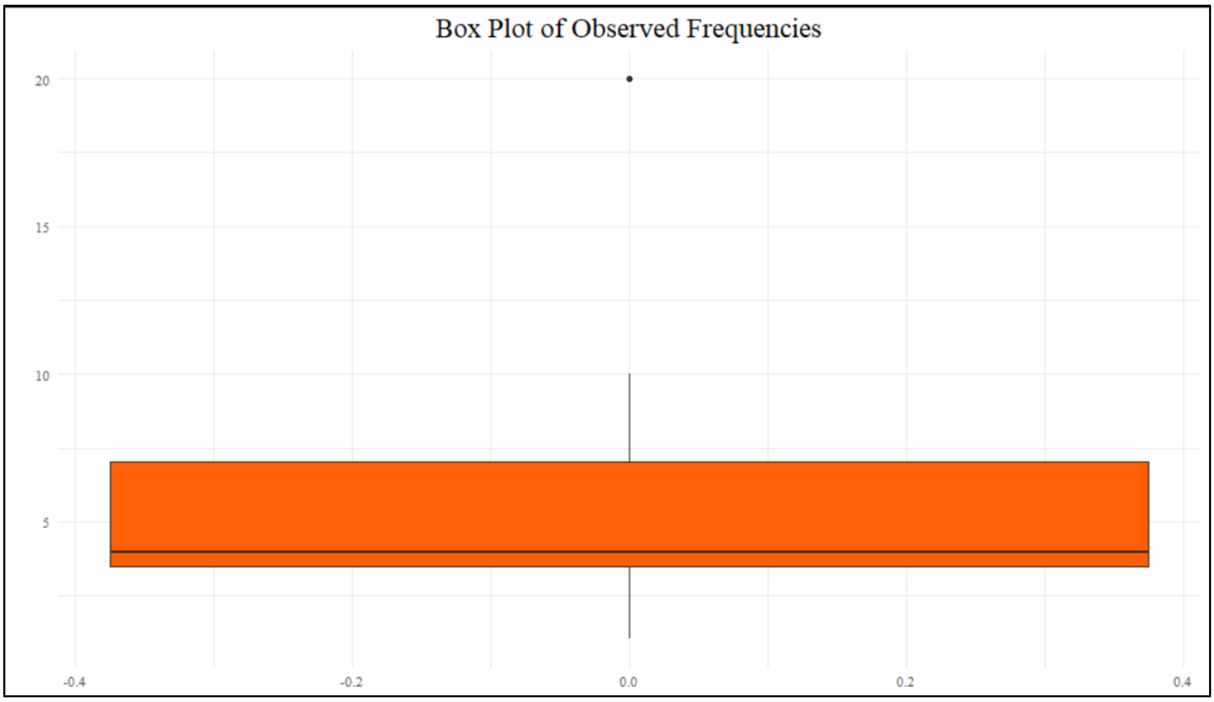
ii.

	(-Inf, -3]	(-3, -2]	(-2, -1]	(-1, 0]	(0, 1]	(1, 2]	(2, 3]	(3, Inf)
Absolute Deviations	9.996	2.958	5.151	13.362	16.320	3.162	4.998	19.992

Table 6: Absolute Deviations

The absolute deviations are high in almost all the groups. The deviations seem to be higher near the mean of the expected distribution and towards the right tail as well.

iii.



Plot 3: Boxplot of Deviations.

We see a significant outlier at the outermost bin i.e., (3, Inf) of 19.992.

iv. As seen in Plot 2, the observed values of z_x follow a fairly symmetric distribution. Albeit about a different mean (2.01343), when compared to the standard normal distribution.

v. <u>Conclusion</u>: There is sufficient evidence to reject H0 at the 5% level. Hence, we conclude that the observed mortality rates do not conform to the model with the rates assumed in the graduation.

b. Signs Test:

H0: There is no bias in the data.

H1: A bias in the data exists.

On performing the signs test, we see that there are a total of 20 values wherein the graduated rates are greater than the crude rates.

P-value: 0.1607796

Conclusion: There is insufficient evidence to reject H0 at the 5% level. Hence, we conclude that no bias exists.

c. Cumulative Deviations Test:

H0: No bias exists in the data.

H1: There is a bias in the data.

Statistic: $\frac{\sum D_x - E_x^c * \mu_x^o}{\sqrt{\sum E_x^c * \mu_x^o}} \sim N(0,1)$

Statistic Value: 18.831

P-Value: 2.105549e-79 ≈ 0

Conclusion: There is sufficient evidence to reject H0 at the 5% level. Hence, we conclude that bias exists.

d. Serial Correlations Test:

H0: There exists a grouping of the signs of the deviations.

H1: There is no grouping of the signs of the deviations.

Statistic: Correlation between the first m-1 values and the sequence with lag 1.

$Statistic\ is: corr(z, z_1) * \sqrt{m}$

Statistic Value: 1.054

P-Value: 0.1458993

Conclusion: There is insufficient evidence to reject H0 at the 5% level. Hence, we conclude that no grouping of data exists.