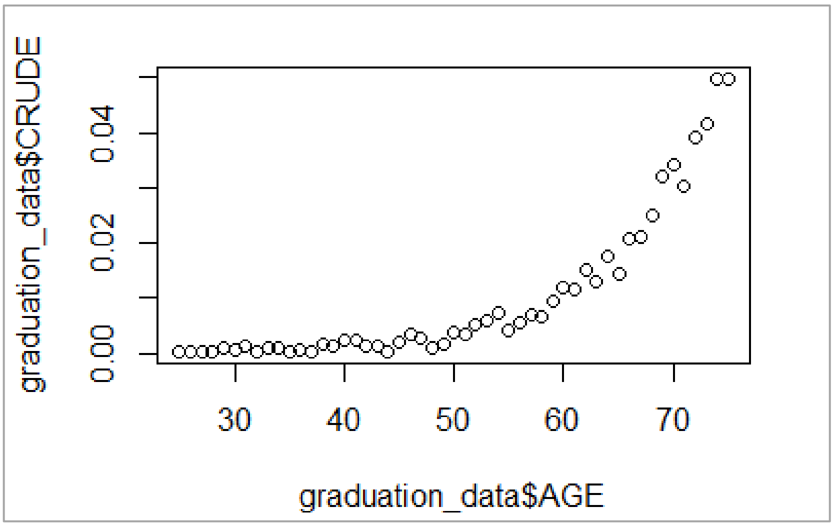
INSTITUTE OF ACTUARIES AND QUANTITATIVE FINANCE

STATISTICAL RISK MODELLING – 1: PROJECT

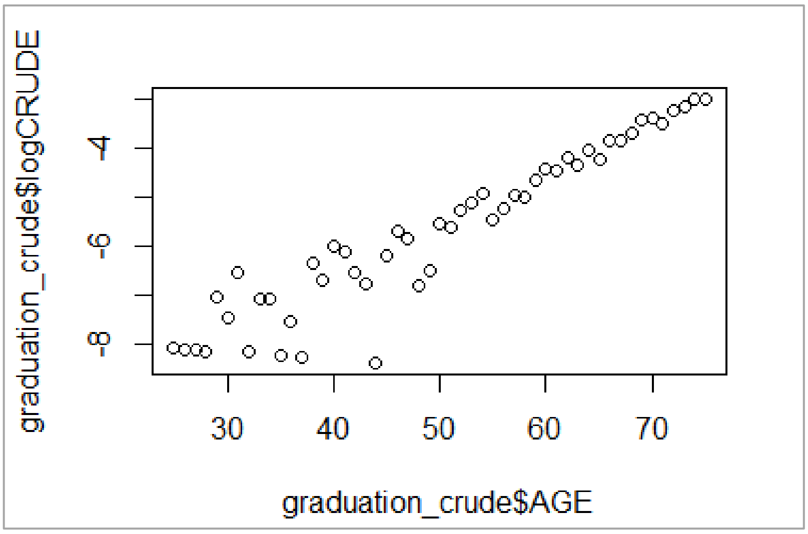
1) Crude Mortality Rate for age x last birthday = $\frac{\text{Number of deaths recorded at age x last birthday}}{\text{Central Exposure to Risk for age x last birthday}}$

2)

The crude mortality rates are not linearly related to age, as seen in the plot below



Hence, we perform a **log transformation** to the crude rates and see that the log of crude rates is linearly related to age. We can now go ahead with a **linear model**.



According to the **Gompertz Law:** $\mu_x = BC^x$

where: μ_x is the mortality rate

: B and C are parameters

: x is the age last birthday

So, **$\log(\mu_x) = \log(B) + x*\log(C)$**

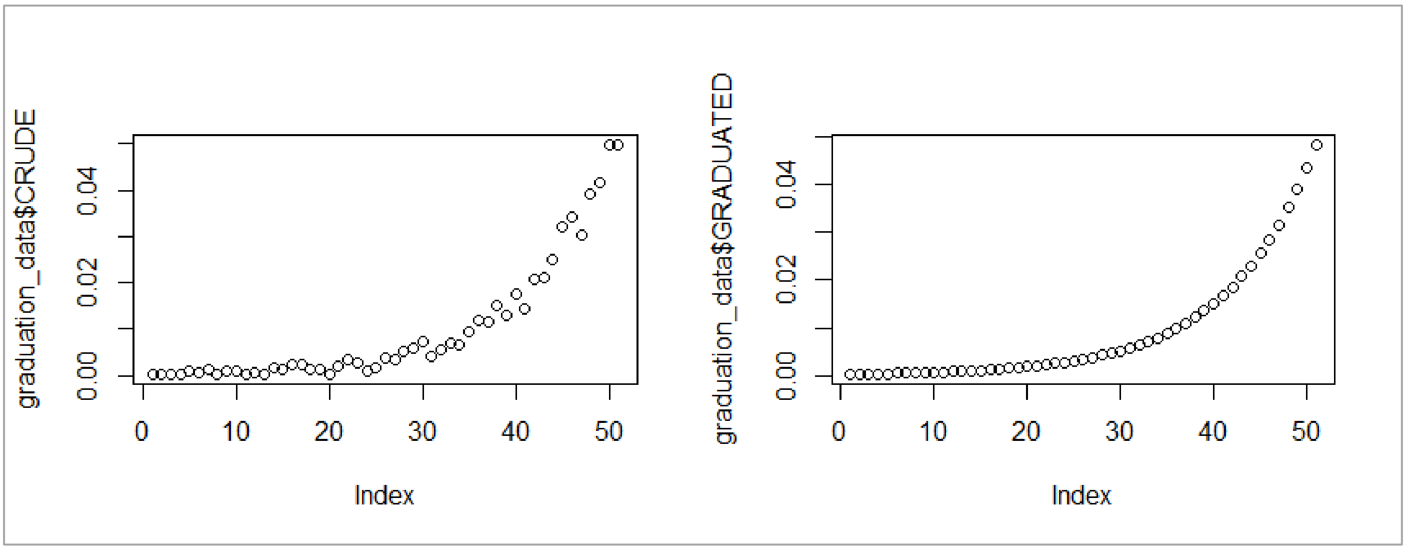
Carrying out linear regression with log of Crude Rate as the response variables and Age as the explanatory variable:

```
Call:
lm(formula = logCRUDE ~ AGE, data = graduation_crude)

Residuals:
     Min       1Q   Median       3Q      Max
-2.07946 -0.09831  0.08351  0.21582  1.13949

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.000864   0.256884  -42.82   <2e-16 ***
AGE           0.106298   0.004929   21.57   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5181 on 49 degrees of freedom
Multiple R-squared:  0.9047,    Adjusted R-squared:  0.9028
F-statistic: 465.2 on 1 and 49 DF,  p-value: < 2.2e-16
```

The graduated mortality rates were then calculated as the fitted values of the model created. The parameters to be estimated are the coefficients of the model.

B = 1.668727e-05

C = 1.112153



The graduated rates have a much smoother exponential curve as compared to the crude rates

3)

Testing the crude rates and graduated rates for smoothness:

The criterion of smoothness is usually that the third differences of the graduated rates should:

i) Be small in magnitude compared with the quantities themselves

ii) progress regularly

On calculating the third differences, the **graduated rates qualify the Smoothness Test** since the third differences are both small in magnitude as compared with the graduated rates as well as progress regularly, whereas the crude rates do not do so. Thus, the graduated rates are much smoother than the crude rates

4)

Expected number of deaths at age x last birthday = central exposure to risk for age x last birthday*graduated rates for age x last birthday

Standardised deviation for age x = (Number of deaths recorded at age x last birthday - Expected number of deaths at age x last birthday)/ (sqrt (Expected number of deaths at age x last birthday)

Performing the Chi-Square Goodness of Fit Test:

H0: The graduated rates are the true underlying mortality rates

H1: The graduated rates are not the true underlying mortality rates

```
        Pearson's Chi-squared test

data:  graduation_chisq
X-squared = 905.26, df = 50, p-value < 2.2e-16
```

The Chi-Square test statistic is **905.26** whereas the p-value is **2.475047e-157 ~ 0**

The degrees of freedom to be used is **49**, since we have 51 ages in the data provided and have estimated 2 parameters while fitting Gompertz Law to the model

Since the p-value is less than the 5% level of significance, we have **sufficient evidence to reject H0** and conclude that the graduated rates are not in line with the actual mortality rates

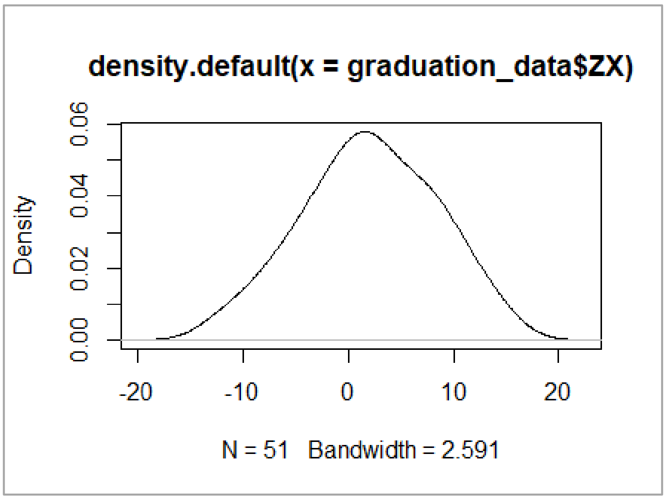
5)

a) Performing the Individual Standardized Deviations Test:

H0: the z(x) values are realisations of a standard normal random variable

Interval	(-20,-16]	(-16,-12]	(-12,-8]	(-8,-4]	(-4,0]	(0,4]	(4,8]	(8,12]	(12,16]	(16,20]
Observed Value	0	1	2	6	11	12	11	5	3	0

i) Overall Shape: The observed graph is **much wider** than the standard normal graph as seen from the observed values as well as the density plot below



ii) Absolute Deviations: The absolute deviations are much **higher** as compared to the expected values

iii) Outliers: The lower bound is -12.124 whereas the upper bound is 14.64. There are no values below and above these bounds respectively and we can say that there are **no significant number of outliers**

iv) Symmetry: The number of positive standard deviations is 31 which is greater than the number of negative standard deviations which is 20. There is a huge trade-off between the counts of positives and negatives which is not favourable since this indicates a **non-symmetrical distribution** (we see that the graph is negatively skewed). The excess of positives indicate that the graduated rates are too low perhaps due to the presence of bias

v) Conclusion: Given the above features, it is seen that the standardised deviations do not appear to conform to a normal distribution. This indicates that the observed mortality rates do not conform to the model with the rates assumed in the graduation. Thus, we have **sufficient evidence to reject the null hypothesis** that the z(x) values are realisations of a standard normal random variable

b) <u>Performing the Signs Test:</u>

The number of positive z(x)s ~ Binomial(51, 0.5). The number of observed positive values is 31 which is greater than the expected number of 25. We get a p-value of **0.1607796**, which is greater than 0.05, so there is **insufficient evidence to reject H0** and we conclude that there is little evidence of bias in the graduated rates.

c) <u>Performing the Cumulative Deviations Test</u>:

H0: Graduated rates are not biased

H1: Graduated rates are biased

Total observed deaths = **66294**

Total expected deaths = **61619.71**

Test Statistic = (Total observed deaths - Total expected deaths)/(sqrt(Total expected deaths))

= **18.83024**

Comparing the test statistic value with the upper and lower 2.5% points of N(0,1) of +/-1.96.
Since 18.83024 > 1.96, there is **sufficient evidence to reject H0** and conclude that the test provides evidence that the **graduated rates are biased** or that the variance is higher than predicted by the model

d) <u>Performing the Serial Correlations Test</u>:

H0: grouping of signs is absent

H1: grouping of signs is present

Serial correlation coefficient at lag 1 = r1 = **0.1477466**

Test Statistic = **1.055122**

Since the test statistic value of 1.055122 is not more than the upper 5% point of N(0,1), there is **insufficient evidence of grouping of deviations** of the same sign