1 Investigation 1: Daily Rainfall

For investigation 1, I explored the rainfall in a location over 40 years and in this time, the daily rainfall exceeded 25mm on 145 days. The data included by how much the rainfall exceeded 25mm in those days and I looked at 2 different distibutions to represent this data.

1.1 Gamma Distribution

The excess rainfall amounts are modelled as realisations of independent and identically distributed random variables. I chose to use the parametric model $X \sim Gamma(\alpha, \beta)$ and therefore the model distribution is $X \sim Gamma(\hat{\alpha}, \hat{\beta})$ as α and β are estimated. Then, to estimate these model parameters, I found the mean and variance of the excess rainfall data using RStudio and used the equations for the gamma distribution:

$$E[X] = \frac{\hat{\alpha}}{\hat{\beta}} = 18.89, Var[X] = \frac{\hat{\alpha}}{\hat{\beta}^2} = 616.11$$

Solving these equations gave the estimations of $\hat{\alpha} = 0.5791$ and $\hat{\beta} = 0.0307$ (both to 4dp) for the model parameters. So this gave the model distribution of $X \sim Gamma(0.5791, 0.0307)$.

Then, to estimate the standard errors of the parameter estimates, I used simulation in RStudio from the command $rgamma(n, \hat{\alpha}, \hat{\beta})$ where n is 145. This gave standard error for $\hat{\alpha}$ as 0.1065 and for $\hat{\beta}$ as 0.0068 (both to 4dp).

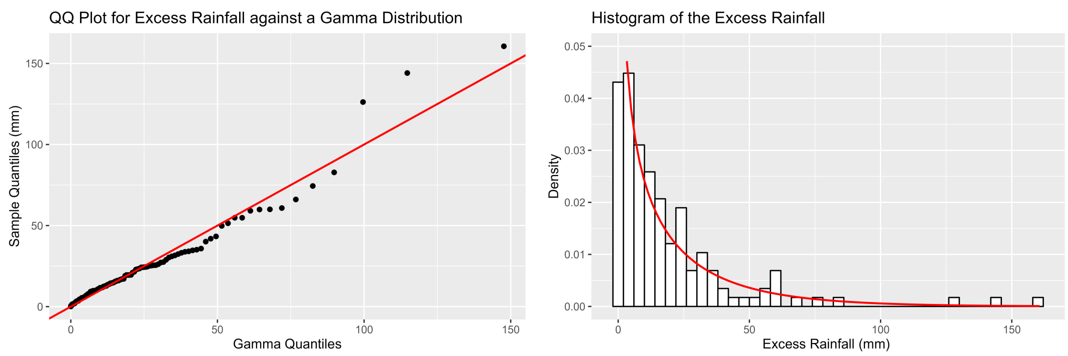

Figure 1: Graphs to show the realism of the Gamma Distribution

Then, to visualise this model's realism, I plotted two graphs on RStudio as seen in figure 1. The graph on the left is a quantile quantile plot for the excess rainfall against the gamma distribution. We know that the closer the points lie to the red line of y=x, the better fit the model is. Therefore this graph shows the gamma distribution is a moderately good fit to the data but the tails of the gamma distribution appear to be too light. The graph on the right is a histogram of the excess rainfall with the red line of a gamma distribution. This also shows a moderately good fit as it fits the general pattern with some data sitting just above or below.

1.2 Model Distribution

Then, for comparison, I modelled the excess rainfall amounts as realisations of independent and identically distributed random variables with a distribution $Y \sim$

$M(\sigma, \gamma)$. **Using the mean and variance found before from the data, I estimated $\hat{\sigma}$ and $\hat{\gamma}$ using the equations given:**

$$E[Y] = \frac{\hat{\sigma}}{(1-\hat{\gamma})} = 18.89, Var[Y] = \frac{\hat{\sigma}^2}{(1-\hat{\gamma})^2(1-2\hat{\gamma})} = 616.11$$

giving $\hat{\sigma} = 14.91$ and $\hat{\gamma} = 0.21$ (both to 2dp). So the model distribution is $Y \sim M(14.91, 0.21)$.

For the standard errors of the parameter estimates in this model, I used a similar simulation in RStudio as the gamma distribution. However I first rearranged equations for σ and γ in terms of the mean and variance for use in the simulation with command rmodel:

$$\sigma = \frac{Mean(y)}{2} + \frac{(Mean(y))^3}{2Var(y)}, \gamma = \frac{1}{2} - \frac{(Mean(y))^2}{2Var(y)}$$

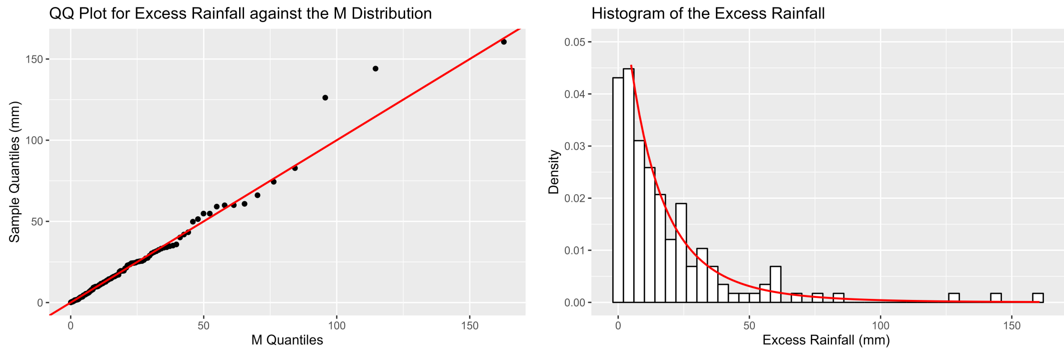

Figure 2: Graphs to show the realism of the Model Distribution

Using these with the simulation gave the standard error for $\hat{\sigma}$ as 1.9131 and for $\hat{\gamma}$ as 0.0896 (both to 4dp). This shows there could be more error in $Y \sim M(14.91, 0.21)$ than in $X \sim Gamma(0.5791, 0.0307)$ for modelling the excess rainfall.

Figure 2 shows the visualisation of the realism of the model distribution in the same way as the gamma distribution. From both graphs, it seems the model distribution is a closer fit than the gamma distribution was. The red line on both graphs sits closer to the data than before, indicating $Y \sim M(14.91, 0.21)$ is a better model than $X \sim Gamma(0.5791, 0.0307)$, despite the standard errors suggesting the other way around.

1.3 10 Year Return Level

The 10 year return level is the daily rainfall total that is exceeded once every 10 years on average. Using my preferred model of $Y \sim M(14.91, 0.21)$, this is computed by $P(Y > y) = \frac{1}{36.5}$ with y being excessive rainfall over 25mm. This is equivalent to $P(Y \leq y) = \frac{71}{73}$ and this format can be used which the cdf function in Rstudio.

Trying some different values for y to find the value closest to but without passing $\frac{71}{72}$ gave y=80mm. This means that the daily rainfall reaches 105mm once every 10 years on average.

2 Investigation 2: Effectiveness of a New Antibiotic

In this investigation, I explored a new antibiotic's effectiveness at treating infections. A study took place over a year which showed whether the infection cleared within 2 weeks in 2 different hospitals using this antibiotic. We know there is an existing antibiotic which clears infections within 2 weeks for 70% of the target population.

2.1 Overall Effectiveness of the Antibiotic

Firstly, to see the antibiotic's overall effectiveness, I ignored which hospital the patients were from and assumed the patients are a simple random sample from the target population.

To get the point estimate, I found the sample mean of the data given from the simple random sample which was 0.7347 (4dp). This suggests that 73.47% of patients using the new antibiotic have their infection cleared within 2 weeks, which is slightly more effective than the existing antibiotic.

However, to find the uncertainty in this point estimate for the mean, I then found the confidence interval. I chose to use the 99% confidence interval as then the mean is more likely to sit in it as it would have high confidence. This gave:

$$P(z_1 \leq \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \leq z_2) = 0.99 \longrightarrow (\bar{Y} - z_2\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y} - z_1\frac{\sigma}{\sqrt{n}}) = 0.99$$

Where $\bar{Y}$ is the point estimate of the mean, σ the standard deviation, n the number of patients (98) and μ the true value of the mean. To find z_1 and z_2 I used the standard normal distribution which gave $P(Z < z_1) = P(Z > z_2) = 0.005$. Then using the statistical tables gave $z_1 = -2.5758$ and $z_2 = 2.5758$. The confidence interval is:

$$(\bar{Y} - z_2\frac{\sigma}{\sqrt{n}}, \bar{Y} - z_1\frac{\sigma}{\sqrt{n}}) \longrightarrow (0.6192, 0.8502)$$

From the found point estimate and the confidence interval, the new antibiotic appears to be effective for most people. The point estimate suggested it is slightly more effective than the existing antibiotic. However, the confidence interval shows that it is between 61.92% and 85.02% effective, so it is uncertain whether it is better than the existing antibiotic.

2.2 Effectiveness of the Antibiotic in each Hospital

Now we know that Hospital A is a general hospital but Hospital B is a specialist hospital (only treats a certain type of patient). Using this information, I reassessed the antibiotic's effectiveness and looked to see the difference between the hospitals.

Hospital A:

This hospital is the general hospital and there were 55 patients here. Firstly, I found the point estimate using the sample mean of just patients from this hospital. This gave 0.836 which suggests that 83.6% of patients found the antibiotic effective.

To find the uncertainty in this point estimate, I found the confidence interval at 99% like before. For this I first found that there were 55 patients in Hospital A and the standard deviation was 0.373. Using these and the same calculations as before, the confidence interval was: $(0.7064, 0.9656)$**.**

From this we can be 99% confident that the antibiotic is between 70.64% and 96.56% effective for patients in the general hospital with a point estimate of 83.6% effective. All of this is better than the effectiveness of the existing antibiotic which is 70%.

Hospital B:

This hospital is the specialist hospital which only treats a certain type of patient and there were 43 patients here. Using the same method as in Hospital A, I found the point estimate of 0.605, suggesting the antibiotic is 60.5% effective in this hospital.

Using the standard deviation of 0.495 and the same calculations again, I found the 99% confidence interval of: $(0.4106, 0.7994)$**.**

From this we can be 99% confident that the antibiotic is between 41.06% and 79.94% effective for patients in the specialist hospital with a point estimate of 60.5% effective. This is significantly lower than in the general hospital.

2.3 Strengths and Weaknesses of the Study

Firstly, looking at the strengths of the study, I concluded that it didn't appear to have any measurement bias. The effectiveness of the antibiotic was checked by a doctor seeing if the infection had cleared and this lacks any measurement errors. Also, the study followed the PICOT system well, ensuring the study was thought through. The population was patients with an infection, intervention the new antibiotic, comparison the existing antibiotic, outcome checking the infection had cleared after 2 weeks and time was the study lasting a year.

When looking closer into the details of the study, I noticed some weaknesses. There was both sampling and generalisation bias. For sampling bias, participants in the study were "consenting patients". This brings a volunteer bias as there would have been some who did not want to participate, potentially changing the overall result as the sample did not represent the whole of the study population. For generalisation bias, 43/98 patients were from specialist hospitals that only treat a certain type of patient. These patients may be at the specialist hospital for other reasons, impacting the antibiotic's ability to work effectively. Also, most UK patients will be in general hospitals not specialist. Both these make it not representative of the target population. These sources of bias bring weaknesses to the study, affecting the outcome and making the effectiveness of the antibiotic more uncertain.

Therefore, using all the data found in the study along with the strengths and weaknesses above, we can conclude the antibiotic is potentially more effective than the existing one. Despite the specialist hospital presenting a lower effectiveness, the general hospital's data suggested that the antibiotic is more effective. It is more useful to look at the general hospital's data as it is more representative of UK patients, however we shouldn't ignore the data from the specialist hospital entirely.