

# Data Analysis

## Slide Data Analysis

- My name is Kennedy Mwavu. I'm going to take you through the data analysis part.

## Slide Descriptive Statistics

- The very first thing we did was to generate some descriptive statistics which gave us the general overview of the data.
- From this table, we can say that the incurred claims data for both classes of auto-insurance is positively skewed.

## Slide Histograms

- Since a picture is worth a thousand words, we went ahead and got a graphical representation of the distribution of our data.
- From the histograms, we can see that not only is the data skewed to the right, but also long-tailed. This gives a good hint of the kind of distributions most appropriate to model the data.

## Slide Normal QQ-Plots

- From the normal qqplots it's clear that most points deviate from the reference line. This implies that the datasets don't match the normal standard distribution.
- This is a good indication that after fitting the distributions, non-parametric tests would be applied as opposed to parametric tests.

## Slide Transformed Data

- The original data contained very large values. We did a cube root transformation to make it easier to work with.
- The transformed data seemed to be more suitable for purposes of fitting the distributions compared to the original data.

## Slide Parameter Estimation

- The next thing was parameter estimation.
- Parameters for the various fitted distributions were obtained using the MLE method.
- The most appropriate distribution is the one with the highest log-likelihood function.

## Slide MLE Table

- From this table we can see that under the motor commercial class, the Weibull distribution had the highest log-likelihood function while under the motor private class it's the Gamma distribution which had the highest log-likelihood function.
- The exponential distribution had the least log-likelihood function.

### **Slide Goodness-Of-Fit Test**

- The next step was to perform goodness-of-fit test.
- K-S and A-D tests were used to determine the appropriateness of the fitted distributions to the claim size data.
- The smaller the K-S and A-D test statistic values, the more suitable a distribution is for fitting the data.
- For the motor commercial class, Weibull distribution had the smallest K-S statistic but the Gamma distribution had the smallest A-D statistic.
- For the motor private class, Gamma distribution had the least K-S statistic as well as A-D statistic.

### **Slide Information Criterion**

- Under the Information Criterion section we used the AIC and BIC.
- Lower values for both AIC and BIC indicate a more appropriate distribution.
- For the motor commercial class, Weibull had the lowest AIC and BIC values.
- For motor private, Gamma distribution had the minimum AIC and BIC values.
- But under both classes, exponential had the highest AIC and BIC values, indicating it's not a good fit compared to the others.

I will now pass it on to Rachael for the conclusion part.