Model Result

## Differencing

Applying linear regression to non-stationary time series data results to a spurious regression model which shows a false relationship due to, for instance, a common trend in otherwise unrelated variables. Also, it makes predictors significant in explaining the dependent variable, while in reality, they are not. This means that the results would not be reliable for inference.

We employed the Augmented Dickey-Fuller (ADF) test to assess the stationarity of our data. The test indicated that none of the time series were stationary initially. To prevent potential issues of spurious regression, we addressed this by differencing the data with an order of 1, resulting in a stationary time series.

## Correlation between human incidence and Animal incidence

We conducted an investigation to explore the correlation or association between the incidence of brucellosis in both human and animal species. Our primary objective was to assess the direction of any identified association. To achieve this, we employed lagged predictor variables, specifically incidences in cattle, goat, sheep, and camel, at various time lags (lag 1 to lag 6). For each lag, we calculated the correlations between the predictors and explanatory variables as well as within the predictors. Since the data was collected on a monthly basis, lag 1 corresponds to a one-month interval.



Correlation between human and animal incidence at different lags

At lag 6, there was a fairly strong positive correlation between human and camel incidences. Also, at this lag, all the predictors had a positive (though weak) correlation with the human incidence, thus, we chose lag 6 as the best lag to model with. There was a significant correlation between goat and cattle incidence, in all the lags.

Linear regression model dictates that the predictors must be independent of each other. Thus, the goat incidence was omitted in the second model.

## Modelling

Initially, we conducted individual time series linear regressions between human incidence and each animal incidence rate separately. Subsequently, we performed a comprehensive model incorporating all animal incidence rates as predictors.

### Indiividual models

#### Cattle and Human incidence

The results were as follows;

| variable | estimate | std.error | statistic | p.value | conf\_low | conf\_high |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | 0.164 | 0.132 | 1.240 | 0.219 | -0.095 | 0.423 |
| Cattle Incidence | 0.199 | 0.251 | 0.794 | 0.430 | -0.292 | 0.691 |

#### Goat and Human Incidence

The results were as follows

| variable | estimate | std.error | statistic | p.value | conf\_low | conf\_high |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | 0.166 | 0.132 | 1.25 | 0.213 | -0.093 | 0.425 |
| Goat Incidence | 0.074 | 0.088 | 0.84 | 0.403 | -0.098 | 0.246 |

#### Sheep and Human Incidence

The results were as follows

| variable | estimate | std.error | statistic | p.value | conf\_low | conf\_high |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | 0.166 | 0.132 | 1.256 | 0.213 | -0.093 | 0.426 |
| Sheep Incidence | 1.056 | 1.723 | 0.613 | 0.542 | -2.321 | 4.434 |

#### Camel and Human Incidence

The results were as follows

| variable | estimate | std.error | statistic | p.value | conf\_low | conf\_high |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | 0.166 | 0.120 | 1.38 | 0.171 | -0.07 | 0.402 |
| Camel Incidence | 2.336 | 0.553 | 4.22 | 0.000 | 1.25 | 3.421 |

### Full model

We first fitted a model with the goat incidence included. The results were as follows;

| variable | estimate | std.error | statistic | p.value | conf\_low | conf\_high |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | 0.167 | 0.120 | 1.389 | 0.169 | -0.069 | 0.403 |
| Cattle incidence | -0.105 | 0.448 | -0.234 | 0.816 | -0.982 | 0.773 |
| Camel incidence | 2.527 | 0.568 | 4.448 | 0.000 | 1.414 | 3.641 |
| Sheep incidence | 2.457 | 1.607 | 1.529 | 0.130 | -0.693 | 5.608 |
| Goat Incidence | 0.098 | 0.157 | 0.626 | 0.533 | -0.209 | 0.405 |

The high correlation between goat and cattle incidence was affecting the model results where the regression coefficient for cattle was negative. Thus, in the second model, the goat incidence was omitted giving the following results;

| variable | estimate | std.error | statistic | p.value | conf\_low | conf\_high |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | 0.165 | 0.120 | 1.376 | 0.173 | -0.070 | 0.399 |
| Cattle incidence | 0.136 | 0.228 | 0.599 | 0.551 | -0.310 | 0.583 |
| Camel incidence | 2.497 | 0.564 | 4.428 | 0.000 | 1.392 | 3.603 |
| Sheep incidence | 2.563 | 1.592 | 1.610 | 0.111 | -0.558 | 5.684 |

## Model validation