moz-cholera-analysis

To run: rmarkdown::render(paste0(rmd\_dir, “/moz-cholera-analysis.qmd”))

## Prepare dataset

df <- cholera\_wash\_sanitation\_district\_tbl %>%   
 dplyr::select(c(ADM1\_PT, ADM2\_PT, pop\_surf\_ratio, pop\_od\_ratio, incidence)) %>%   
 mutate(across(c(ADM1\_PT, ADM2\_PT), factor))

## 0. Test data distribution

Test if dependat variable (incidence) is normally distributed using Shapiro. P-value < 0.05 -> We reject the null hypothesis (that is, that data was normally distributed):

shapiro.test(df$incidence)

##   
## Shapiro-Wilk normality test  
##   
## data: df$incidence  
## W = 0.67315, p-value = 8.864e-07

Test if Poisson distribution. Mean and variance should be similar, which is not the case -> not Poisson distribution

mean(df$incidence)

## [1] 123.7361

var(df$incidence)

## [1] 25057.96

## 1. Mixed-effects model (hierarchical model)

Using only ADM1\_PT as random effect

# Using only ADM1\_PT as random effect  
model <- lme4::lmer(incidence ~ pop\_surf\_ratio + pop\_od\_ratio + (1 | ADM1\_PT), data = df)

## boundary (singular) fit: see help('isSingular')

summary(model)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: incidence ~ pop\_surf\_ratio + pop\_od\_ratio + (1 | ADM1\_PT)  
## Data: df  
##   
## REML criterion at convergence: 335  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.5294 -0.4833 -0.1997 0.3772 3.6733   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## ADM1\_PT (Intercept) 0 0.0   
## Residual 20676 143.8   
## Number of obs: 29, groups: ADM1\_PT, 6  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 55.19 49.00 1.126  
## pop\_surf\_ratio -128.72 226.27 -0.569  
## pop\_od\_ratio 350.04 125.81 2.782  
##   
## Correlation of Fixed Effects:  
## (Intr) pp\_sr\_  
## pop\_surf\_rt -0.561   
## pop\_od\_rati -0.385 -0.352  
## optimizer (nloptwrap) convergence code: 0 (OK)  
## boundary (singular) fit: see help('isSingular')

## 2. Multivariate linear regression model

Fit the linear regression model

# Fit the linear regression model  
model <- lm(incidence ~ pop\_surf\_ratio + pop\_od\_ratio, data = df)  
  
# Output the summary of the model  
summary(model)

##   
## Call:  
## lm(formula = incidence ~ pop\_surf\_ratio + pop\_od\_ratio, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -219.92 -69.50 -28.72 54.23 528.20   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 55.19 49.00 1.126 0.27035   
## pop\_surf\_ratio -128.72 226.27 -0.569 0.57432   
## pop\_od\_ratio 350.04 125.81 2.782 0.00992 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 143.8 on 26 degrees of freedom  
## Multiple R-squared: 0.2338, Adjusted R-squared: 0.1749   
## F-statistic: 3.967 on 2 and 26 DF, p-value: 0.03136