Dairy Farm Management and Percentage of Farm Samples Containing Tetracycline Resistant E. coli

Abstract

My goal is to identify possible farm management causes of prevalent tetracycline resistant E. coli samples. The data consists of farm management and longitudinal data on antibiotic use and antibiotic resistant E. coli for 53 dairy farms in the South-West of England between 2017-2019. I used one-way anova tests to determine if the percentage of samples for a given farm found to contain tetracycline resistant E. coli is significantly different among 'first choice/most commonly used class of antibiotic for clinical mastitis' levels and significantly different among 'antibiotic most commonly used for mastitis at the start of the project' levels. I also used a Pearson's product-moment correlation test to determine if the percentage of samples for a given farm found to contain tetracycline resistant E. coli is positively correlated with 'Amount of tetracycline used in mg/PCU.' I conclude that 'Amount of tetracycline used in mg/PCU' is positively correlated with the percentage of samples for a given farm found to contain tetracycline resistant E. coli; There is no significant difference in percentage of samples containing tetracycline resistant E. coli among 'first choice/most commonly used class of antibiotic for clinical mastitis' factor levels. And there is a significant difference in percentage of samples containing tetracycline resistant E. coli among 'antibiotic that was most commonly used for mastitis at the start of the project' factor levels.

Introduction

Farm management and longitudinal data on antibiotic use and antibiotic resistant E. coli for 53 dairy farms in the South-West of England between 2017-2019 data used in this analysis was part of the OH-STAR (One Health Selection and Transmission of Antimicrobial Resistance) project at the University of Bristol, funded by the Antimicrobial Resistance Cross Council Initiative supported by the seven United Kingdom research councils. Samples and data were collected from 53 dairy farms in the South-West of England between January 2017 and December 2019. The data consists of farm demographic and management data collected from questionnaires administered to the farmer participants, data recorded at the time of collecting each sample, lab results from screening of samples for presence of E. coli resistant to the named antibiotics and from genetic analysis of selected isolates from the samples, and prescription records acquired from veterinary practices and farm records. Farms were visited monthly between January 2017 and December 2018. Samples were collected from various areas of each farm: Samples were collected using sterile overshoes (over-boot socks) traversing farm areas; Where access was restricted (e.g. for pens containing single or pairs of calves), samples were collected directly from the ground using gloved hands. Samples were refrigerated from collection to processing, prepared, spread onto TBX agar containing no antibiotic or containing tetracycline, amoxicillin, ciprofloxacin, streptomycin, or cephalexin. After incubation, the number of blue colonies (indicating E. coli) counted (Schubert et al.).

Antimicrobial resistance, and particularly antibacterial resistance (ABR), is a significant global challenge. Many countries are implementing plans to reduce the use of antibacterial drugs (ABs) in food-producing animals. One reason for reducing AB use in farming is to reduce the prevalence of ABR bacteria carried by farm animals. Escherichia coli is a species commonly found in animal feces and considered one of the most significant potential zoonotic ABR threats to humans (Schubert et al.).

Since the introduction of penicillin in the 1940s, which started the era of antibiotics, these agents have been recognized as one of the greatest advances in modern medicine and a turning point in human history. In

1900, infectious disease was a leading cause of death; in 2000, infectious diseases were responsible for only a small percentage of deaths in developed nations [3]. Unfortunately for humans, bacteria have evolved different mechanisms that have rendered them resistant to antibiotics, to the point that since not long ago antimicrobial resistance has become a global threat to public health systems worldwide (Galindo-Méndez et al.).

My hypothesis is that the percentage of samples for a given farm found to contain tetracycline resistant E. coli is significantly different among 'first choice/most commonly used class of antibiotic for clinical mastitis' levels and significantly different among 'antibiotic most commonly used for mastitis at the start of the project' levels. I also hypothesize that the percentage of samples for a given farm found to contain tetracycline resistant E. coli is positively correlated with 'Amount of tetracycline used in mg/PCU.'

```
#read in data
dairyFarms = read.csv("/Users/merej/Downloads/S22 EEMB 146/S22 EEMB Final
↪  Project/dairy_farms_dataset.csv")
```

```
resistance <- dairyFarms %>%
  select(ctx_m, amp_c:amoxR, plain_undiluted) #select subset
print(paste('The dimensions of the selected subset of the dairy farms data are',
↪  nrow(resistance), 'rows and', ncol(resistance), 'columns.')) #display dimensions of
↪  set
```

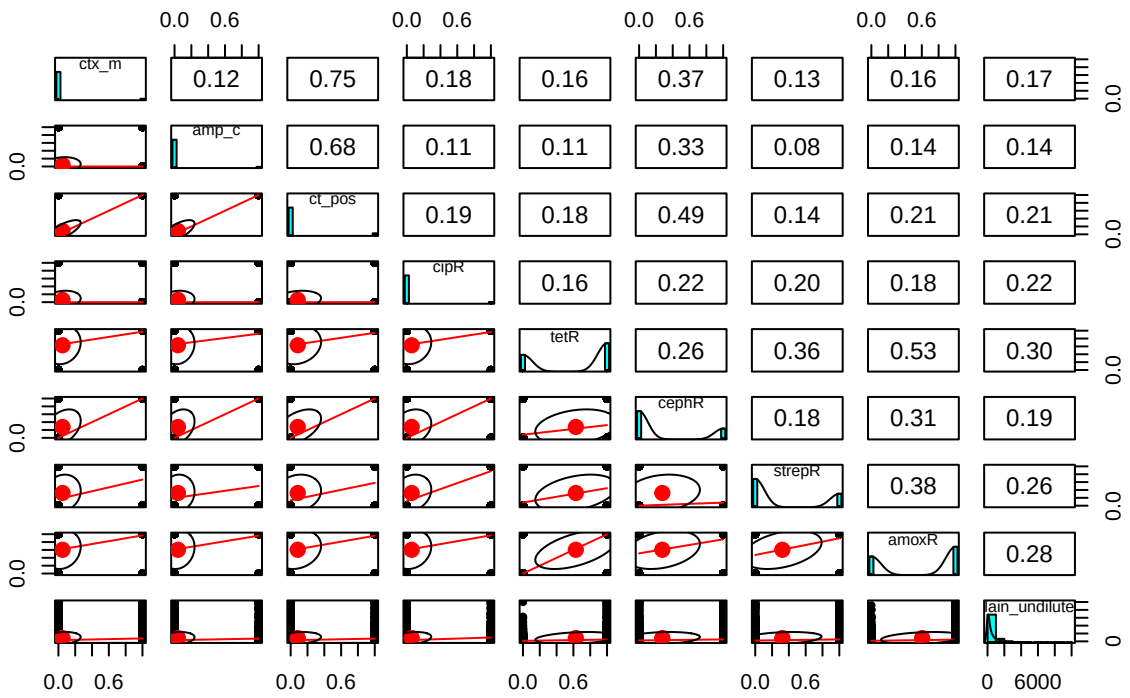
```
## [1] "The dimensions of the selected subset of the dairy farms data are 4578 rows and 9 columns."
```

```
head(resistance)
```

```
##   ctx_m amp_c ct_pos cipR tetR cephR strepR amoxR plain_undiluted
## 1     0     0      0    0    0     0      0     0               0
## 2     0     0      0    0    0     0      0     0               0
## 3     0     0      0    0    0     0      0     0               0
## 4     0     0      0    0    1     0      0     1             350
## 5     0     0      0    0    1     0      0     1            3760
## 6     0     0      0    0    1     0      0     1             120
```

```
pairs.panels(resistance, cor = TRUE, main = "Correlation Coefficents") #display
↪  correlation coefficients
```

Correlation Coefficents




```
resistance_fit <- lm(log(plain_undiluted + 1)~., resistance) #run linear model #log
↪  transformation
summary(resistance_fit) #display summary statistics
```

```
##
## Call:
## lm(formula = log(plain_undiluted + 1) ~ ., data = resistance)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.4008 -1.1847  0.1652  1.2441  5.6777
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.06761    0.04798  63.938  < 2e-16 ***
## ctx_m        0.08336    0.32195   0.259    0.796
## amp_c        0.28172    0.31712   0.888    0.374
## ct_pos      -0.04480    0.34320  -0.131    0.896
## cipR         0.50645    0.12137   4.173 3.06e-05 ***
## tetR         1.57713    0.06783  23.250  < 2e-16 ***
## cephR        0.51427    0.07210   7.133 1.14e-12 ***
## strepR       0.53519    0.06441   8.309  < 2e-16 ***
## amoxR        1.16161    0.06836  16.993  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

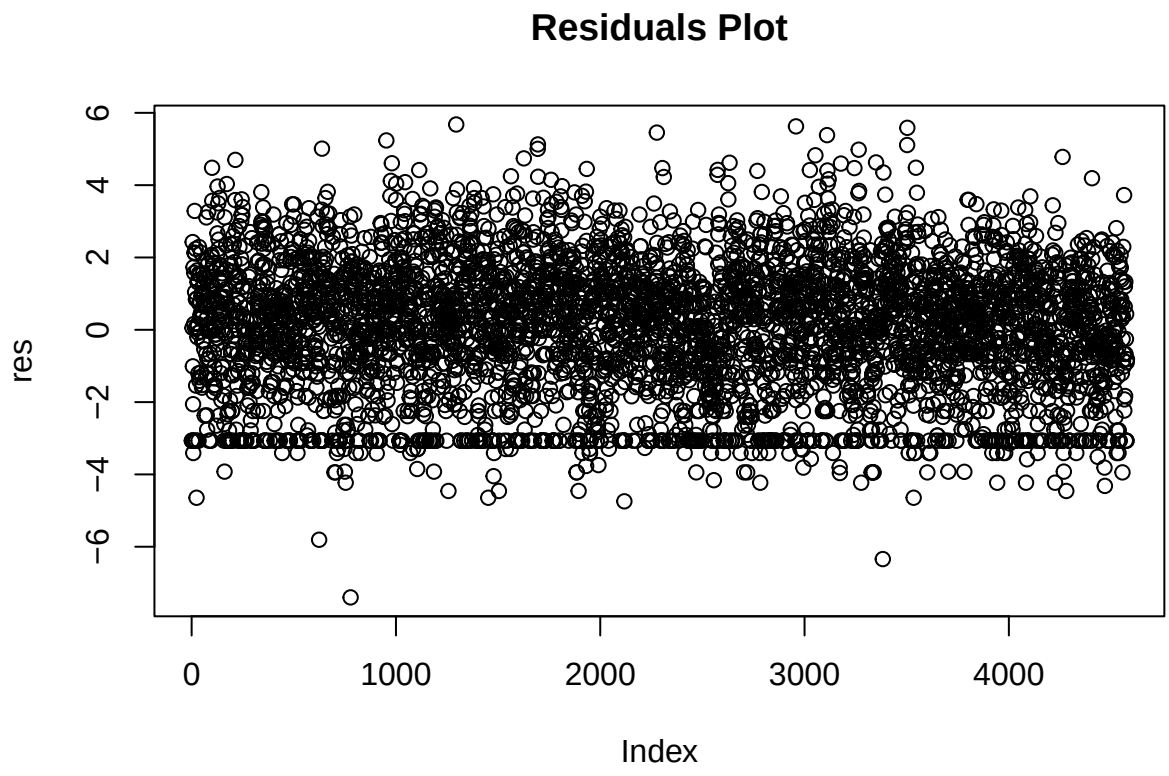
```
##
## Residual standard error: 1.827 on 4569 degrees of freedom
## Multiple R-squared:   0.38,  Adjusted R-squared:  0.3789
## F-statistic: 350.1 on 8 and 4569 DF,  p-value: < 2.2e-16
```

```
res <- resistance_fit$residuals #get residuals

#Check normality assumption
shapiro.test(res) #Shapiro_Wilk test on residuals
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.98308, p-value < 2.2e-16
```

```
#Check homogeneity of variance assumption
plot(res, main = "Residuals Plot") #residuals plot
```



The percentage of samples for a given farm found to contain tetracycline resistant E. coli are examined in the following sections due to the results of this linear model. The summary statistics indicate that the presence of tetracycline resistant E. coli in a sample is the most significant predictor of 'Total E. coli count in the plated sample on selective agar without antibiotic' which is stored in the variable plain_undiluted, with a test statistic of 23.250 and p-value < 2e-16. The assumptions of linear regression are met by this model: no two predictors are perfectly co-linear; the residuals are not found to be normal however, the

data contains 4578 points which is enough to assume normality through the Central Limit Theorem. The even spread of points of the residuals plot demonstrates the homogeneity of variance assumption. This assumption was met through the use of a log transformation.

Exploratory Data Analysis

```
tetRresistance <- dairyFarms %>%
  select(tetR, farm) %>%
  group_by(farm) %>%
  summarise(tetRPercent = sum(tetR)/n()) #subset, groupby, and summarize a new variable
  #tetRPercent = percentage of samples for a given farm found to contain tetracycline
  ↪  resistant E. coli

predictorsDup <- dairyFarms %>%
  select(farm, whichclinresp, firstmastitis, tet) #subset desired variables

#remove duplicate observations
predictors <- predictorsDup[!duplicated(predictorsDup$farm),] #all variables contained in
↪  predictorsDup are farm specific (samples from the same farm will have the same value
↪  for each of these variables)

#change these variable types to factor
predictors$whichclinresp <- as.factor(predictors$whichclinresp)
predictors$firstmastitis <- as.factor(predictors$firstmastitis)

tetRresistance <- merge(tetRresistance, predictors) %>% select(-farm) %>% na.omit()
↪  #merge sets, deselect farm, and omit observations containing NAs
```

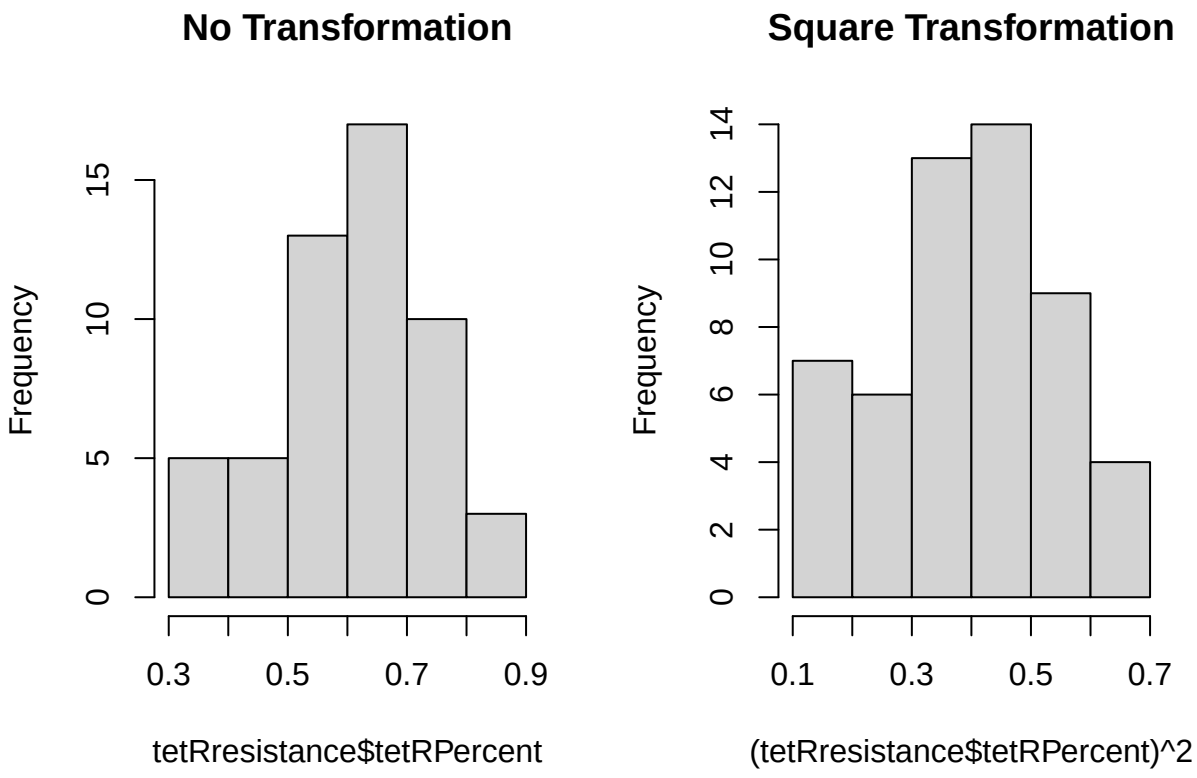
A Shapiro-Wilk normality test p-value > 0.05 is an indication of normality. (The null hypothesis states that the data is normally distributed and a p-value > 0.05 allows one to fail to reject the null hypothesis.)

```
par(mfcol = c(1,2))

hist(tetRresistance$tetRPercent, main = 'No Transformation') #to visualize normality
shapiro.test(tetRresistance$tetRPercent) #Shapiro-Wilk Normality Test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  tetRresistance$tetRPercent
## W = 0.95846, p-value = 0.0628
```

```
hist((tetRresistance$tetRPercent)^2, main = 'Square Transformation') #to visualize
↪  normality
```

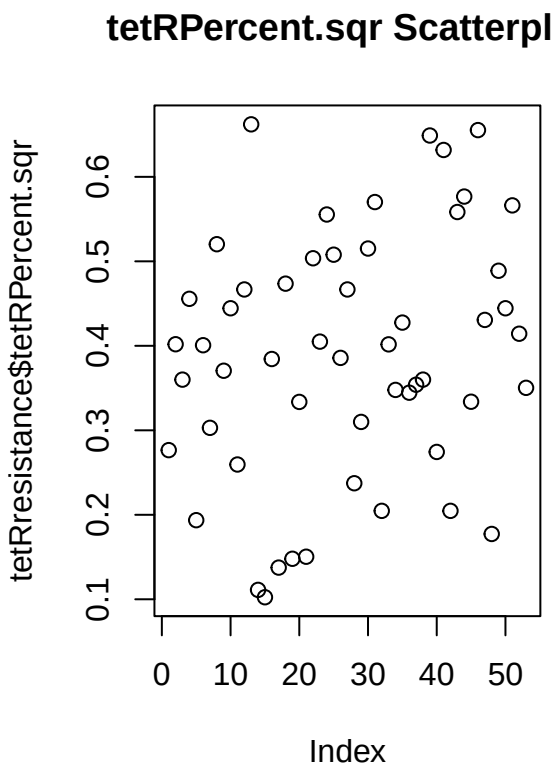



```
shapiro.test((tetRresistance$tetRPercent)^2) #Shapiro-Wilk Normality Test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  (tetRresistance$tetRPercent)^2
## W = 0.97762, p-value = 0.4177
```

```
tetRresistance$tetRPercent.sqr <- (tetRresistance$tetRPercent)^2 #make new variable for
↪  the square of tetRPercent

plot(tetRresistance$tetRPercent.sqr, main = 'tetRPercent.sqr Scatterplot')
```



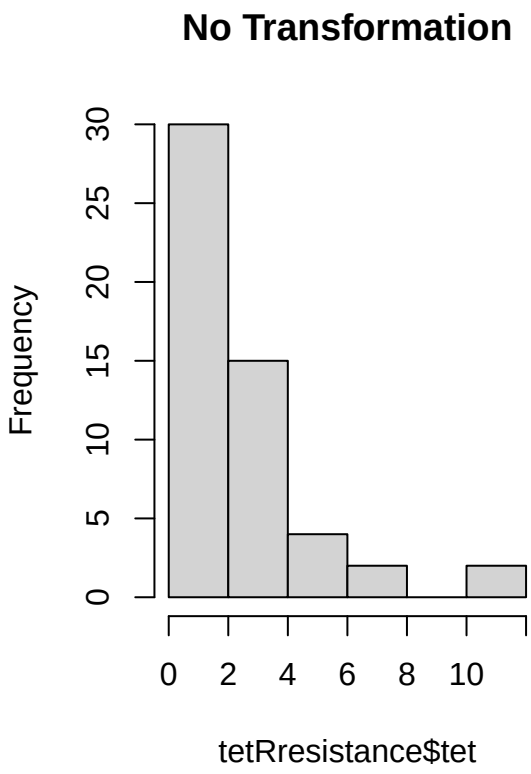
The Shapiro-Wilk Normality Test and histogram of the percentage of samples for a given farm found to contain tetracycline resistant E. coli indicate normality while the histogram is slightly left skewed. A Square transformation makes the data more normal as indicated and depicted through the Shapiro-Wilk Normality Test and histogram of the transformed data. The transformed data also appears to have equal variance as displayed by the even distribution of points in the scatter plot.

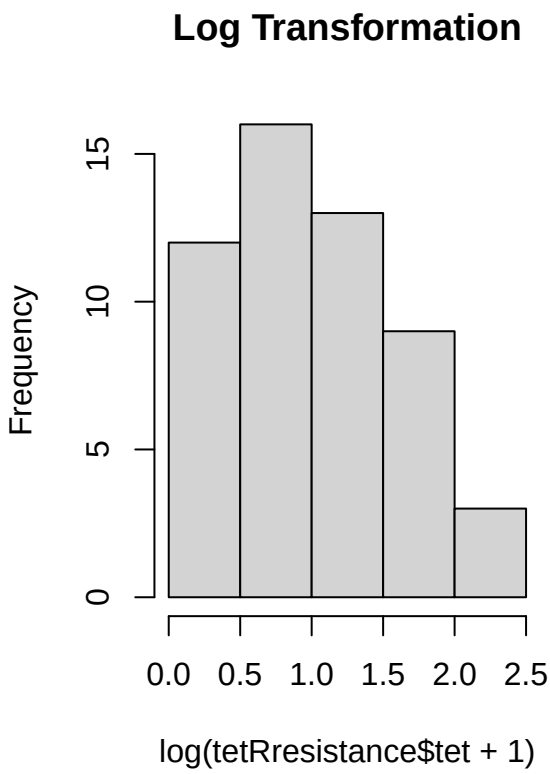
```
par(mfcol = c(1,2))

hist(tetRresistance$tet, main = 'No Transformation') #to visualize normality
shapiro.test(tetRresistance$tet) #Shapiro-Wilk Normality Test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  tetRresistance$tet
## W = 0.80226, p-value = 5.565e-07
```

```
hist(log(tetRresistance$tet + 1), main = 'Log Transformation') #to visualize normality
```





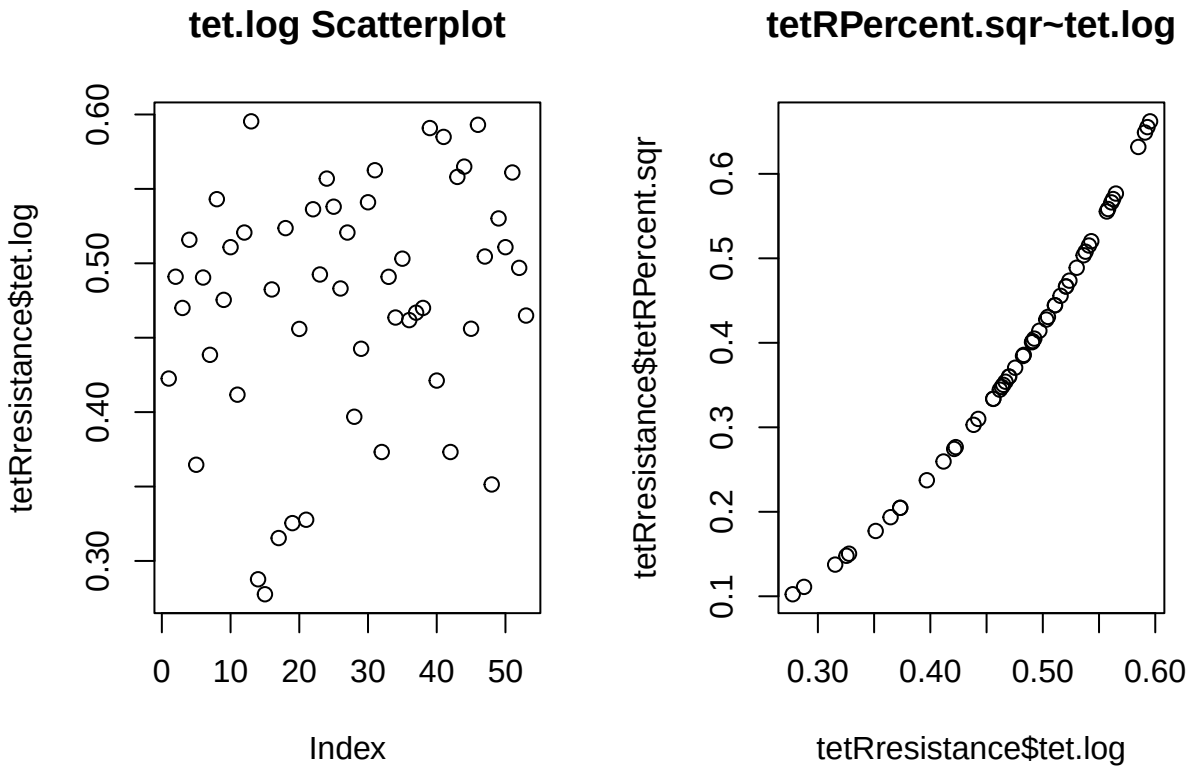

```
shapiro.test(log(tetRresistance$tet + 1)) #Shapiro-Wilk Normality Test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  log(tetRresistance$tet + 1)
## W = 0.9683, p-value = 0.1703
```

```
tetRresistance$tet.log <- log(tetRresistance$tetRPercent + 1) #make new variable for the
↪  log transformation of tet

plot(tetRresistance$tet.log, main = 'tet.log Scatterplot')

plot(tetRresistance$tet.log, tetRresistance$tetRPercent.sqr, main =
↪  'tetRPercent.sqr~tet.log')
```

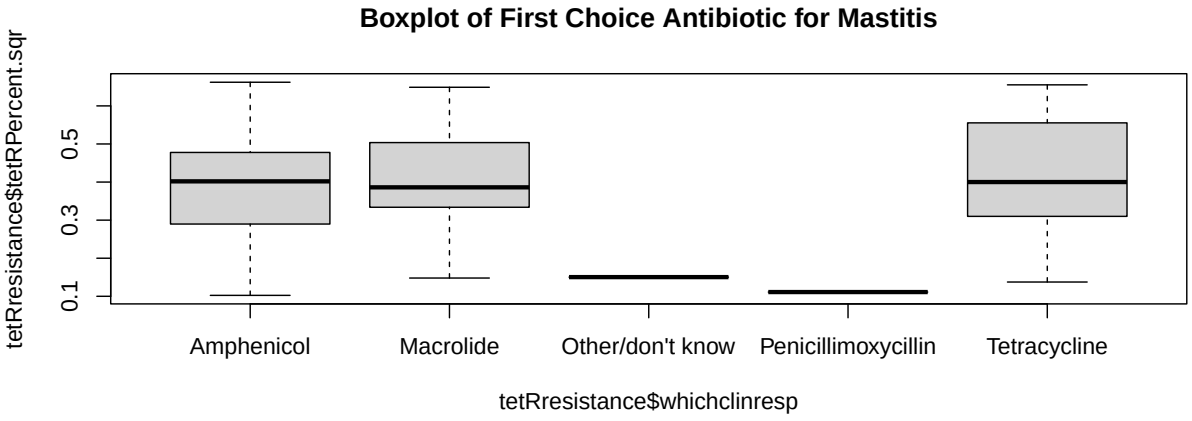


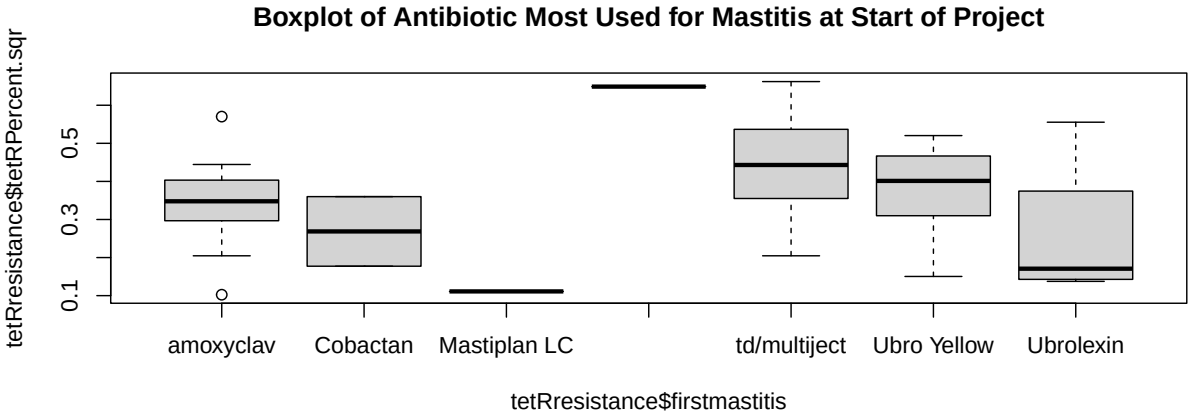
The Amount of tetracycline used in mg/PCU appears is right skewed and not normally distributed from the histogram and Shapiro-Wilk Normality test of the non-transformed data. A log transformation appears to make the data normally distributed from the histogram and Shapiro-Wilk Normality test of the transformed data. The transformed data also has equal variance as displayed by the even distribution of points in the scatter plot. The transformed data has a strong linear relationship with the square of the percentage of samples for a given farm found to contain tetracycline resistant E. coli as depicted in the tetRPercent~tet.log plot.

```
par(mfrow = c(2, 1))

boxplot(tetRresistance$tetRPercent.sqr~tetRresistance$whichclinresp, main = 'Boxplot of
↪  First Choice Antibiotic for Mastitis', outline = TRUE)

boxplot(tetRresistance$tetRPercent.sqr~tetRresistance$firstmastitis, main = 'Boxplot of
↪  Antibiotic Most Used for Mastitis at Start of Project', outline = TRUE)
```





The 'first choice/most commonly used class of antibiotic for clinical mastitis' data has no outliers and the factor levels have similar mean and variance except for factor levels 'Other/don't know' and 'Penicillimoxycillin' which appear to correspond with farms that have very low percentages of samples containing tetracycline resistant E. coli. The 'antibiotic that was most commonly used for mastitis at the start of the project' data has outliers on the 'amoxyclav' factor level. 'Mastiplan LC' appears to correspond with farms that have very low percentages of samples containing tetracycline resistant E. coli; While 'Orbenin LA' corresponds with farms that have very high percentages of samples containing tetracycline resistant E. coli. All the factor levels have different means and variances.

Statistical Methods

```
cor.test(tetRresistance$tet.log, tetRresistance$tetRPercent.sqr)
```

```
##
##  Pearson's product-moment correlation
##
## data:  tetRresistance$tet.log and tetRresistance$tetRPercent.sqr
## t = 41.368, df = 51, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9747620 0.9916009
```

```
## sample estimates:
##      cor
## 0.985424
```

Here I have run a Pearson's product-moment correlation test on the log transformed amount of tetracycline used in mg/PCU and square transformed percentage of samples containing tetracycline resistant E. coli. Null hypothesis: Amount of tetracycline used in mg/PCU and percentage of samples containing tetracycline resistant E. coli have no relationship ($\rho = 0$). Alternative hypothesis: the variables are correlated ($\rho \neq 0$). The assumptions of a Pearson's product-moment correlation test include that the data is a random sample as well as that X is normally distributed with equal variance for all values of Y, and Y is normally distributed with equal variance for all values of X. The tetRPercent~tet.log plot demonstrates this assumption. The Pearson's product-moment correlation test yields a p-value < 0.05; Therefore, we may reject the null hypothesis and conclude that the amount of tetracycline used in mg/PCU and percentage of samples containing tetracycline resistant E. coli are correlated ($\rho \neq 0$). The estimated correlation coefficient is 0.985424, which is very high.

```
whichclinresp_fit <- lm(tetRPercent.sqr~whichclinresp, tetRresistance) #run a linear
↪  model
whichclinresp_res <- whichclinresp_fit$residuals #retrieve the residuals from the model
shapiro.test(whichclinresp_res) #run a Shapiro-Wilk to test the normality of residuals
↪  assumption
```

```
##
##  Shapiro-Wilk normality test
##
## data:  whichclinresp_res
## W = 0.98492, p-value = 0.7378
```

```
leveneTest(tetRresistance$tetRPercent.sqr, tetRresistance$whichclinresp) #run a
↪  leveneTest to test the equal variance of residuals assumption
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  4  0.9135 0.4637
##       48
```

```
whichclinresp_aov <- aov(tetRPercent.sqr~whichclinresp, tetRresistance) #perform anova
↪  when all assumptions are met
summary(whichclinresp_aov) #display summary statistics
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## whichclinresp  4 0.1418 0.03546   1.748  0.155
## Residuals     48 0.9733 0.02028
```

```
TukeyHSD(whichclinresp_aov, conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tetRPercent.sqr ~ whichclinresp, data = tetRresistance)
```

```
##
## $whichclinresp
##                                           diff        lwr       upr
## Macrolide-Amphenicol                   0.015947655 -0.1110599 0.1429552
## Other/don't know-Amphenicol           -0.241846057 -0.6541139 0.1704218
## Penicillimoxycillin-Amphenicol        -0.281088965 -0.6933568 0.1311789
## Tetracycline-Amphenicol                0.005816732 -0.1470563 0.1586897
## Other/don't know-Macrolide            -0.257793712 -0.6724405 0.1568531
## Penicillimoxycillin-Macrolide         -0.297036620 -0.7116834 0.1176102
## Tetracycline-Macrolide                -0.010130923 -0.1693080 0.1490462
## Penicillimoxycillin-Other/don't know -0.039242908 -0.6100019 0.5315161
## Tetracycline-Other/don't know          0.247662789 -0.1756234 0.6709490
## Tetracycline-Penicillimoxycillin       0.286905697 -0.1363805 0.7101919
##                                           p adj
## Macrolide-Amphenicol                  0.9964481
## Other/don't know-Amphenicol           0.4661849
## Penicillimoxycillin-Amphenicol        0.3146652
## Tetracycline-Amphenicol               0.9999683
## Other/don't know-Macrolide            0.4071354
## Penicillimoxycillin-Macrolide         0.2675192
## Tetracycline-Macrolide                0.9997542
## Penicillimoxycillin-Other/don't know 0.9996662
## Tetracycline-Other/don't know         0.4688216
## Tetracycline-Penicillimoxycillin      0.3204322
```

```
firstmastitis_fit <- lm(tetRPercent.sqr~firstmastitis, tetRresistance) #run a linear
↪  model
firstmastitis_res <- firstmastitis_fit$residuals #retrieve the residuals from the model
shapiro.test(firstmastitis_res) #run a Shapiro-Wilk to test the normality of residuals
↪  assumption
```

```
##
##  Shapiro-Wilk normality test
##
## data:  firstmastitis_res
## W = 0.98641, p-value = 0.805
```

```
leveneTest(tetRresistance$tetRPercent.sqr, tetRresistance$firstmastitis) #run a
↪  leveneTest to test the equal variance of residuals assumption
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  6  0.5938 0.7336
##       46
```

```
firstmastitis_aov <- aov(tetRPercent.sqr~firstmastitis, tetRresistance) #perform anova
↪  when all assumptions are met
summary(firstmastitis_aov) #display summary statistics
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## firstmastitis  6 0.3405 0.05675   3.369 0.00775 **
```

```
## Residuals     46 0.7747 0.01684
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(firstmastitis_aov, conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tetRPercent.sqr ~ firstmastitis, data = tetRresistance)
##
## $firstmastitis
##                                 diff         lwr        upr     p adj
## Cobactan-amoxyclav        -0.07389843 -0.38139401 0.23359714 0.9891435
## Mastiplan LC-amoxyclav    -0.23142998 -0.64923420 0.18637424 0.6147161
## Orbenin LA-amoxyclav       0.30637866 -0.11142556 0.72418288 0.2850073
## td/multiject-amoxyclav     0.10223802 -0.04341202 0.24788806 0.3346654
## Ubro Yellow-amoxyclav      0.04090579 -0.13387425 0.21568583 0.9905569
## Ubrolexin-amoxyclav       -0.08392649 -0.31748615 0.14963317 0.9225663
## Mastiplan LC-Cobactan     -0.15753155 -0.64745042 0.33238732 0.9534295
## Orbenin LA-Cobactan        0.38027709 -0.10964178 0.87019597 0.2250271
## td/multiject-Cobactan      0.17613645 -0.11826815 0.47054105 0.5263350
## Ubro Yellow-Cobactan       0.11480422 -0.19504768 0.42465612 0.9114244
## Ubrolexin-Cobactan        -0.01002806 -0.35645301 0.33639690 1.0000000
## Orbenin LA-Mastiplan LC    0.53780864 -0.02790094 1.10351823 0.0721330
## td/multiject-Mastiplan LC  0.33366800 -0.07459773 0.74193373 0.1764168
## Ubro Yellow-Mastiplan LC   0.27233577 -0.14720569 0.69187723 0.4280461
## Ubrolexin-Mastiplan LC     0.14750349 -0.29972920 0.59473619 0.9475512
## td/multiject-Orbenin LA   -0.20414064 -0.61240637 0.20412509 0.7188859
## Ubro Yellow-Orbenin LA    -0.26547287 -0.68501433 0.15406859 0.4589909
## Ubrolexin-Orbenin LA      -0.39030515 -0.83753785 0.05692755 0.1239163
## Ubro Yellow-td/multiject  -0.06133223 -0.21189319 0.08922873 0.8680570
## Ubrolexin-td/multiject    -0.18616451 -0.40219842 0.02986941 0.1331814
## Ubrolexin-Ubro Yellow     -0.12483228 -0.36148558 0.11182102 0.6666878
```

Here I have run two one-way ANOVA tests. First null hypothesis: The mean percentage of samples containing tetracycline resistant E. coli for all factor levels of 'first choice/most commonly used class of antibiotic for clinical mastitis' are equal. Alternative hypothesis: at least one factor level mean is not equal. The assumptions of an ANOVA test include the data is a random sample, the residuals are normally distributed, and the residuals have equal variance. The Shapiro-Wilk Normality test of the residuals returns a p-value $>$ 0.05; Therefore we may fail to reject the null hypothesis and assume normality of the residuals. Levene's Test for Homogeneity of Variance of the residuals returns a p-value $>$ 0.05; Therefore we may fail to reject the null hypothesis and assume equal variance of the residuals. All of the ANOVA test assumptions are met. The ANOVA test of the relationship between 'percentages of samples containing tetracycline resistant E. coli' and 'first choice/most commonly used class of antibiotic for clinical mastitis' yields a p-value $>$ 0.05; Therefore, we may fail to reject the null hypothesis and conclude that the factor level means of 'first choice/most commonly used class of antibiotic for clinical mastitis' are equal and there is no significant difference in percentage of samples containing tetracycline resistant E. coli among 'first choice/most commonly used class of antibiotic for clinical mastitis' factor levels.

Second null hypothesis: The mean percentage of samples containing tetracycline resistant E. coli for all factor levels of 'antibiotic that was most commonly used for mastitis at the start of the project' are equal. Alternative hypothesis: at least one factor level mean is not equal. The Shapiro-Wilk Normality test of

the residuals returns a p-value > 0.05; Therefore we may fail to reject the null hypothesis and assume normality of the residuals. Levene's Test for Homogeneity of Variance of the residuals returns a p-value > 0.05; Therefore we may fail to reject the null hypothesis and assume equal variance of the residuals. All of the ANOVA test assumptions are met. The ANOVA test of the relationship between 'percentages of samples containing tetracycline resistant E. coli' and 'antibiotic that was most commonly used for mastitis at the start of the project' yields a p-value < 0.05; Therefore, we may reject the null hypothesis and conclude that the factor level means of 'antibiotic that was most commonly used for mastitis at the start of the project' are not equal and there is a significant difference in percentage of samples containing tetracycline resistant E. coli among 'antibiotic that was most commonly used for mastitis at the start of the project' factor levels.

Results

The null hypothesis was rejected in favor of my hypothesis that the percentage of samples for a given farm found to contain tetracycline resistant E. coli is correlated with 'Amount of tetracycline used in mg/PCU.' by the Pearson's product-moment correlation test: ($\rho \neq 0$)(t = 41.368, df = 51, p-value < 2.2e-16, $\alpha = 0.05$, 95% CI: 0.9747620 0.9916009) I also hypothesized a positive correlation which is consistent with the results of the Pearson's product-moment correlation test: ($\rho = 0.985424$)

The ANOVA test failed to reject the null hypothesis that the percentage of samples for a given farm found to contain tetracycline resistant E. coli is significantly different among 'first choice/most commonly used class of antibiotic for clinical mastitis' levels: (mean percentage of samples for a given farm found to contain tetracycline resistant E. are equal among 'first choice/most commonly used class of antibiotic for clinical mastitis' levels)(F-statistic = 1.748, p-value = 0.155, $\alpha = 0.05$, df = 52). This result is inconsistent with my given hypothesis.

The ANOVA test rejected the null hypothesis in favor of my hypothesis that the percentage of samples for a given farm found to contain tetracycline resistant E. coli is significantly different among 'antibiotic most commonly used for mastitis at the start of the project' levels: (at least one mean percentage of samples for a given farm found to contain tetracycline resistant E. is not equal among 'antibiotic most commonly used for mastitis at the start of the project' levels)(f = 3.369, p-value = 0.00775, $\alpha = 0.05$, df = 52).

Discussion

My results suggest that use of Orbenin LA may be a cause of prevalent tetracycline resistant E. coli while use of Mastiplan LC and Penicillimoxycillin may be a cause of scarce tetracycline resistant E. coli; However, each of these antibiotics were not prevalent compared to the other antibiotics which is an indication that the results may be a consequence of small sample size or other farm management factors. My results also suggest that amount of tetracycline used in mg/PCU may be a cause of prevalence of samples for a given farm found to contain tetracycline resistant E. coli. This is a very logical conclusion considering the large correlation coefficient and that bacteria are known to become resistant to antibiotics used against them; However, correlation does not imply causation.

In the future, I would choose to analyze 'first choice/most commonly used class of antibiotic for clinical mastitis' and 'antibiotic most commonly used for mastitis at the start of the project' levels given more farms with preference of the antibiotics that had fewer observations. I would also take data from no tetracycline use to several increments of tetracycline used in mg/PCU and run two-sample T-tests between the increments and no tetracycline use data to determine at what increment percentage of samples for a given farm found to contain tetracycline resistant E. coli becomes significantly different. I would also use a chi-squared tests in the future to account for different farms as a factor in addition to my treatments.

References

```
citation(package = 'readr')
```

```
##
## To cite package 'readr' in publications use:
##
##   Wickham H, Hester J, Bryan J (2022). _readr: Read Rectangular Text
##   Data_. R package version 2.1.3,
##   <https://CRAN.R-project.org/package=readr>.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {readr: Read Rectangular Text Data},
##     author = {Hadley Wickham and Jim Hester and Jennifer Bryan},
##     year = {2022},
##     note = {R package version 2.1.3},
##     url = {https://CRAN.R-project.org/package=readr},
##   }
```

```
citation(package = 'here')
```

```
##
## To cite package 'here' in publications use:
##
##   Müller K (2020). _here: A Simpler Way to Find Your Files_. R package
##   version 1.0.1, <https://CRAN.R-project.org/package=here>.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {here: A Simpler Way to Find Your Files},
##     author = {Kirill Müller},
##     year = {2020},
##     note = {R package version 1.0.1},
##     url = {https://CRAN.R-project.org/package=here},
##   }
```

```
citation(package = 'ggplot2')
```

```
##
## To cite ggplot2 in publications, please use:
##
##   H. Wickham. ggplot2: Elegant Graphics for Data Analysis.
##   Springer-Verlag New York, 2016.
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     author = {Hadley Wickham},
##     title = {ggplot2: Elegant Graphics for Data Analysis},
```

```
##     publisher = {Springer-Verlag New York},
##     year = {2016},
##     isbn = {978-3-319-24277-4},
##     url = {https://ggplot2.tidyverse.org},
##   }
```

```
citation(package = 'car')
```

```
##
## To cite the car package in publications use:
##
##   John Fox and Sanford Weisberg (2019). An {R} Companion to Applied
##   Regression, Third Edition. Thousand Oaks CA: Sage. URL:
##   https://socialsciences.mcmaster.ca/jfox/Books/Companion/
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     title = {An {R} Companion to Applied Regression},
##     edition = {Third},
##     author = {John Fox and Sanford Weisberg},
##     year = {2019},
##     publisher = {Sage},
##     address = {Thousand Oaks {CA}},
##     url = {https://socialsciences.mcmaster.ca/jfox/Books/Companion/},
##   }
```

```
citation(package = 'tidyverse')
```

```
##
## To cite package 'tidyverse' in publications use:
##
##   Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R,
##   Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller
##   E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V,
##   Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to
##   the tidyverse." _Journal of Open Source Software_, *4*(43), 1686.
##   doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {Welcome to the {tidyverse}},
##     author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostin
##     year = {2019},
##     journal = {Journal of Open Source Software},
##     volume = {4},
##     number = {43},
##     pages = {1686},
##     doi = {10.21105/joss.01686},
##   }
```

```
citation(package = 'psych')
```

```
##
## To cite the psych package in publications use:
##
##   Revelle, W. (2022) psych: Procedures for Personality and
##   Psychological Research, Northwestern University, Evanston, Illinois,
##   USA, https://CRAN.R-project.org/package=psych Version = 2.2.9.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {psych: Procedures for Psychological, Psychometric, and Personality Research},
##     author = {William Revelle},
##     organization = { Northwestern University},
##     address = { Evanston, Illinois},
##     year = {2022},
##     note = {R package version 2.2.9},
##     url = {https://CRAN.R-project.org/package=psych},
##   }
```

Schubert H, Morley K, Puddy EF et al. Reduced Antibacterial Drug Resistance and blaCTX-M β-Lactamase Gene Carriage in Cattle-Associated Escherichia coli at Low Temperatures, at Sites Dominated by Older Animals, and on Pastureland: Implications for Surveillance. Applied and Environmental Microbiology 2021; 87: e01468-20

Galindo-Méndez, Mario. "Antimicrobial Resistance in Escherichia coli". E. Coli Infections - Importance of Early Diagnosis and Efficient Treatment, edited by Luis Rodrigo, IntechOpen, 2020. 10.5772/intechopen.93115.