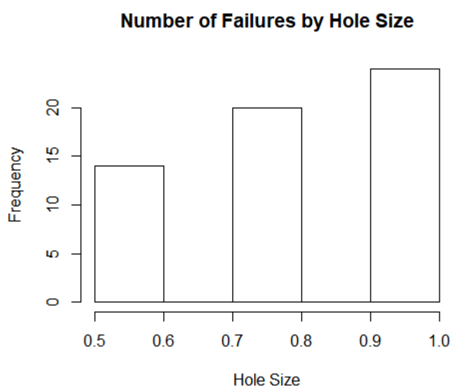
FINAL PROJECT

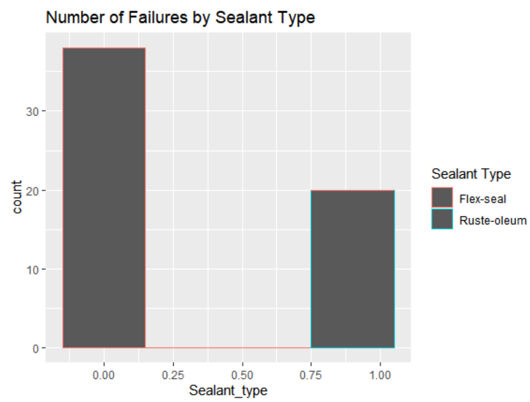
Romnikh Ortega

APRIL 14, 2020 Statistics 466

Exploratory Data Analysis

Based on the histogram below, there is evidence that hole size has an effect on failure rate since cups with smaller holes failed less often that cups with bigger holes. Also note that flex seal seems to be less reliable than Ruste-oleum since it had more failures.





Model for Data Analysis

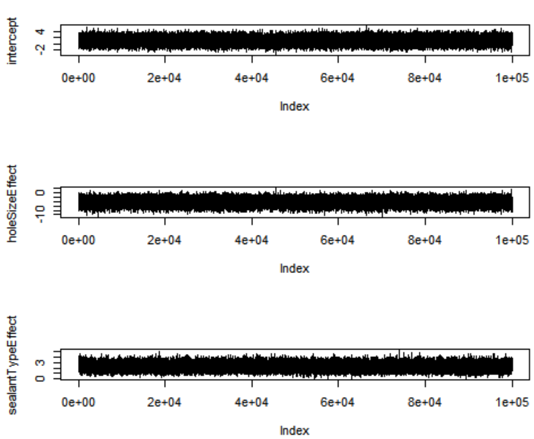
$$g(\pi) = \beta_0 + \beta_1 \times Hole\_size + B_2 \times Sealant\_Type$$

Parameter Definitions:

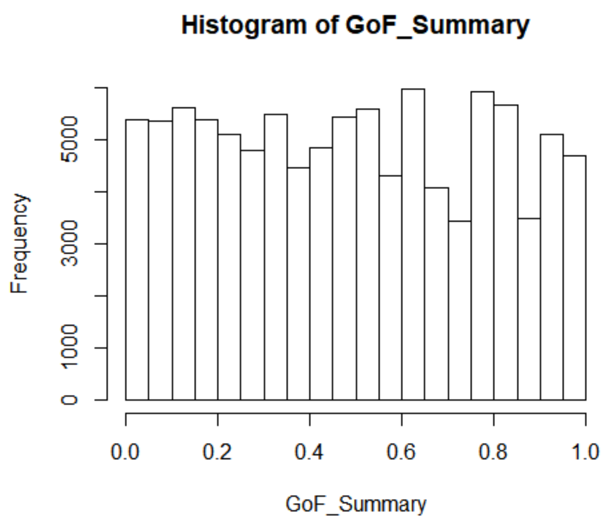
- π – Vector of probabilities
- g – link function, in our case it was a logit link.
- $\beta_{0,1,2}$ – Beta Parameter estimates for the explanatory variable.

Convergence and Goodness of fit of model

The trace plots below suggest that the MCMC samples have converged. Since the upper limit for the Gelman Diagnostic test was one, we have further evidence that the MCMC chains have converged.



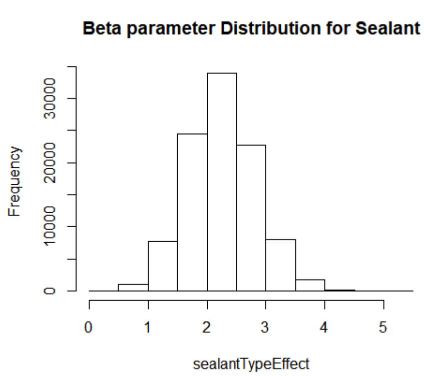
Note that the bins in the histogram below are for the most part, uniform in height. This combined with the fact that the proportion of posterior models with a p-value less than 0.05 was 0.05375 suggests that the model is a very good fit for the data.



Data Analysis

Research Question 1: Is one sealant more reliable when compared to another?

Note that 0 is not contained in the beta parameter distribution for β_2 below and zero is also not in interval (1.18, 3.44) which represents the 95% credible interval for the true posterior mean value of β_2. This implies that sealant has an effect on the probability of failure when it comes to sealing leaks.

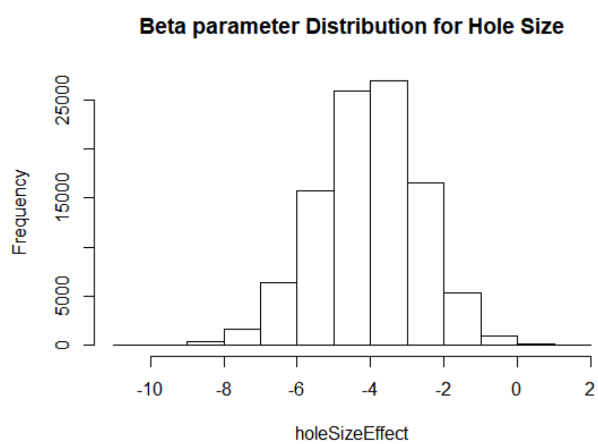


The table below represents the posterior predictive probability that a particular sealant will successfully seal a hole of a certain size. Based on this table, its clear that the Ruste-oleum sealant is expected to perform much better than Flex-seal. Hence, when it comes to sealing leaks, Ruste-oleum should be the sealant of choice.

Hole Size	Sealant Type	Probability of Success Estimate
0.5	Flex-seal	0.28
0.75	Flex-seal	0.13
1.0	Flex-seal	0.06
0.5	Ruste-oleum	0.77
0.75	Ruste-oleum	0.57
1.0	Ruste-oleum	0.33

Research Question 2: Does hole size have an effect on failure rate? If there is an effect, Interpret it.

Note that 0 is not contained in beta parameter distribution for β_1 and it's not in the interval (-6.88,-1.43) which represents the 95% credible interval for the true posterior mean value of β_1. This implies that the size of the hole has an effect on the probability of failures when it comes to sealing leaks.



Using the equation $\frac{1}{1+e^{-x^t\beta}}$, I calculated the expected probability of success for plugging a cup with hole sizes 0.5 and 1 inch respectively. By subtracting the result, I obtained the value 0.226 which will aid in interpreting the value of β_1 . In R this was accomplished using the following code:

```
mean(ilogit(intercept + holeSizeEffect*0.5)) - mean(ilogit(intercept + holeSizeEffect*1.0))
```

In the context of the project, the value above means that holding all else constant, we expect a decline in the probability of the success of plugging a leak by 0.226 for every 0.5 unit increase in hole size.

R-appendix

```
#####################
##Required Library##
#####################
library(R2jags)
library(ggplot2)
#####################
##Required Functions#
#####################
GoF_Test = function(fitted_quantiles) {
  n  = length(fitted_quantiles)
  K  = round((n)^(0.4))
  mK = table(cut(fitted_quantiles,(0:K)/K))
  np = n/K
  RB = sum(((mK-np)^2)/np)
  return(1-pchisq(RB,K-1))
}
######################
#Reads in Data#######
####################
cupsData = read.table("C:/Users/rpo22/Documents/466Files/data/cupsData.txt",
                      header = TRUE,sep = ",")
####################
#Data preparation###
####################
cupsData$Sealant_type = as.character(cupsData$Sealant_type)
cupsData$Seal[cupsData$Sealant_type == "F"] = "Flex-seal"
cupsData$Seal[cupsData$Sealant_type == "R"] = "Ruste-oleum"
cupsData$Sealant_type[cupsData$Sealant_type == "F"] = "0"
cupsData$Sealant_type[cupsData$Sealant_type == "R"] = "1"
cupsData$Sealant_type = as.integer(cupsData$Sealant_type)
###########
###EDA#####
###########
failures = cupsData[cupsData$Result == 0 ,]
successes = cupsData[cupsData$Result == 1 ,]
hist(failures$Hole_Size, main = "Number of Failures by Hole Size", xlab = "Hole Size")
```

```
hist(failures$Sealant_type, main = "Number of Failures by Sealant type", xlab = "Sealant
Type")
legend("topright", c("Flex-seal", "Ruste-oleum"), lwd=1)

ggplot(failures, aes(x = Sealant_type,color = factor(Seal))) +
  geom_histogram(binwidth = 0.3) +
  scale_color_discrete(name = "Sealant Type") +
  ggtitle("Number of Failures by Sealant Type")

# Setting up the variables we will use in JAGS
n =  length(cupsData[,1])
yFailure = cupsData[,3]
sealantType = cupsData[,4]
holeSize = cupsData[,2]

# Defines the logit and inverse logit functions
logit = function(x) {log(x/(1-x))}
ilogit = function(x) {1/(1+exp(-x))}
#Setting up the Model
cupModel = "model {
  for(i in 1:n){
    yFailure[i] ~ dbin(pi[i],1)
    logit(pi[i]) = beta[1] + beta[2]*holeSize[i] + beta[3]*sealantType[i]
  }
  beta[1] ~ dnorm(0,1/1000)
  beta[2] ~ dnorm(0,1/1000)
  beta[3] ~ dnorm(0,1/1000)
}
"
#Performs MCMC Simulation
cupSim = jags(
  data=c('yFailure','sealantType','holeSize','n'),
  parameters.to.save=c('beta'),
  model.file=textConnection(cupModel),
  n.iter=21000,
  n.burnin=1000,
  n.chains=5,
  n.thin=1
)
############################
##Checking For Convergence#
```

```
###########################
head(cupSim$BUGSoutput$sims.matrix)
intercept  = cupSim$BUGSoutput$sims.matrix[,1]
holeSizeEffect  = cupSim$BUGSoutput$sims.matrix[,2]
sealantTypeEffect = cupSim$BUGSoutput$sims.matrix[,3]
#Checks mixing: Mixing looks good
par(mfrow=c(3,1))
plot(intercept,type="l")
plot(holeSizeEffect ,type="l")
plot(sealantTypeEffect,type="l")
par(mfrow=c(1,1))
#Effective sample size: Effective sample size is the same as obtained sample
effectiveSize(cupSim$BUGSoutput$sims.matrix[,1])
effectiveSize(cupSim$BUGSoutput$sims.matrix[,2])
effectiveSize(cupSim$BUGSoutput$sims.matrix[,3])
#Gelman Diagnostic Test:
gelman.diag(cupSim$BUGSoutput)
########################
##Goodness of fit test##
#######################
GoF = matrix(NA,ncol=length(yFailure),nrow=length(intercept))
for (i in 1:length(intercept)) {
  GoF[i,] = runif(length(yFailure),
                 pbinom(yFailure-1,1,ilogit(intercept[i] + holeSizeEffect[i]*holeSize
                  + sealantTypeEffect[i]*sealantType)),
                pbinom(yFailure,1,ilogit(intercept[i] + holeSizeEffect[i]*holeSize
                  + sealantTypeEffect[i]*sealantType))
  )
}
# Calculating the p-values for each posterior model
GoF_Summary = apply(GoF,1,GoF_Test)
# Histogram of posterior model p-values
hist(GoF_Summary,xlim=c(0,1))
# Percent of posterior models with p-value less than 0.05
mean(GoF_Summary < 0.05)
####
#Questions, What to include in Data Analysis.
#Research Question 1
quantile(sealantTypeEffect ,c(0.025,0.975))
hist(sealantTypeEffect, main = "Beta parameter Distribution for Sealant")
########
```

```
#PPD###
#######
piFlexSeal = ilogit(intercept + holeSizeEffect*1 + sealantTypeEffect*0)
PPDFlex  =  rbinom(length(piFlexSeal),1,piFlexSeal)
piRust = ilogit(intercept + holeSizeEffect*1 + sealantTypeEffect*1)
PPDRust = rbinom(length(piRust),1,piRust)
sum(PPDRust)/length(PPDRust)
sum(PPDFlex)/length(PPDFlex)
#Research Question 2
quantile(holeSizeEffect,c(0.025,0.975))
hist(holeSizeEffect , main = "Beta parameter Distribution for Hole Size")
#Interpretation of B1
mean(ilogit(intercept + holeSizeEffect*0.5)) -
mean(ilogit(intercept + holeSizeEffect*1.0))
```