McKade Thomas

Stat466 Project Report

Dr. Richard Warr

Which Paper Towel Brand Really is the Strongest?

Introduction to the Problem

Every brand of paper towel claim to be the best, the strongest, and the most absorbent. Bounty, for example, claims to be the most awarded of any paper towel brand. With so many claims and choices which one should you pick? This report will analyze the strength of three paper towel brands to investigate this question.

Questions to Tackle

More specific than just which brand should we pick, this study investigated the following:

- Which paper towel brand holds the most weight before tearing?
- What affect does water have on the strength of the towel?
- Do some brands perform better than others overall or just under certain conditions?
- Based on the results from the previous questions, which paper towel brand should we buy?

Collection of the Data

In collecting the data, three paper towel brands were tested: Costco's store brand, Sam's Club's store brand, and Bounty. Thus, brand was the first factor to be considered. The second factor that would help address our questions was water. All three brands of towels were tested 5 times without water (dry) and 5 times with the same amount of water (wet).

The test itself consisted of adding weight (in grams) slowly to the middle of one section (section constituting the area between two perforated edges) of the towel while it was clipped in place at either end. A failure occurred when the towel tore or could no longer support the weight. Upon failure, the amount of weight the towel was holding when the failure occurred was recorded.

To collect the data, many controls needed to put in place. Below is a list of these controls that helped ensure consistency across all replications and factors:

- The towel was placed curved side up for each replicate
- Each trial started with a 194-gram glass cup which was dried after each replicate
- 1-inch lines were drawn on either end of the towel so that weight placement was aligned for each replicate
- The water temperature was kept constant (about 72 degrees F) for each water replicate
- Same orientation of the clips (45 degrees) the held the towel in place on either end for each replicate
- Each towel was saturated for 30 seconds for each water replicate
- Flow of water was consistent across all replicates

- Water placement was the same for each replicate (there was minor drippage)
- Randomization was used to determine the order in which trials would be conducted

Initial Exploratory Data Analysis

A table of the data can be found in appendix 1. After collecting the data, a brief exploration was conducted to see if assumptions were correct. Some initial hypotheses about the data would be that Bounty performed best out of the three paper towels (held the most weight) both wet and dry and that all three paper towel brands would perform better when dry.

Initially, the data revealed that, on average, Costco paper towels held about 251.8 more grams when wet, Sam's Club paper towels held 314 more grams when wet, and Bounty paper towels held 102.6 more grams when wet. This was extremely interesting and a direct contradiction to our original hypotheses. At least initially, it appears that the towels had more strength when water was added, not less.

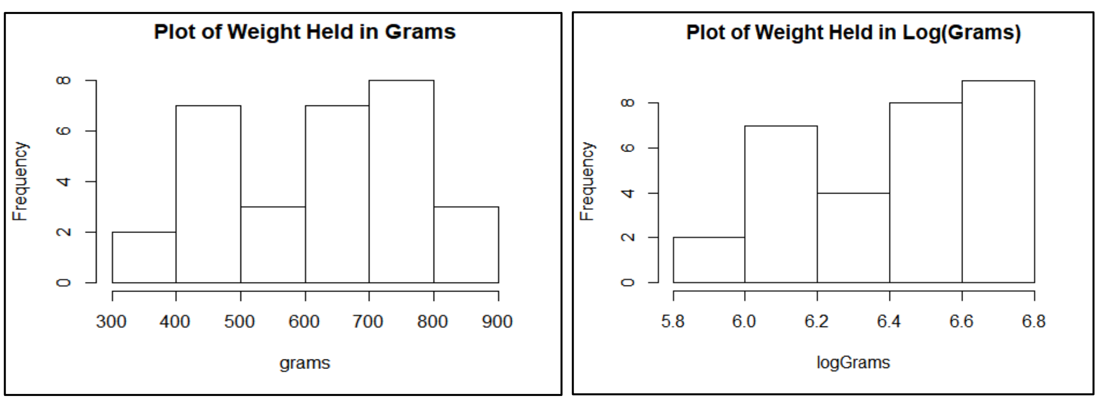
The data also seemed to suggest that when dry, Bounty paper towels seemed to perform the best (mean: 621.6g) followed by Sams (mean: 472g), then Costco (mean: 441g). However, after water was added to the towels Sams seemed to perform the best (mean: 786g) followed by Bounty (mean: 724.2g), and Costco last (mean: 692.8). Thus, when dry Bounty paper towels seemed to hold the most weight while Sams outperformed the rest when wet.

Analyze the Data Using a Gibbs Sampler

1. <u>Selecting a Likelihood and Priors</u>

In selecting a likelihood model for the data, a lognormal distribution seemed to be the best fit (compared with generalized gamma, gamma, Weibull, and exponential) with log(Grams) being used for the outcome. This model also had the lowest DIC of any other model considered (-22.3692). Figure 1 illustrates the difference created by logging the data:

Figure 1 – Comparison of Logged and Un-Logged Data



Diffuse normal priors were used with a mean of 0 and precision of 1/100 for all covariates including the interaction terms. This was done since the amount of information we had prior to assessing the model was very limited. In this manner, the priors will not weigh into the model as much as the likelihood.

2. Fitting a Model

To further analyze the data, the following model was created in JAGS:

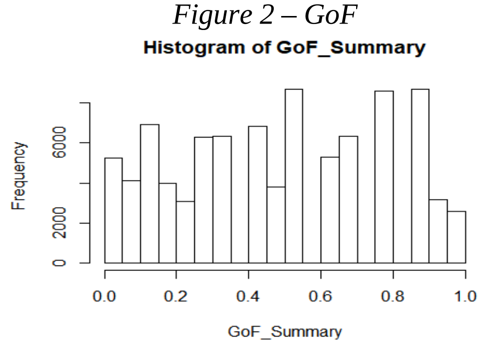
> *logGrams[i] ~ dnorm(mu[i],1/(sigma)^2)*
> *mu[i] <- beta0 + beta1*water[i] + beta2*Sams[i] + beta3*Bounty[i] + beta4*water[i]*Sams[i] + beta5*water[i]*Bounty[i]*

Here beta0 represents dry Costco paper towels which were served as a base. From there, the effect of water and each of the brands (beta1, beta2, and beta3) as well as the interaction effects (beta4 and beta5) were considered.

The model used 50,000 iterations with 3 chains and a burn-in of 2,000. After checking for convergence and mixing of the data, all diagnostics looked good and indicated that the algorithm was converging to the correct values and was mixing well. The Gelman diagnostic for the model was 1 for both the point estimate and upper C.I. which was a further indication that model was converging correctly. For a further breakdown of the diagnostics, see Appendix 2.

A Goodness of Fit test was conducted to assess how well the model above appeared to fit the data. The results of the goodness of fit indicated that about 5.9% of p-values would've been rejected for this model. Because this diagnostic should be around 5%, this is a great indication that the model was a good fit for the data. See Appendix 2 for code. Figure 2 is a histogram of the GoF test (note that the shape is fairly uniform which is another indication of good fit:

Figure 2 – GoF



3. Analyzing the Model

After finding that the model was a good fit for the data, an analysis was conducted first on each of the covariates to determine their significance in the model. Histograms of each covariate can be found in Appendix 2. Table 1 lists the probability associated with each beta value in the model. These probabilities suggest that beta1, beta3, and beta5 had the greatest effect on weight held. This means that water as well as the Bounty brand and the interaction between Bounty and brand seem to have the greatest difference in determining the strength of the paper towels.

Table 1: Model Covariate Probabilities

Covariate	Probability Pt. Estimate
Beta1	99.99%
Beta2	81.47%
Beta3	99.97%

Beta4	33.86%
Beta5	98.70%

4. Assessing Posterior Predictive Probabilities

Comparing our results to those we found prior to the model, we find first that all three paper towels performed better when wet. Sams towels had the biggest difference on average (MSTF: 316.17g) followed by Costco (MSTF: 253.42g), then Bounty (MSTF: 106.06g). Bounty seemed to be much less affected by water overall while Sams towels could hold much more weight when wet.

The previous results seem to answer the question of the affect of water on strength of paper towels, at least for these three brands. The next question then is which brand is best? Continuing our analysis of the PPD's for the model, on average, Bounty paper towels performed the best when dry (MSTF: 632.54g) followed by Sams (MTSF: 477.42g), then Costco (MTSF: 440.23g). Bounty, when dry, also had the greatest variance. While Bounty towels performed the best when dry however, Sam's paper towels performed the best when wet (MTSF: 793.59g) followed by Bounty (MTSF: 729.60), then Costco (MTSF: 693.65g). See below for the posterior predictive distributions of each brand both when dry (Figure 3) and wet (Figure 4).

Figure 3 – PPD's for Dry Paper Towels

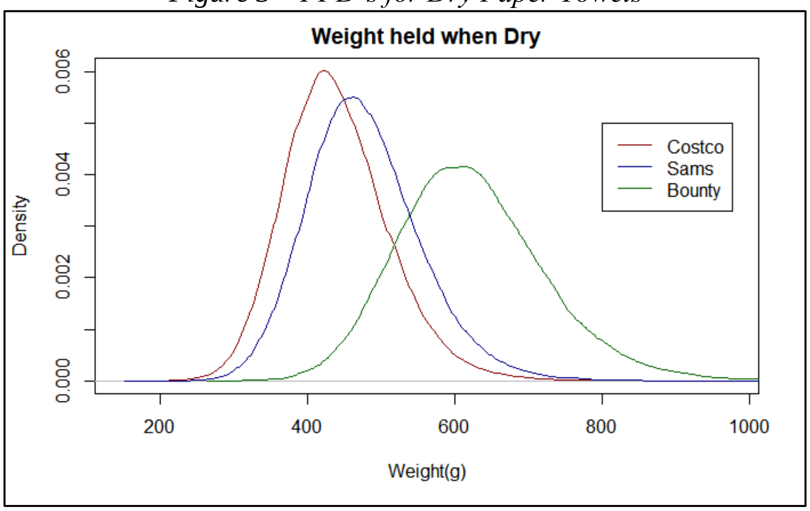
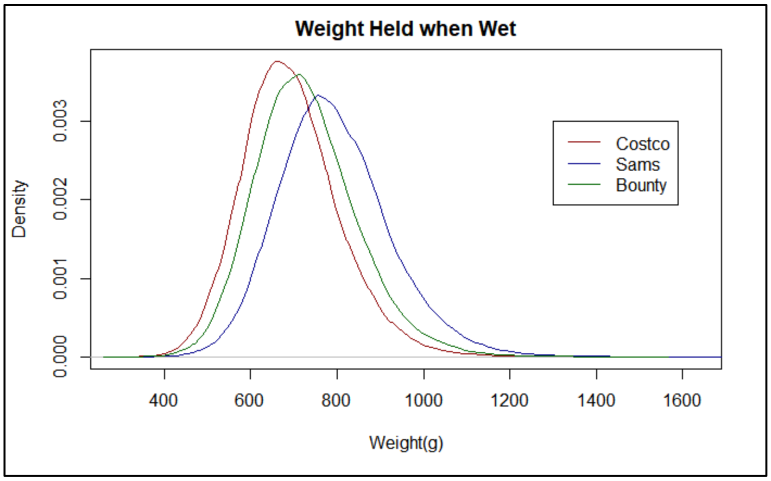


Figure 4 – PPD's for Wet Paper Towels

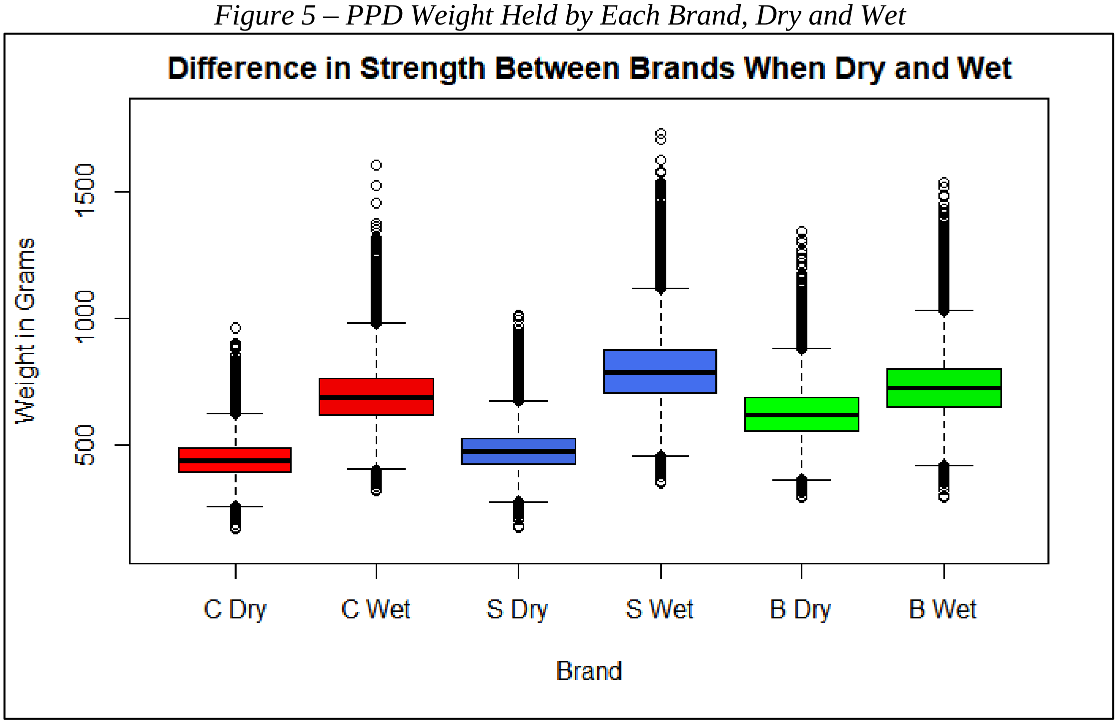


Conclusion

Overall, it appears that despite our previous beliefs about the strength of paper towels, they can hold more weight when wet than dry. The reason for this is not yet clear. Perhaps adding more replicates would illustrate a clearer pattern as to when and why the towels hold more weight on average when wet.

Also, while Bounty paper towels tend to hold more weight on average when dry, Sams paper towels held more on average when wet. This was also a surprising conclusion since Bounty claims to be the most awarded of any other paper towel brand. Figure 5 gives a final illustration of each brand, dry and wet (note that the plot does not show one outlier for Sams and that the second plot is for wet paper towels).

Figure 5 – PPD Weight Held by Each Brand, Dry and Wet



Appendix 1
Contents: Original Data Table

Data

Trial Number	ID	Brand	Water_Yes	Grams	Brand1_Y	Brand2_Y
7	1	Costco	1	601	0	0
9	2	Costco	1	633	0	0
30	3	Costco	1	747	0	0
5	4	Costco	1	605	0	0
10	5	Costco	1	878	0	0
23	6	Costco	0	583	0	0
13	7	Costco	0	463	0	0
17	8	Costco	0	396	0	0
2	9	Costco	0	406	0	0
8	10	Costco	0	357	0	0
19	11	Sams	1	758	1	0
6	12	Sams	1	888	1	0
20	13	Sams	1	758	1	0
14	14	Sams	1	735	1	0
15	15	Sams	1	791	1	0
29	16	Sams	0	471	1	0
21	17	Sams	0	515	1	0
26	18	Sams	0	452	1	0
25	19	Sams	0	482	1	0
4	20	Sams	0	440	1	0
1	21	Bounty	1	643	0	1
27	22	Bounty	1	799	0	1
16	23	Bounty	1	606	0	1

18	24	Bounty	1	764	0	1
3	25	Bounty	1	809	0	1
12	26	Bounty	0	635	0	1
28	27	Bounty	0	695	0	1
11	28	Bounty	0	477	0	1
22	29	Bounty	0	594	0	1
24	30	Bounty	0	707	0	1

Appendix 2

Contents – Code used to obtain results

```
#Subset the data
C_nw <- subset(all_paper, all_paper[,2]=="Costco" & all_paper[,3]==0)
C_w <- subset(all_paper, all_paper[,2]=="Costco" & all_paper[,3]==1)
S_nw <- subset(all_paper, all_paper[,2]=="Sams" & all_paper[,3]==0)
S_w <- subset(all_paper, all_paper[,2]=="Sams" & all_paper[,3]==1)
B_nw <- subset(all_paper, all_paper[,2]=="Bounty" & all_paper[,3]==0)
B_w <- subset(all_paper, all_paper[,2]=="Bounty" & all_paper[,3]==1)

#Difference in water Vs. no water
(C_initial_diff <- mean(C_nw$Grams) - mean(C_w$Grams))

## [1] -251.8

#Costco paper towels on average held 251.8 more grams when wet.
(S_initial_diff <- mean(S_nw$Grams) - mean(S_w$Grams))

## [1] -314

#Sams paper towels on average held 314 more grams when wet.
(B_initial_diff <- mean(B_nw$Grams) - mean(B_w$Grams))

## [1] -102.6

#Bounty paper towels on average held 102.6 more grams when wet.


#Difference in Brands
#No water
c("Costco w/out water:", mean(C_nw$Grams),
  "Bounty w/out water:", mean(B_nw$Grams),
  "Sams w/out water:", mean(S_nw$Grams))
```

```
## [1] "Costco w/out water:" "441"                 "Bounty w/out water:"
## [4] "621.6"               "Sams w/out water:"   "472"

#For dry paper towels Sams Club performed best.

#No water
c("Costco w/ water:", mean(C_w$Grams),
  "Bounty w/ water:", mean(B_w$Grams),
  "Sams w/ water:", mean(S_w$Grams))

## [1] "Costco w/ water:" "692.8"            "Bounty w/ water:"
## [4] "724.2"            "Sams w/ water:"   "786"

#For wet paper towels Bounty performed best.


water <- all_paper[,3]
grams <- all_paper[,4]
Sams <- all_paper[,5]
Bounty <- all_paper[,6]

logGrams <- log(grams)
n <- length(grams)

hist(grams, main="Plot of Weight Held in Grams")

hist(logGrams,main="Plot of Weight Held in Log(Grams)")



##################
# Lognormal Model
library(R2jags)
#####
# Lognormal Model
Papermdl <- "model {
for(i in 1:n){
logGrams[i] ~ dnorm(mu[i],1/(sigma)^2)
mu[i] <- beta0 + beta1*water[i] + beta2*Sams[i] + beta3*Bounty[i] + beta4*wat
er[i]*Sams[i] + beta5*water[i]*Bounty[i]
}
sigma ~ dexp(1)
beta0 ~ dnorm(0,1/100)
beta1 ~ dnorm(0,1/100)
beta2 ~ dnorm(0,1/100)
beta3 ~ dnorm(0,1/100)
beta4 ~ dnorm(0,1/100)
beta5 ~ dnorm(0,1/100)
PPDCostcoNW ~ dnorm(beta0,1/(sigma)^2)
PPDCostcoW ~ dnorm(beta0+beta1,1/(sigma)^2)
PPDSamsNW ~ dnorm(beta0+beta2,1/(sigma)^2)
```

```
PPDSamsW ~ dnorm(beta0+beta1+beta2+beta4,1/(sigma)^2)
PPDBountyNW ~ dnorm(beta0+beta3,1/(sigma)^2)
PPDBountyW ~ dnorm(beta0+beta1+beta3+beta5,1/(sigma)^2)
}
"

water.sim <- jags(
  data=c('logGrams','n','water','Sams','Bounty'),
  parameters.to.save=c('beta0','beta1','beta2','beta3','beta4','beta5','sigma
','PPDCostcoNW','PPDCostcoW',
                       'PPDSamsNW','PPDSamsW','PPDBountyNW','PPDBountyW'),
  model.file=textConnection(Papermdl),
  n.iter=20000,
  n.burnin=2000,
  n.chains=5,
  n.thin=1
)

## module glm loaded

## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 30
##    Unobserved stochastic nodes: 13
##    Total graph size: 166
##
## Initializing model

#Assigning Variables
PPDB_nw <- water.sim$BUGSoutput$sims.matrix[,1]
PPDB_w <- water.sim$BUGSoutput$sims.matrix[,2]
PPDC_nw <- water.sim$BUGSoutput$sims.matrix[,3]
PPDC_w <- water.sim$BUGSoutput$sims.matrix[,4]
PPDS_nw <- water.sim$BUGSoutput$sims.matrix[,5]
PPDS_w <- water.sim$BUGSoutput$sims.matrix[,6]
plot(PPDB_nw,type="l")

beta0 <- water.sim$BUGSoutput$sims.matrix[,7]
beta1 <- water.sim$BUGSoutput$sims.matrix[,8]
beta2 <- water.sim$BUGSoutput$sims.matrix[,9]
beta3 <- water.sim$BUGSoutput$sims.matrix[,10]
beta4 <- water.sim$BUGSoutput$sims.matrix[,11]
beta5 <- water.sim$BUGSoutput$sims.matrix[,12]
sigma <- water.sim$BUGSoutput$sims.matrix[,14]

#Check mixing
plot(PPDB_nw,type="l")

plot(PPDB_w,type="l")
```

```
plot(PPDC_nw,type="l")

plot(PPDC_w,type="l")

plot(PPDS_nw,type="l")

plot(PPDS_w,type="l")

plot(beta0,type="l")

plot(beta1,type="l")

plot(beta2,type="l")

plot(beta3,type="l")

plot(beta4,type="l")

plot(beta5,type="l")

plot(sigma,type="l")

acf(beta0)

acf(beta1)

acf(beta2)

acf(beta3)

acf(beta4)

acf(beta5)

acf(sigma)

#Check convergence and model fit
gelman.diag(water.sim$BUGSoutput)

## Potential scale reduction factors:
##
##              Point est. Upper C.I.
## beta0                 1          1
## beta1                 1          1
## beta2                 1          1
## beta3                 1          1
## beta4                 1          1
## beta5                 1          1
## deviance              1          1
## PPDBountyNW           1          1
## PPDBountyW            1          1
## PPDCostcoNW           1          1
## PPDCostcoW            1          1
## PPDSamsNW             1          1
```

```
## PPDSamsW              1          1
## sigma                 1          1
##
## Multivariate psrf
##
## 1
```

```
water.sim$BUGSoutput$DIC
```

```
## [1] -22.3692
```

Check Goodness of Fit

```
#######################
# Check Model Fit w/ the GoF Test found in Section 3.4

# Calculating the fitted quantiles for each posterior model
# and adjusting for a discrete model
GoF <- matrix(NA,ncol=length(logGrams),nrow=length(beta0))
for (i in 1:length(beta0)) {
  for (j in 1:length(water)) {
    GoF[i,j] <- pnorm(logGrams[j],(beta0[i] + beta1[i]*water[j] + beta2[i]*Sa
ms[j] + beta3[i]*Bounty[j] + beta4[i]*water[j]*Sams[j] + beta5[i]*water[j]*Bo
unty[j]),sd=sigma[i])
  }
}

# Function requires fitted quantiles and returns a p-value
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))
}

# 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)
```

```
## [1] 0.05846667
```

```
#About 4.9% of p-values would be rejected which gives us greta indication tha
t this model
#   is a good fit for the data!
```

Analysis of the Model

```
#############################################
#Assessing the signficance of each covariate in the model
hist(beta1)

mean(beta1>0)

## [1] 0.9999778

#Prob: 0.780

hist(beta2)

mean(beta2>0)

## [1] 0.8146778

#Prob: .823

hist(beta3)

mean(beta3>0)

## [1] 0.9997444

#Prob: .835

hist(beta4)

mean(beta4<0)

## [1] 0.3386111

#Prob: .803

hist(beta5)

mean(beta5<0)

## [1] 0.9870222

#Prob: 0.663

#Based on these probabilities, it appears that beta2, beta3, and beta4 have t
he most significance in the model.
#   This suggests that Sams and Bounty as well as the interaction between wat
er and Sams have the most effect
#     on determining strength of paper towels/weight that they can hold.



##############################################
```

```r
#Checking out the Posterior Predictive Models
#Unlog the data
EPPDC_nw <- exp(PPDC_nw)
EPPDC_w <- exp(PPDC_w)
EPPDS_nw <- exp(PPDS_nw)
EPPDS_w <- exp(PPDS_w)
EPPDB_nw <- exp(PPDB_nw)
EPPDB_w <- exp(PPDB_w)


mean(EPPDC_nw)

## [1] 440.2316

mean(EPPDC_w)

## [1] 693.6467

mean(EPPDS_nw)

## [1] 477.4166

mean(EPPDS_w)

## [1] 793.5913

mean(EPPDB_nw)

## [1] 623.5446

mean(EPPDB_w)

## [1] 729.6001

hist(EPPDC_nw)

hist(EPPDC_w)

hist(EPPDS_nw)

hist(EPPDS_w)

hist(EPPDB_nw)

hist(EPPDB_w)

#All plots look fairly normal with a little bit of right skew.

##################################
#Difference in Water Vs. No water
(C_water_diff <- mean(EPPDC_nw) - mean(EPPDC_w))

## [1] -253.4151
```

```
#The paper towels from Costco held about 253.4 more grams on average when wet
.
(S_water_diff <- mean(EPPDS_nw) - mean(EPPDS_w))

## [1] -316.1747

#The paper towels from Sams Club held about 316.2 more grams on average when
wet.
(B_water_diff <- mean(EPPDB_nw) - mean(EPPDB_w))

## [1] -106.0555

#The Bounty paper towls held about 106.1 more grams on average when wet.

#All brands actually held more weight when wet.

#####################
#Difference in Brands
#No water
c("Costco w/out water:", mean(EPPDC_nw),
  "Sams w/out water:", mean(EPPDS_nw),
  "Bounty w/out water:", mean(EPPDB_nw))

## [1] "Costco w/out water:" "440.231578809463"    "Sams w/out water:"
## [4] "477.416589833656"    "Bounty w/out water:" "623.544584985782"

#On average, Bounty paper towels performed the best when dry followed by Sams
, then Costco

#Water
c("Costco w/ water:", mean(EPPDC_w),
  "Sams w/ water:", mean(EPPDS_w),
  "Bounty w/ water:", mean(EPPDB_w))

## [1] "Costco w/ water:" "693.646669568635" "Sams w/ water:"
## [4] "793.591300385218" "Bounty w/ water:" "729.600106218221"

#On average, Sams paper towels performed the best when wet followed by Bounty
, then Costco

#In both cases, wet and dry, Bounty paper towels held the greatest amount of
weight before failing.


#90% Confidence intervals for mean Grams
c("Costco w/out water:", quantile(EPPDC_nw,c(0.05,0.95)),
  "Costco w/ water:", quantile(EPPDC_w,c(0.05,0.95)),
  "Sams w/out water:", quantile(EPPDS_nw,c(0.05,0.95)),
  "Sams w/ water:", quantile(EPPDS_w,c(0.05,0.95)),
  "Bounty w/out water:", quantile(EPPDB_nw,c(0.05,0.95)),
  "Bounty w/ water:", quantile(EPPDB_w,c(0.05,0.95)))
```

```
##                                            5%                    95% 
## "Costco w/out water:"    "333.738208654629"    "565.824366847741" 
##                                            5%                    95% 
##    "Costco w/ water:"      "525.3212703377"    "891.582021162635" 
##                                            5%                    95% 
##   "Sams w/out water:"    "361.978984183091"    "613.027385615106" 
##                                            5%                    95% 
##      "Sams w/ water:"    "602.018648303268"     "1020.6559172873" 
##                                            5%                    95% 
## "Bounty w/out water:"    "473.394225094506"    "800.950366531393" 
##                                            5%                    95% 
##    "Bounty w/ water:"    "552.897296611629"    "938.393305112231"
```

Graphics to illustrate Relationships between Covariates

```
#Distributions for dry
d1 <- density(EPPDC_nw)
d2 <- density(EPPDS_nw)
d3 <- density(EPPDB_nw)
plot(d1,col="darkred",xlab="Weight(g)",main="Weight Held when Dry")
lines(d2,col="darkblue")
lines(d3,col="darkgreen")
#Add a legend
legend(800, .005, legend=c("Costco", "Sams","Bounty"),
       col=c("darkred", "darkblue","darkgreen"),lty=1, cex=1)
```

```
#Distributions for Wet
d4 <- density(EPPDC_w)
d5 <- density(EPPDS_w)
d6 <- density(EPPDB_w)
plot(d4,col="darkred",xlab="Weight(g)",main="Weight Held when Wet")
lines(d5,col="darkblue")
lines(d6,col="darkgreen")
#Add a legend
legend(1300, .003, legend=c("Costco", "Sams","Bounty"),
       col=c("darkred", "darkblue","darkgreen"),lty=1, cex=1)
```

Final Graphics

```
#Differences in water Vs. no water
#Costco
boxplot(EPPDC_nw,EPPDC_w,names=c("Dry","Wet"),xlab="Brand",ylab="Grams",main=
"Difference for Costco")

#Sams Club
boxplot(EPPDS_nw,EPPDS_w,names=c("Dry","Wet"),xlab="Brand",ylab="Grams",main=
"Difference for Sams")
```

```
#Bounty
boxplot(EPPDB_nw,EPPDB_w,names=c("Dry","Wet"),xlab="Brand",ylab="Grams",main=
"Difference for Bounty")

#Difference in Brands
#No Water
boxplot(EPPDC_nw,EPPDS_nw,EPPDB_nw,names=c("Costco","Sams","Bounty"),xlab="Br
and",ylab="Grams",main="Difference in Brand when Dry")

#Water
boxplot(EPPDC_w,EPPDS_w,EPPDB_w,names=c("Costco","Sams","Bounty"),xlab="Brand
",ylab="Grams",main="Difference in Brand when Wet")

#All Together
boxplot(EPPDC_nw,EPPDC_w,EPPDS_nw,EPPDS_w,EPPDB_nw,EPPDB_w,at=c(1,2, 3,4, 5,6
),
        col=c("red","red2","blue","royalblue2","green","green2"),
        names=c("Cost Dry","Cost Wet","Sams Dry","Sams Wet","Bount Dry", "Bou
nt Wet"),xlab="Brand",
        ylab="Weight in Grams",main="Difference in Strength Between Brands Wh
en Dry and Wet")

#All Together Zoomed In
boxplot(EPPDC_nw,EPPDC_w,EPPDS_nw,EPPDS_w,EPPDB_nw,EPPDB_w,at=c(1,2, 3,4, 5,6
),
        col=c("red","red2","royalblue","royalblue2","green","green2"),
        names=c("Costco Dry","Costco Wet","Sams Dry","Sams Wet","Bounty Dry",
"Bounty Wet"),xlab="Brand",
        ylab="Weight in Grams",main="Difference in Strength Between Brands Wh
en Dry and Wet",ylim=c(100,1800))
```