

Final Project

Diptanshu Singh

5/7/2019

(1) Statistics and the Law

Data Sufficiency problem

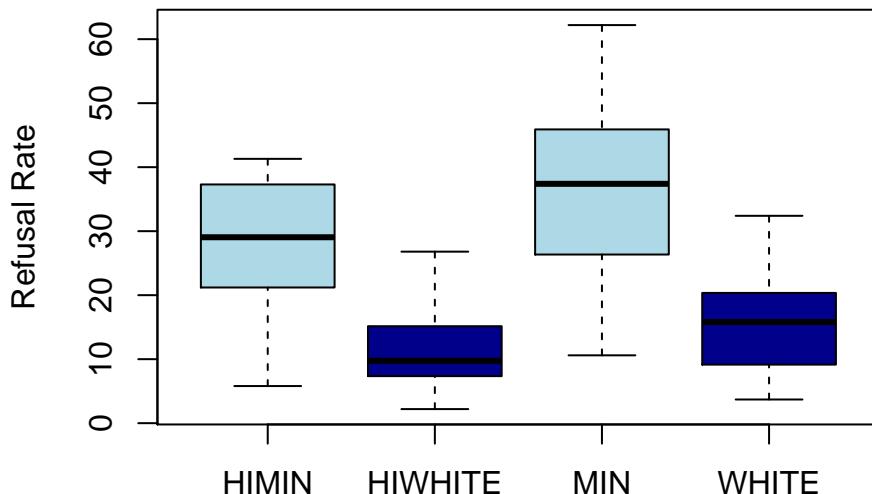
In order to get an initial look at the data, lets plot the data to get visual clues.

```
acorn <- read_csv("acorn.csv")

acorn_data <- acorn %>%
  gather("group", "refusal" = c(MIN, WHITE, HIMIN, HIWHITE))

boxplot(value ~ group, data = acorn_data,
        main = "Refusal Rates for Minorities and Whites", ylab = "Refusal Rate",
        col = c("lightblue", "darkblue"))
```

Refusal Rates for Minorities and Whites



We can see a clear difference in the means of the refusal rates for minorities and whites. To check if this data has enough evidence to show that there is discrimination in terms of refusal rates, we could do a statistical test.

Since the sample size is small, we will apply t-test and test for surprises.

Making the assumption that the rejections for minorities and white people across the banks are independent, we could perform independent 2-sample t-test to prove that there is sufficient evidence of discrimination. Additionally, the Z-test would be one-tailed since we are testing for the refusal rate in minorities being higher than the refusal rate for white people.

Our hypotheses would be as follows: *Null : $\mu_{min} > \mu_{white}$ Alternative : $\mu_{min} \leq \mu_{white}$*

The t-statistic we would need for this test is:

$$t = \frac{\mu_{min} - \mu_{white}}{\sqrt{\frac{s_{min}^2 + s_{white}^2}{n}}}$$

```
#mean of refusal rates for minorities
min_mean <- mean(acorn$MIN)
#mean of refusal rates for whites
white_mean <- mean(acorn$WHITE)

#standard deviation for refusal rates of minorities
min_sd <- sd(acorn$MIN)
#standard deviation for refusal rates of whites
white_sd <- sd(acorn$WHITE)

#calculating t-statistic
t <- (min_mean - white_mean)/sqrt((min_sd^2 + white_sd^2)/20)
t

## [1] 6.253257
```

The t-statistic is 6.2533, which is greater than 2. Hence the hypothesis can be rejected at level of significance = 0.05

Next we test if similar difference in refusal rates is also observed for the high income group as well

Our hypotheses would be as follows: Null : $\mu_{highIncomeMinority} > \mu_{HighIncomeWhite}$ Alternative : $\mu_{highIncomeMinority} \leq \mu_{HighIncomeWhite}$

The t-statistic we would need for this test is:

$$t = \frac{\mu_{highIncomeMinority} - \mu_{HighIncomeWhite}}{\sqrt{\frac{s_{highIncomeMinority}^2 + s_{HighIncomeWhite}^2}{n}}}$$

```
#mean of refusal rates for high income minorities
min_mean <- mean(acorn$HIMIN)
#mean of refusal rates for high income whites
white_mean <- mean(acorn$HIWHITE)

#standard deviation for refusal rates of high income minorities
min_sd <- sd(acorn$HIMIN)
#standard deviation for refusal rates of high income whites
white_sd <- sd(acorn$HIWHITE)

#calculating t-statistic
t <- (min_mean - white_mean)/sqrt((min_sd^2 + white_sd^2)/20)
t
```

[1] 5.701701

For the high income minorities and whites, the t-statistic is 5.7017 and the p-value is much smaller than 0.05. Therefore, we can reject the null hypothesis at level of significance = 0.05.

Additional Critique of the tests

Paired T-test

There are some additional constraints to think about. Since the data is paired; we should be performing paired t test. I have done the following test using the t-test function and not by calculating the t-test statistics.

```
t.test(acorn$MIN, acorn$WHITE, paired = TRUE, alternative = "greater")
```

```
##  
##  Paired t-test  
##  
## data: acorn$MIN and acorn$WHITE  
## t = 11.46, df = 19, p-value = 2.81e-10  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 18.04912      Inf  
## sample estimates:  
## mean of the differences  
##                      21.2565
```

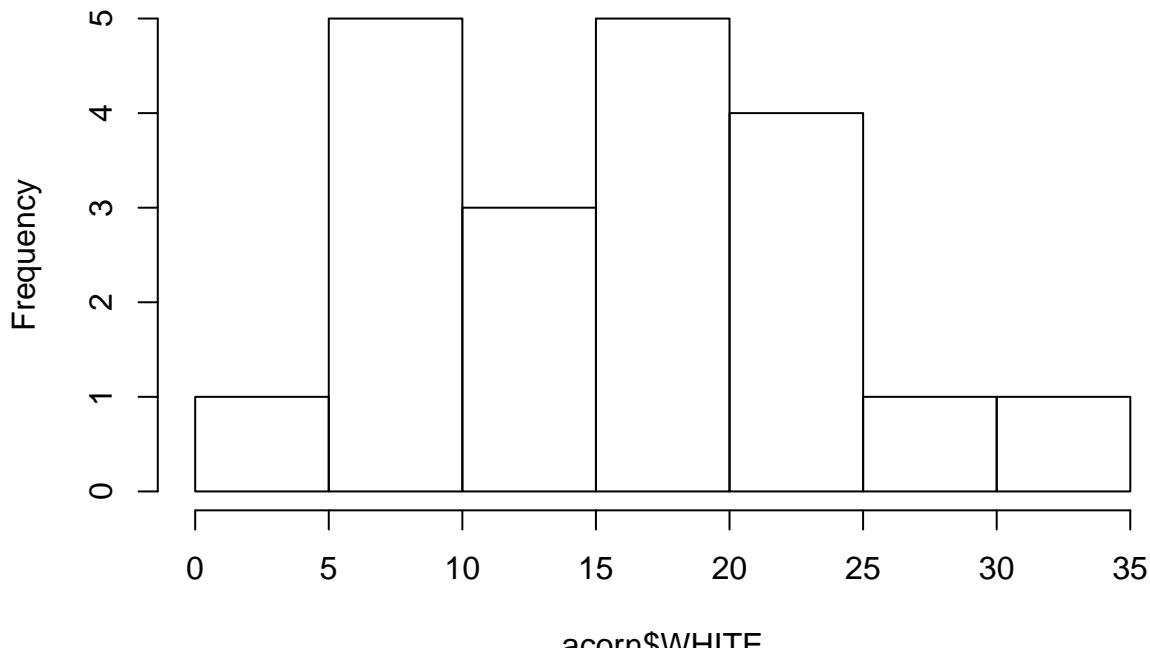
Even in the paired setting, the result is significant at level = 0.05

Normality of the data

We would also need to check for the normality assumption of the t-test.

```
hist(acorn$WHITE, breaks = 10)
```

Histogram of acorn\$WHITE



Although this kind of distribution can be generated from a normal distribution; just to be sure , I have also performed a two sample z-test.

```

z.test = function(a, b, var.a, var.b){
  n.a = length(a)
  n.b = length(b)
  zeta = (mean(a) - mean(b)) / (sqrt(var.a/n.a + var.b/n.b))
  return(zeta)
}

#calculating z-statistic
z.test(acorn$MIN, acorn$WHITE, var(acorn$MIN), var(acorn$WHITE))

```

[1] 6.253257

The Z-statistics is > 1.96 and hence the difference is still significant. ##### We can hence safely conclude that the data is sufficient for the analysis.

(2)Comparing Suppliers

Is there a difference between the schools in the quality of work they produce? We are comparing only the quality and not the revenue generated.

Making data frame from the question

	Dead	Display	Art
Area 51	12	23	89
BDV	8	12	62
Giffen	21	30	119

We will check if there is independence between the quality of birds and the schools. In an independent setting, all the probability in the cells can be written as product of marginal probability. However if there is significant difference between the actual and independent probaility, we can see that in the chi-square test.

```

## data
ex2

##      [,1] [,2] [,3]
## [1,]    12   23   89
## [2,]     8   12   62
## [3,]    21   30  119

## margins
m1 <- rowSums(ex2)
m1

## [1] 124 82 170
m2 <- colSums(ex2)
m2

## [1] 41 65 270
## Expected
ex2.expt <- outer(m1,m2, '*')/376 #376 is the sum of the columns and the rows
ex2.expt

##          [,1]      [,2]      [,3]
## [1,] 13.521277 21.43617 89.04255

```

```

## [2,] 8.941489 14.17553 58.88298
## [3,] 18.537234 29.38830 122.07447
## Normalized differences

ex2.dif <- (ex2 - ex2.expt)^2/ex2.expt

ex2.dif

## [,1]      [,2]      [,3]
## [1,] 0.17115858 0.11408585 2.033605e-05
## [2,] 0.09913362 0.33388088 1.650022e-01
## [3,] 0.32719100 0.01273226 7.743105e-02
## test stat

Q <- sum(ex2.dif)

Q

## [1] 1.300636
## Chi-sq

chi2 <- 1 - pchisq(q = Q, df = 4) # n-1 * n-1

chi2

## [1] 0.8612677
chisq.test(ex2)

##
## Pearson's Chi-squared test
##
## data: ex2
## X-squared = 1.3006, df = 4, p-value = 0.8613

```

From the results of the chi-squared test, we can see that the p-value associated with the chi-squared value is 0.9. This suggests that we cannot reject the null hypothesis that there is no difference in the quality of the ornithopters produced by the three schools. Hence we do not have sufficient data to answer this question.

(3) How deadly are sharks?

```

sharks <- read_csv("sharkattack.csv")
shark <- sharks[, -1]

Shark <- shark %>%
  subset(select = -c(1)) %>%
  filter(Country == "United States" | Country == "Australia")

sharkattack <- shark %>%
  filter(Country %in% c("United States", "Australia")) %>%
  group_by(Country, Fatal)%>%

```

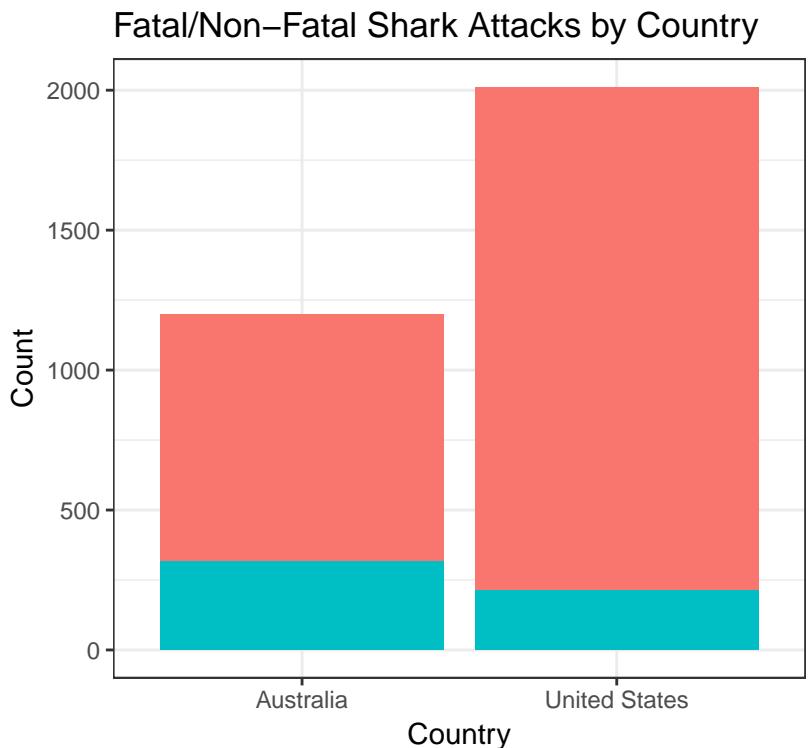
```

summarize(total = n())

sharkattack <- sharkattack[!sharkattack$Fatal %in% "UNKNOWN",]

ggplot(sharkattack)+ 
  aes(x = Country, y = total, fill = Fatal)+ 
  geom_bar(aes(fill = Fatal), stat = "identity")+
  labs(y = "Count", title = "Fatal/Non-Fatal Shark Attacks by Country")+
  theme_bw()

```



##EDA

2032 attacks were in the US, 1224 were in Australia 217 fatal attacks were in the US, 318 fatal attacks were in Australia So 10.6% of US attacks are fatal, 25.9% of Australian attacks are fatal

Statistical Testing 1

Are Australian sharks and American sharks equal in their deadliness?

Chi Square Test

```

Australia <- c(318, 906)
US <- c(217, 1815)
df <- as.data.frame(cbind(Australia, US))
rownames(df) <- c("Fatal", "Non-Fatal")
df

##          Australia   US
## Fatal           318  217
## Non-Fatal       906 1815

```

```

chisq <- chisq.test(df)
chisq #Answer

## 
## Pearson's Chi-squared test with Yates' continuity correction
## 
## data: df
## X-squared = 129.13, df = 1, p-value < 2.2e-16

```

The chi-square test indicates that there is a difference between the deadliness of American sharks in comparison to Australian sharks. While the chi-square test does not indicate which is more deadly, the proportion of deadly attacks within each country calculated in the EDA indicates that Australian sharks are more deadly than American sharks.

Power test

```

N <- sum(df)

## convert observations into frequencies
ex31 <- df/sum(df)

## observed frequencies
ex31

##          Australia      US
## Fatal      0.09766585 0.06664619
## Non-Fatal 0.27825553 0.55743243

## margins
m13 <- rowSums(ex31)
m13

##      Fatal Non-Fatal
## 0.164312 0.835688

m23 <- colSums(ex31)
m23

## Australia      US
## 0.3759214 0.6240786

## expected given independence
ex33.expt <- outer(m13,m23, '*')
ex33.expt

##          Australia      US
## Fatal      0.06176841 0.1025436
## Non-Fatal 0.31415297 0.5215350

ex23.dif <- (ex31 - ex33.expt)^2/ex33.expt
ex23.dif

##          Australia      US
## Fatal      0.020862221 0.012566613
## Non-Fatal 0.004101907 0.002470834

eeses <- sqrt(sum(ex23.dif))

ES.w2(ex31)

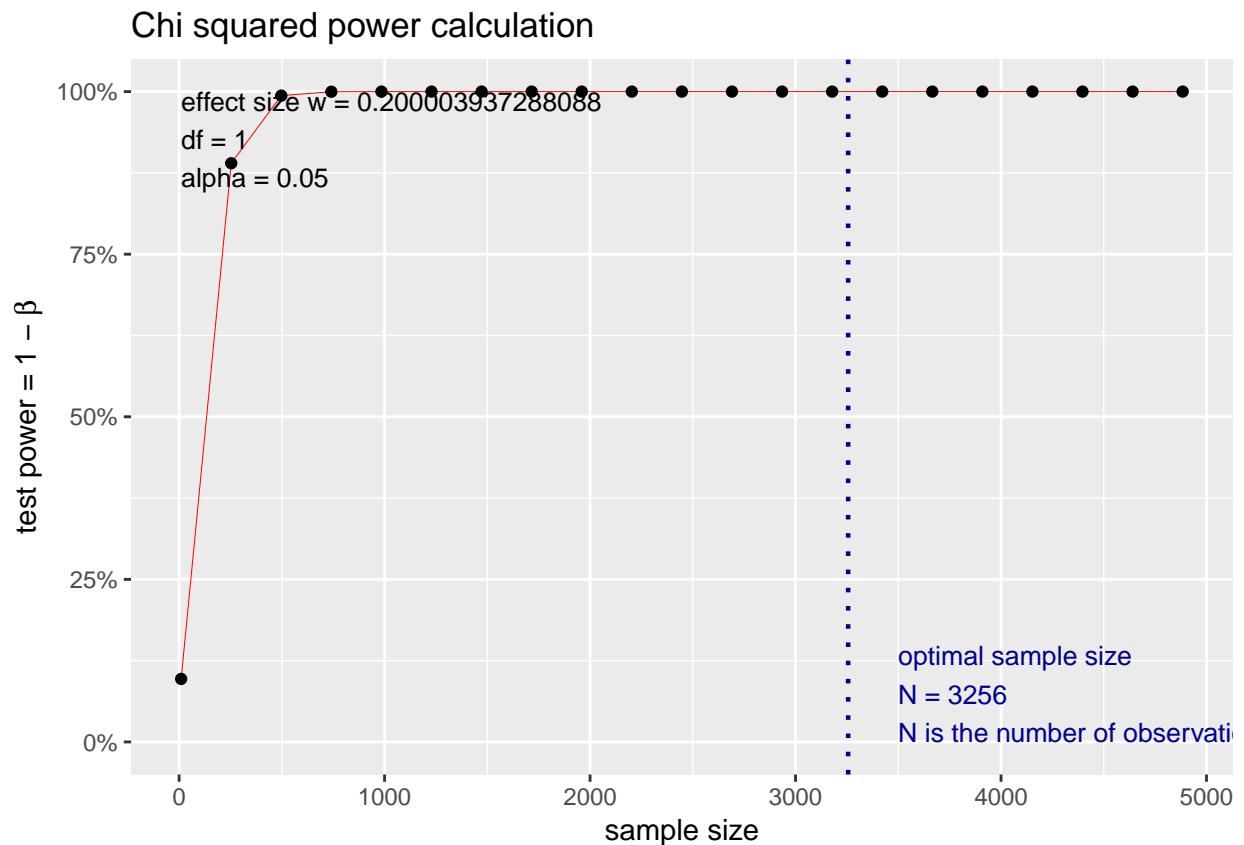
```

```

## [1] 0.2000039
pw3 <- pwr.chisq.test(w=ES.w2(ex31), N=N, df=1, sig.level=.05)
pw3

##
##      Chi squared power calculation
##
##      w = 0.2000039
##      N = 3256
##      df = 1
##      sig.level = 0.05
##      power = 1
##
## NOTE: N is the number of observations
plot(pw3)

```



(4) Power Analysis

Explain the use of the arcsine transformation. How does it work? Why does it work?

Arcsine Transformation

Cohen's 'h' is measure of distance between 2 proportions. It can be used for →

- a) Discretizing Z-proportions into high, med. low.
- b) Calculate sample size for future study.

It can be used along with Null hypothesis tests for proportions. We would like to transfer the difference in probabilities into an equally split scale.

We would like to have an equally split scale as the results would be more interpretable.

However, probability scale doesn't provide arcsine transformⁿ.
this. So we perform

The use of arc sine transform:

- ① Difference in arc sine scale are equally detectable.
- ② It stabilizes the variance

The arcsine transformation reduces

$$P_1 - P_2 \text{ to } \phi_1 - \phi_2 \quad \text{s.t. } \begin{aligned} \phi_{1i} &= \phi_{2i} \\ &= \phi_{1j} = \phi_{2j} \end{aligned}$$

Given a proportion 'p'

$$\text{arc sine transform } \phi = z \arcsin \sqrt{p}$$

The motivation behind using this transform is that it reduces variance after transform.

$\text{var}(\arcsin \sqrt{p}) \sim \text{Delta method.}$

$$\approx \frac{\text{Var}(p)}{4p(1-p)} = \frac{p(1-p)}{4n p(1-p)} = \underline{\underline{\frac{1}{4n}}}$$

where $p = x/n$

Q6 Using MOM and MLE, find estimate of the following -

i) exponential distribution

$$x_1 - x_n \sim \exp(\lambda)$$

MLE :

$$\text{Joint distribution } L(\lambda; x_1 - x_n) = \prod_{i=1}^n f_x(x_i; \lambda)$$

$$= \prod_{i=1}^n \lambda \exp(-\lambda x_i)$$

$$= \lambda^n \exp(-\lambda \sum x_i)$$

Taking log likelihood.

$$l \rightarrow n \ln(\lambda) - \lambda \sum x_i$$

Using $\frac{dl}{d\lambda} = 0$ (for gradient descent
- using slope = 0)

$$\frac{d}{d\lambda} \left(n \ln(\lambda) - \lambda \sum_{i=1}^n x_i \right)$$

$$= \frac{n}{\lambda} - \sum x_i = 0 \quad \rightarrow \text{we get:}$$

$$\lambda = \frac{\sum x_i}{n}$$

⑥ A new distribution

$$f(x) = \begin{cases} (1-\theta) + 2\theta x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

mom for θ

First moment \rightarrow

$$E(x) = \int_0^1 x [1-\theta + 2\theta x] dx$$

$$= \frac{\theta+3}{6}$$

Setting sample mean = proportion mean

$$\bar{x} = \frac{\theta+3}{6}$$

Solving we get.

$$\theta = 6 \frac{\sum x}{n} - 3$$

mle for θ

$$L(\theta) = \prod_{i=1}^n [(1-\theta) + 2\theta x_i]$$

$$\ell(\theta) = \log L(\theta)$$
$$= \sum_{i=1}^N \ln (1 - \theta + 2\theta x_i)$$

$$\frac{d(\ell(\theta))}{d\theta} = \sum_{i=1}^n \frac{-1 + 2x_i}{1 - \theta + 2\theta x_i}$$
$$= \sum_{i=1}^n \frac{-1 + 2x_i}{1 - \theta(-1 + 2x_i)} = 0.$$

This doesn't have a closed form solution.

$$\hat{\theta} = \arg \max \left(\sum_{i=1}^n \frac{1}{\frac{1}{-1 + 2x_i} - \theta} \right) \text{ subject to } 0 < \theta < 1$$

This can be solved by gradient descent -

(6) Rainfall in Illionis

What distribution best represents the rainfall data ?

```
setwd('//Users//diptanshu//Downloads/illinois storms')
ill_60 <- read.delim("ill-60.txt", header = FALSE)
ill_61 <- read.delim("ill-61.txt", header = FALSE)
ill_62 <- read.delim("ill-62.txt", header = FALSE)
ill_63 <- read.delim("ill-63.txt", header = FALSE)
ill_64 <- read.delim("ill-64.txt", header = FALSE)

colnames(ill_60) <- c("rain")
colnames(ill_61) <- c("rain")
colnames(ill_62) <- c("rain")
colnames(ill_63) <- c("rain")
colnames(ill_64) <- c("rain")

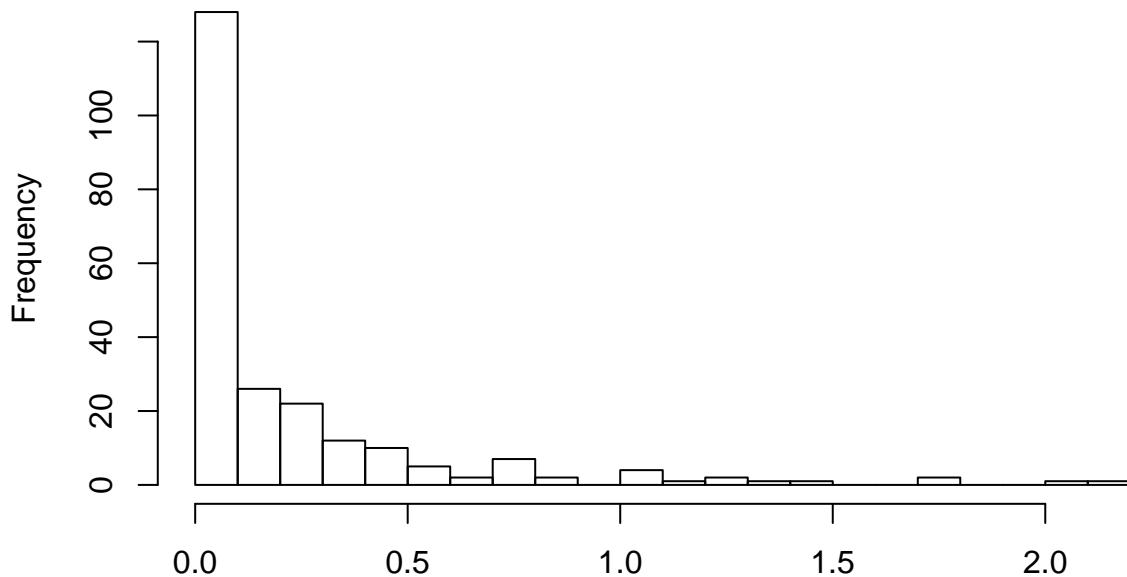
ill_60$year <- c("1960")
ill_61$year <- c("1961")
ill_62$year <- c("1962")
ill_63$year <- c("1963")
ill_64$year <- c("1964")

ill_60$obs <- seq.int(nrow(ill_60))
ill_61$obs <- seq.int(nrow(ill_61))
ill_62$obs <- seq.int(nrow(ill_62))
ill_63$obs <- seq.int(nrow(ill_63))
ill_64$obs <- seq.int(nrow(ill_64))

rain <- rbind(ill_60, ill_61, ill_62, ill_63, ill_64)
# Histogram

hist(rain$rain, breaks = 20)
```

Histogram of rain\$rain



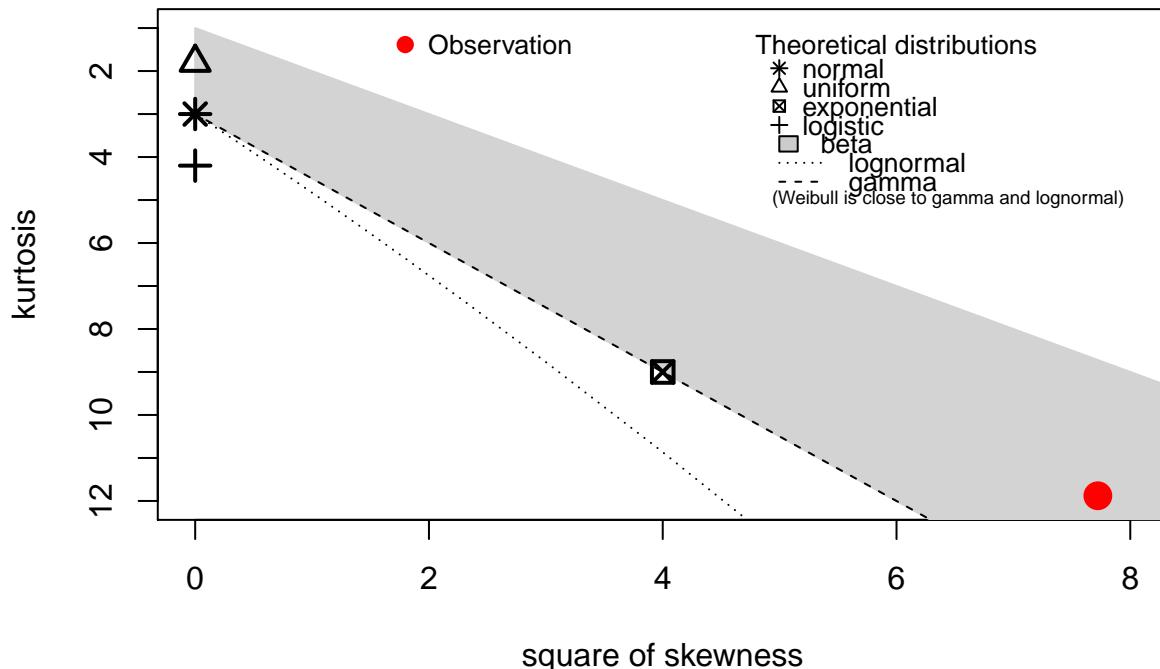
rain\$rain

The data

is right skewed distribution – beta, gamma, Weibull, or lognormal can be used. There is a concentrated density at zero; so we can even think about hurdle model densities

```
descdist(rain$rain, obs.col = "red")
```

Cullen and Frey graph



```
## summary statistics
```

```

## -----
## min: 0.001 max: 2.13
## median: 0.07
## mean: 0.2243921
## estimated sd: 0.3658212
## estimated skewness: 2.778925
## estimated kurtosis: 11.87935

```

In its present form, this cannot be a beta distribution as range is not constrained in [0,1] Also, the average of rainfall is not intuitively realted to probability; hence any transformations also do not make sense

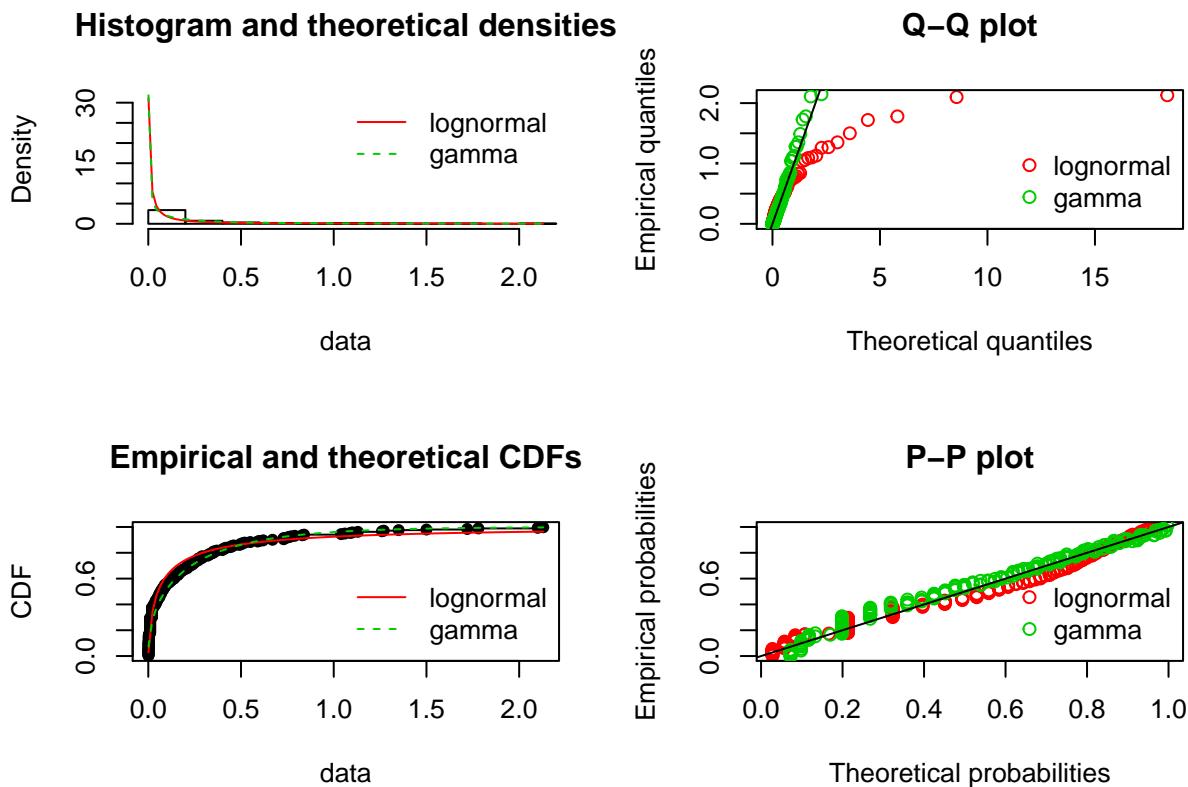
Checking between Gamma and Lognormal Since Weibull is closer to Gamma - I am not considering it currently

```

fit_ln <- fitdist(rain$rain, "lnorm")
fit_g <- fitdist(rain$rain, "gamma")

plot.legend <- c( "lognormal", "gamma")
par(mfrow = c(2,2))
denscomp(list( fit_ln, fit_g), legendtext = plot.legend)
qqcomp(list( fit_ln, fit_g), legendtext = plot.legend)
cdfcomp(list(fit_ln, fit_g), legendtext = plot.legend)
ppcomp(list( fit_ln, fit_g), legendtext = plot.legend)

```



From the empirical cdf graph - We see that towards the higher values, log normal is a better approximation; whereas at the mid region and starting region - gamma is closer.

```

# Analyzing the distances between theoretical and emprirical distributions
fit_w <- fitdist(rain$rain, "weibull")

#Draw from weibull distribution

```

```
w_dist <- rweibull(n = 227, shape = 0.5690986, scale = 0.1394868)
g_dist <- rgamma(n = 227, shape = 0.4408386, rate = 1.9648409)

ks.test(w_dist, rain$rain)

## Warning in ks.test(w_dist, rain$rain): p-value will be approximate in the
## presence of ties

##
## Two-sample Kolmogorov-Smirnov test
##
## data: w_dist and rain$rain
## D = 0.12335, p-value = 0.06325
## alternative hypothesis: two-sided

ks.test(g_dist, rain$rain)

## Warning in ks.test(g_dist, rain$rain): p-value will be approximate in the
## presence of ties

##
## Two-sample Kolmogorov-Smirnov test
##
## data: g_dist and rain$rain
## D = 0.12335, p-value = 0.06325
## alternative hypothesis: two-sided
```

Weibull distribution will be a better distribution for modelling this distribution
Probability value that both samples to arise from same distribution is higher for weibull

Estimation of parameters of gamma distribution using Maximum Likelihood estimation and bootstrap

```
# How estimation for a single step will be done - mle_gamma
# Creating a function to pass into the bootstrap

mle_gamma <- function(x){
  parameter <- fitdistr(x, "gamma", start=list(shape=1, rate=1))$estimate
  return(parameter)
}

# Bootstrap sample from data frame `x` with statistics calculated in `g` for `B` samples
bootstrap <- function (x, g, B = 100) {
  n <- nrow(x)

  theta.shape <- numeric(B)
  theta.rate <- numeric(B)

  for (i in 1:B) {

    x.star <- x[sample.int(n, replace = TRUE), ]

    parameter <- g(x.star)
    theta.shape[i] <- as.numeric(parameter['shape'])
    theta.rate[i] <- as.numeric(parameter['rate'])
```

```

}

parameters <- list(theta.shape, theta.rate)
return(parameters)
}

parameters <- bootstrap(data.frame(rain$rain), g = mle_gamma)
shape <- mean(parameters[[1]])
shape_err <- sd(parameters[[1]])
rate <- mean(parameters[[2]])
rate_err <- sd(parameters[[2]])

shape

## [1] 0.4389764
shape_err

## [1] 0.02646655
rate

## [1] 1.994587
rate_err

## [1] 0.2564691

# Using method of moments for bootstrapped calculation of parameters
mom_gamma <- function(x){
  mu <- mean(x)
  variance <- var(x)
  parameter = list()
  parameter['rate'] <- as.numeric(mu/variance)
  parameter['shape'] <- as.numeric(mu^2/variance)

  return(parameter)
}

parameters <- bootstrap(data.frame(rain$rain), g = mom_gamma)
shape <- mean(parameters[[1]])
shape_err <- sd(parameters[[1]])
rate <- mean(parameters[[2]])
rate_err <- sd(parameters[[2]])

shape

## [1] 0.3894388
shape_err

## [1] 0.04466532
rate

## [1] 1.720981

```

```

rate_err

## [1] 0.2546841

# To calculate which is a better estimate, we can use KS test again
r_mom <- rgamma(n = 227, shape = 0.387184, scale = 1.739256)
r_mle <- rgamma(n = 227, shape = 0.4428879, rate = 1.976928)

ks.test(r_mom, rain$rain)

## 
##  Two-sample Kolmogorov-Smirnov test
##
## data: r_mom and rain$rain
## D = 0.23789, p-value = 5.274e-06
## alternative hypothesis: two-sided
ks.test(r_mle, rain$rain)

## 
##  Two-sample Kolmogorov-Smirnov test
##
## data: r_mle and rain$rain
## D = 0.14537, p-value = 0.0165
## alternative hypothesis: two-sided

```

Maximum Likelihood Estimation does a better job at estimating the parameters compared to the Method of Moments. Also, compared to standard error values compared in the paper, these standard error values are more reliable

(7) Statistical Decision Theory

Proof after table formation

```

# Minimax regret rules for small sample size

# Function to find value given alpha and N
mm <- read.csv('minimax.csv')

tresh <- mm

n_calc <- function(alpha, N){
  if(N == 0){
    return (0)
  }

  if(0.5/N >= alpha){
    return(0)
  }
  for (i in c(1:N)){
    if(i/N >= alpha)
      return(i)
  }
}

```

```

    return(0)
}

for (i in c(1:5)){
  for( j in c(2:12)){
    mm[i,j] <- n_calc(mm[i,1], j-2)
  }
}

kable(mm)

```

alpha	n_0	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_10
0.10	0	0	0	0	0	0	1	1	1	1	1
0.25	0	0	0	1	1	2	2	2	2	3	3
0.50	0	0	1	2	2	3	3	4	4	5	5
0.75	0	1	2	3	3	4	5	6	6	7	8
0.90	0	1	2	3	4	5	6	7	8	9	9

1. To derive :

Given Prior of $\beta \sim \text{Beta}(c, d)$

Then Posterior mean for β is

$$\mu = \frac{(c+n)}{(c+d+N)}.$$

The resulting bayes rules is :

$$S(n) = 0 \quad \text{for} \quad \mu < \alpha \quad 10-a$$

$$S(n) = \lambda \quad \text{for} \quad \mu = \alpha \quad 0 \leq \lambda \leq 1 \quad 10-b$$

$$S(n) = 1 \quad \text{for} \quad \mu > \alpha \quad 10-c$$

where $\alpha : P(y=1|A)$ and $\beta = P(y=1|B)$.

Definitions

1. We have 2 treatments A, B $A \equiv \text{control}$
 $B \equiv \text{test}$

outcome $Y_i \begin{cases} 0 \\ 1 \end{cases}$ and $\alpha = P(Y(A) = 1)$
 $\beta = P(Y(B) = 1)$

2. N subjects for trial of B

in which $\begin{cases} n & \text{success} \\ N-n & \text{failure} \end{cases}$

$$S(n) = \frac{n}{N} \quad \text{which lies in set } [0, 1]$$

3. Welfare function

$$W(S, P, N) = \alpha + (\beta - \alpha) (E(S(n)))$$

where $n \sim \text{Binomial}(\hat{\beta}, N)$

The prior on $\hat{\beta} \sim \text{Beta}(c, d)$

So to calc. W :

$$E[S(n)] = \sum_{i=0}^N S(i) f(n=i; \beta, N)$$

$$\text{where } f(n=i, \beta, N) = \sum_{i=0}^N \binom{N}{i} \beta^i (1-\beta)^{N-i}$$

Only unknown determinant is β

$$\hat{\beta} \sim \text{Beta}(c, d)$$

$$p(\hat{\beta}) = \frac{\Gamma(c+d)}{\Gamma(c) \Gamma(d)} \hat{\beta}^{c-1} (1-\hat{\beta})^{d-1}$$

By law of lazy statistician:

$$\begin{aligned} E[S(n)] &= \sum_{i=0}^N S(i) \binom{N}{i} \hat{\beta}^i (1-\hat{\beta})^{N-i} \cdot p(\hat{\beta}) \\ &= \sum_{i=0}^N \frac{i}{N} \binom{N}{i} \hat{\beta}^i (1-\hat{\beta})^{N-i} \cdot \frac{1}{B(c, d)} \hat{\beta}^{c-1} (1-\hat{\beta})^{d-1} \\ &= \sum_{i=0}^N \frac{i}{N} \binom{N}{i} \hat{\beta}^{(i+c-1)} \frac{(1-\hat{\beta})^{N-i+d-1+c}}{B(c, d)} \\ &= \sum_{i=0}^N \frac{i}{N} \left(\frac{N \cdot N-1 \cdot N-2 \dots N-i}{B(c, d)} \right) \end{aligned}$$

$$\text{Since } \beta(c, d) = \frac{\Gamma(c+d)}{\Gamma(c)\Gamma(d)} = \frac{(c+d)(r(c+d-1))}{c(r(c-1))d(r(d-1))}$$

$$\text{Hence } \frac{(c+d)!}{c! d!}$$

$$= \sum \frac{i^*}{N} \frac{N!}{(N-i)! i!} \frac{(c+d)!}{c! d!}$$

$$= \sum \frac{(N-i)!}{(N-i)! (i-1)!} \frac{(c+d)!}{c! d!}$$

$$E_s[\delta(n)] = f(n > n_0, \beta_s, N) + \lambda f(n = n_0, \beta_s, N)$$

$$E_s[\delta(n)] \propto \left(\hat{\beta}\right)^{i+c-1} (1-\hat{\beta})^{N-i+c+d-1}.$$

$$\text{Mean of beta} \Rightarrow \frac{c+i^*n}{c+d+N-n+m} = \frac{c+n}{c+d+N}.$$

Hence we can replace this $\hat{\beta}$.

$$\text{So if } \beta > \alpha \text{ ie } \frac{c+n}{c+d+N} > \alpha \text{ then } \delta = 1$$

$$\text{if } \beta = \alpha \text{ ie } " = " \text{ then } \delta = \lambda$$

$$\text{if } \beta < \alpha \text{ ie } " = " \text{ then } \delta = 0.$$