Homework Assignment 1

Kalvin Goode and Hedi Xia

October 12, 2018

1.

(a) Population of interest is the group of set items relevent to the question.

Population quantity of interest is

Sampling units is

(b) Useful estimand for writing style including transition words between sentences and .

(c) $L(\lambda : y_1, ...y_n) = \prod_{i=1}^{n} Poisson(y_i|\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{y_i}}{y_i!}$

$\propto e^{-n\lambda} \lambda^{n\bar{y}}$

(d) $l(\lambda : y_1, ...y_n) = ln(L(\lambda : y_1, ...y_n)) = ln(\prod_{i=1}^{n} Poisson(y_i|\lambda)) = ln(\prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{y_i}}{y_i!}) =$

$= \sum_{i=1}^{n} ln(e^{-\lambda} \frac{\lambda^{y_i}}{y_i!})$

$= \sum_{i=1}^{n} (-\lambda + y_i ln\lambda - ln y_i!) = -n\lambda + ln(\lambda \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} ln y_i)$

Differentiate, we have

$l'(\lambda) = \frac{n\bar{y}}{\lambda} - n$

So, MLE is $\lambda = \bar{y}$

(e)

```r
suppressMessages(library(tidyverse)) # data manipulation & plotting
library(stringr) # text cleaning and regular expressions
library(tidytext) # provides additional text mining functions
library(harrypotter) # text for the seven novels of the Harry Potter series
text_tb<-tibble(chapter=seq_along(deathly_hallows),text=deathly_hallows)
tokens<-text_tb%>%
  unnest_tokens(word, text)
word_counts<-tokens%>%
  group_by(chapter) %>%
  count(word, sort = TRUE) %>% ungroup

word_counts_mat<-word_counts %>% spread(key=word,value=n,fill=0)
dark_counts<-word_counts_mat$dark

text_tb%>%
  ggplot(aes(dark_counts))+geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
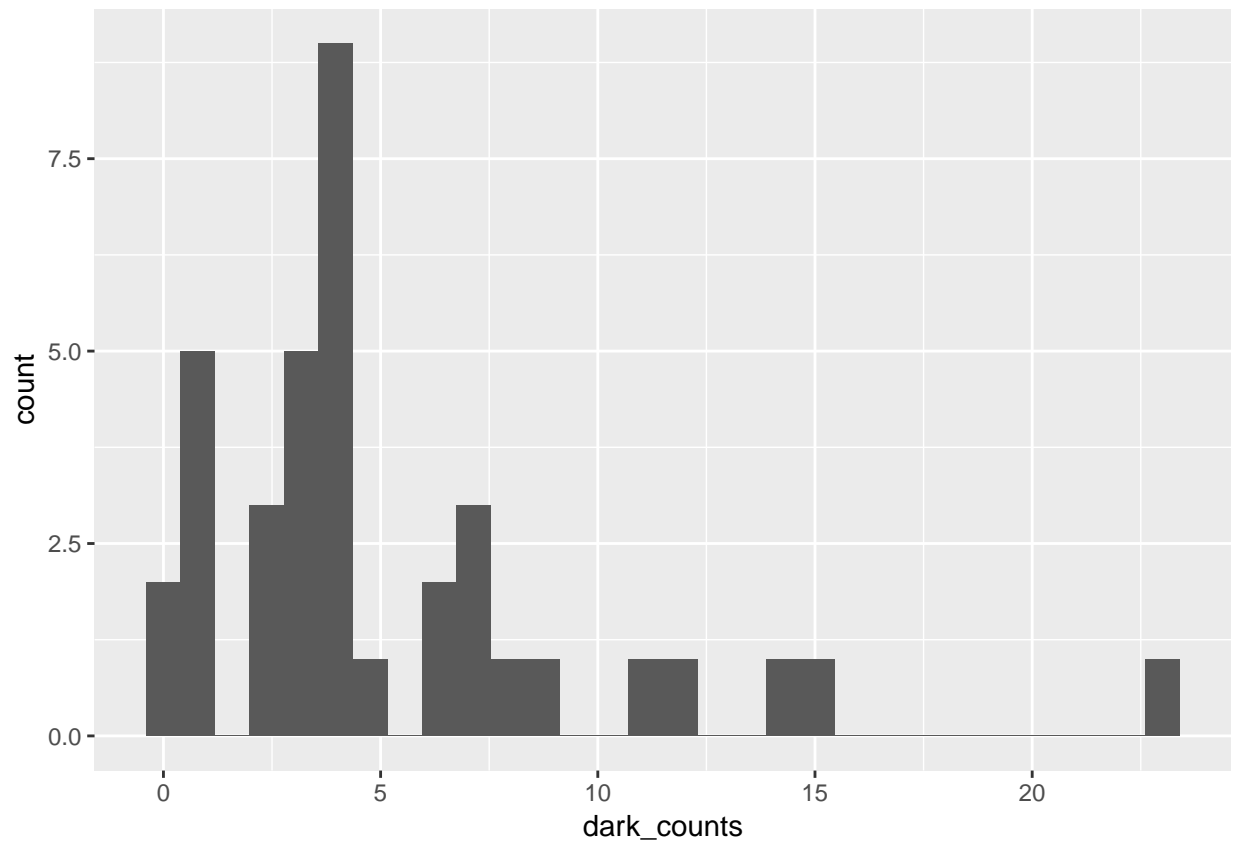
```
#barplot(dark_counts, main="Dark Counts", xlab="Chapter")
```

  (f)

2

```
chapter_lengths <- word_counts %>% group_by(chapter) %>%
summarize(chapter_length = sum(n)) %>%
ungroup %>% select(chapter_length) %>% unlist %>% as.numeric
```

  (a)

  (b)