

# Assignment 1

*Angela Zhai*

*September 14, 2018*

## Document assignment

Consider the following situation:

A sloppy printer produces books with an average of 2 misprints per page. You want to know how many pages have more than  $k$  misprints in a book of  $n$  pages. Make an  $n \times k$  table that shows the relationship between the total number of pages in a book and the number of pages with  $k$  misprints.

Show and explain your work. Include equations and calculations to teach the reader how to solve the problem. Include an image of a book.

Push your solution to a github repository and submit the url for repository on blackboard. Be sure your repo includes your document as a pdf file and as an RMD file. Include other files needed to recompile your document.

## Code

```
library(knitr)
library(kableExtra)

nrow = 6
ncol = 6

k <- c(2,3,4,5,6,7)
n <- c(100,200,300,400,500,600)

error <- data.frame()
error[3,1] <- "Total"
error[4,1] <- "Pages"

for (m in 1:nrow){
  error[m,2] <- n[m]
}

misprints <- rpois(n = 10000,lambda = 2)

for (sel in 1:nrow){
  for (selc in 1:ncol){
    error[sel,selc+2] <- length(which(misprints[]>=k[sel]))/length(misprints)
  }
}

colnames(error) <- c(" ", " ", "2", "3", "4", "5", "6", "7")

kable(x = error, booktabs = T) %>%
  kable_styling(position = "center") %>%
  add_header_above(c(" " = 2, "Least Misprints" = 6))
```

		Least Misprints					
		2	3	4	5	6	7
Total Pages	100	0.5882	0.5882	0.5882	0.5882	0.5882	0.5882
	200	0.3147	0.3147	0.3147	0.3147	0.3147	0.3147
	300	0.1434	0.1434	0.1434	0.1434	0.1434	0.1434
	400	0.0533	0.0533	0.0533	0.0533	0.0533	0.0533
	500	0.0165	0.0165	0.0165	0.0165	0.0165	0.0165
	600	0.0040	0.0040	0.0040	0.0040	0.0040	0.0040

```
nrow = 8
ncol = 6

k <- c(100,200,300,400,500,600,700,800)
n <- c(100,200,300,400,500,600)

error <- data.frame()
error[4,1] <- "Least"
error[5,1] <- "Errors"

for (m in 1:nrow){
  error[m,2] <- k[m]
}

for(i in 1:nrow){
  for (j in 1:ncol){
    freq <- k[i]/n[j]
    error[i,j+2] <- ppois(q = freq, lambda = 2, lower.tail = FALSE)
  }
}

colnames(error) <- c(" ", " ", "100", "200", "300", "400", "500", "600")

kable(x = error, booktabs = T) %>%
  kable_styling(position = "center") %>%
  add_header_above(c(" " = 2, "pages" = 6))
```

		pages					
		100	200	300	400	500	600
Least Errors	100	0.5939942	0.8646647	0.8646647	0.8646647	0.8646647	0.8646647
	200	0.3233236	0.5939942	0.8646647	0.8646647	0.8646647	0.8646647
	300	0.1428765	0.5939942	0.5939942	0.8646647	0.8646647	0.8646647
	400	0.0526530	0.3233236	0.5939942	0.5939942	0.8646647	0.8646647
	500	0.0165636	0.3233236	0.5939942	0.5939942	0.5939942	0.8646647
	600	0.0045338	0.1428765	0.3233236	0.5939942	0.5939942	0.5939942
	700	0.0010967	0.1428765	0.3233236	0.5939942	0.5939942	0.5939942
	800	0.0002374	0.0526530	0.3233236	0.3233236	0.5939942	0.5939942

The question is about least errors appear within n pages, so we need to get cumulative probability of poisson distribution.

The equation to get CDF of poisson distribution is:  $e^{-\lambda} \sum_{i=0}^k \frac{\lambda^i}{i!}$

Like if we get 100 misprints in a book of

The main part of the code is: `#ppois(q=freq, lamda=2, lower.tail=FALSE)` Function `ppois` is used to calculate CDF for poisson distribution; `q` assigns the average errors for one page in a book of `n` pages; `lamda` assigns average misprints per page of this printer; `lower.tail=FALSE` means we try to get the cumulated value for errors more than `q`.