S520 Instructor's Solutions

Spring 2023 STAT-S 520

February 5th, 2023

1.

a. 64 of accepted student decides to attend other college so the probability of a success (attend the college) is $p = 1 - 0.64 = 0.36$. Let X be the number of students attending college, so $X \sim binomial(225, 0.36)$. The expected number of students to be accommodated: is $EX = np = 225 * 0.36 = 81$

b.$P(X > 95) = 1 - P(X <= 95) = 1 - F(95)$

```r
1-pbinom(95,225,0.36)
```

```
## [1] 0.02291658
```

2.

a. The probability of correctly guessing (success) is $p = 1/5 = 0.2$. The number of trials is $n = 25$, and we can define Y as the random variable that assigns the number of correct guesses, so $Y \sim binomial(25, 0.2)$. The expected number of correct guesses is $EX = np = 25 * .2 = 5$

b. Probability of getting a score greater than 7 is $P(Y > 7) = 1 - P(Y <= 7) = 1 - F(7)$

```r
1-pbinom(7,25,0.2)
```

```
## [1] 0.1091228
```

c. From part b, let's define $p = P(Y > 7) \approx 0.11$ as the probability of getting a score indicative of ESP. Moreover, let Z be the random variable that assigns the number of receivers getting a score indicative of ESP, so $Z \sim binomial(20, 0.11)$. So $P(Z \geq 1) = 1 - P(Z < 1) = 1 - P(Z \leq 0) = 1 - F_z(0)$. Using R:

```r
p = 1-pbinom(7,25,0.2)
1 - pbinom(0,20,p)
```

```
## [1] 0.9008353
```

3.

We first find the probability that a someone observes no more than two `Heads` out of 89. This is a binomial by itself:

```r
p = pbinom(2, 89, 0.3)
p
```

```
## [1] 1.240591e-11
```

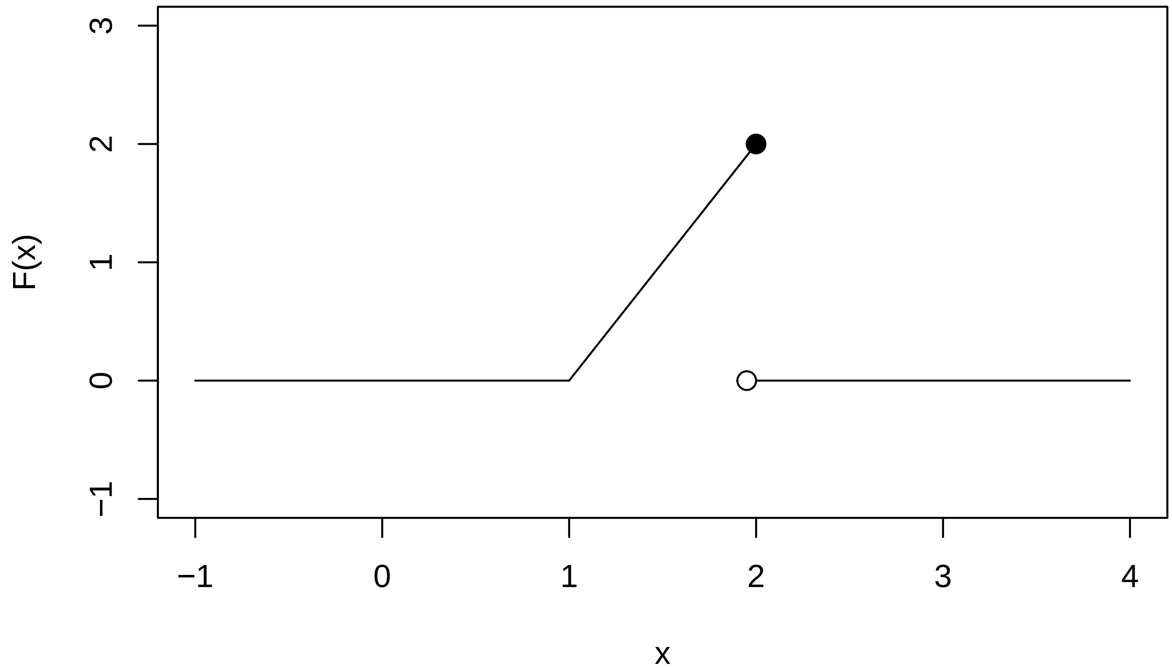
A very small probability of success. Now, we let X represent the number of students (out of 1500) that observe no more than two `Heads` out of 89. Using p obtained above, we have $X \sim binomial(1500, p)$ and we need to find $P(X \geq 1) = 1 - F(0)$

```r
1 - pbinom(0, 1500, p)
```

```
## [1] 1.860886e-08
```

4.

a.



b. Note that f assigns values that are all greater than or equal to zero. The the area under f for the relevant region (between 1 and 2) is like the area of a triangle. So

$$Area = 1/2 * base * height = 1/2 * (2 - 1) * 2 = 1$$

So f is indeed a PDF.

c. Using geometry, this is the difference of areas between two triangles given by: $P(1.5 < X < 1.75) = P(X < 1.75) - P(X \leq 1.5) = F(1.75) - F(1.5)$. So we get $(1.75 - 1) * 2 * (1.75 - 1)/2 - (1.75 - 1) * 2 * (1.75 - 1)/2$

```
(1.75 - 1)*2*(1.75 - 1)/2 - (1.5 - 1)*2*(1.5 - 1)/2
```

```
## [1] 0.3125
```

Or using integrals:

$P(1.5 < X < 1.75) = \int_{1.5}^{1.75} 2(x-1)dx$
$= [x^2 - 2x]_{1.5}^{1.75}$
$= [1.75^2 - 1.5^2] - [2 \times 1.75 - 2 \times 1.5]$
$= 0.3125$

5.

```
library(fivethirtyeight)
```

```
## Warning: package 'fivethirtyeight' was built under R version 4.2.2

## Some larger datasets need to be installed separately, like senators and
## house_district_forecast. To install these, we recommend you install the
## fivethirtyeightdata package by running:
## install.packages('fivethirtyeightdata', repos =
## 'https://fivethirtyeightdata.github.io/drat/', type = 'source')
```

```
library(tidyverse)
```

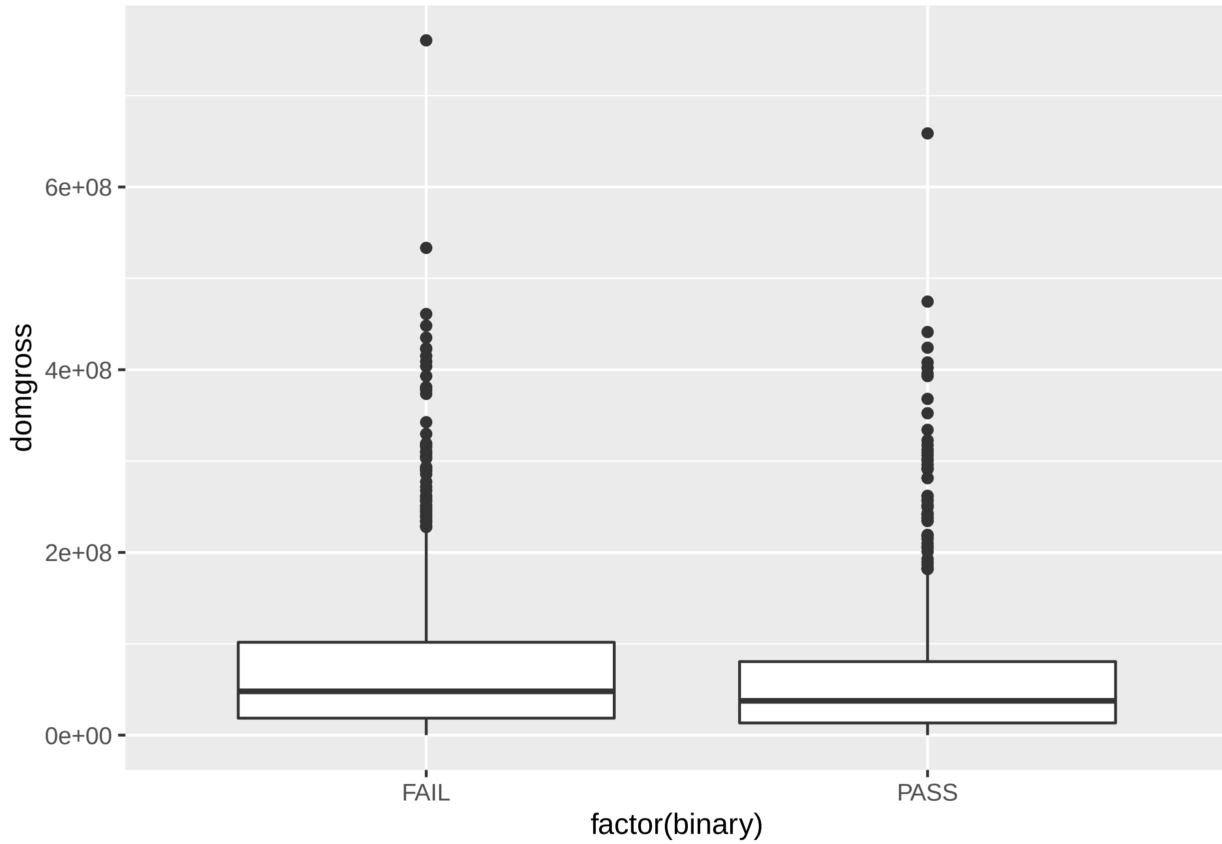
```
## Warning: package 'tidyverse' was built under R version 4.2.2

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.3     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
bechdel %>%
  group_by(binary) %>%
  summarize(mean=mean(domgross,na.rm=T),s=sd(domgross,na.rm=T))
```

```
## # A tibble: 2 x 3
##   binary      mean         s
##   <chr>      <dbl>     <dbl>
## 1 FAIL   74985189. 83484962.
## 2 PASS   61885653. 75758965.
```

```
library(ggplot2)
ggplot(bechdel, mapping = aes(x= factor(binary), y= domgross))+
  geom_boxplot()
```

```
## Warning: Removed 17 rows containing non-finite values (stat_boxplot).
```




```
bechdel %>%
  group_by(period_code) %>%
  summarize(count=n())
```

```
## # A tibble: 6 x 2
##   period_code count
##         <int> <int>
## 1           1   438
## 2           2   488
## 3           3   352
## 4           4   247
## 5           5    90
## 6          NA   179
```