

practiceMidterm1

Jingwen GAO

2025-10-08

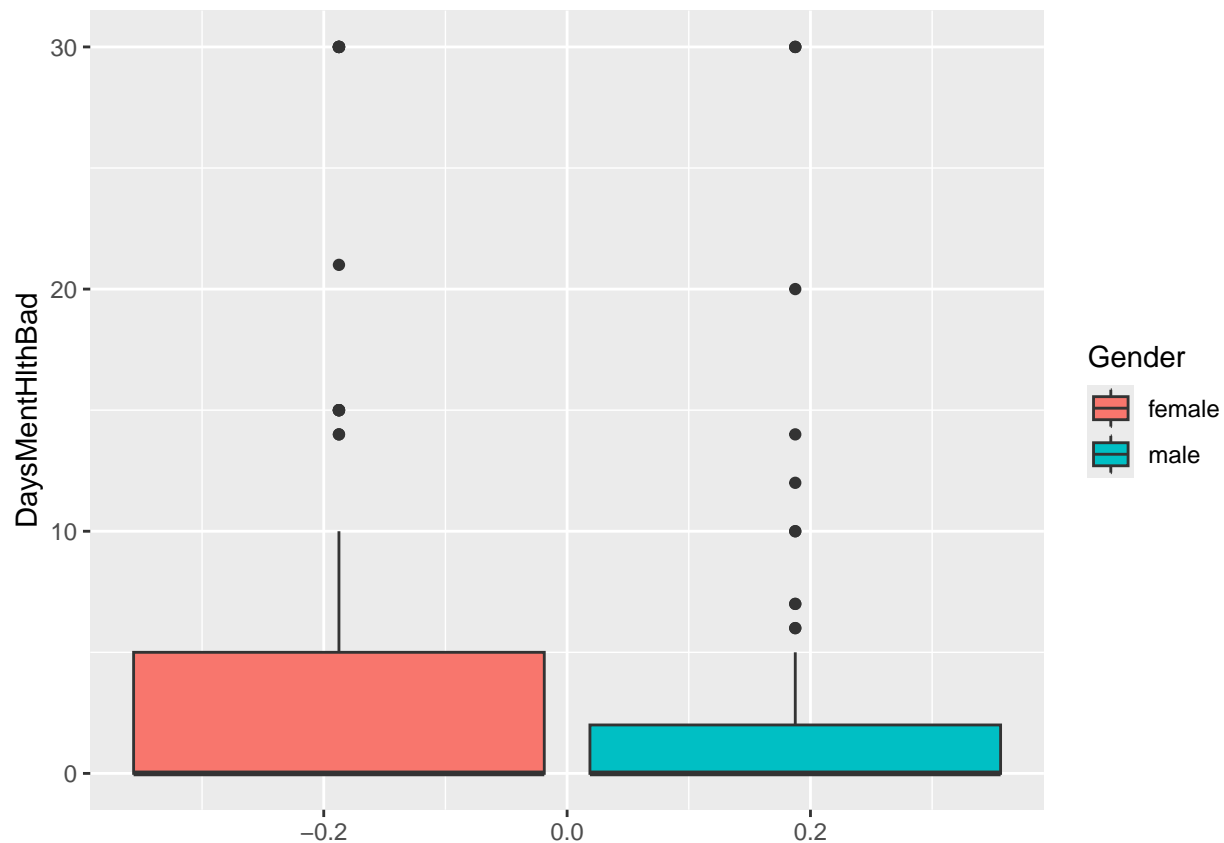
1. NHANES data. Find the mean reported days/month of having bad mental health for participants living at or below 2 (the ratio of income to poverty level). Does reported poor mental health among impoverished participants differ by gender?

```
library(NHANES)
library(dplyr)
NHANES |>
  mutate(LivingCond = as.numeric(HHIncome)/Poverty, .before = 2) |>
  filter(LivingCond <= 2) |>
  summarise(mean = mean(DaysMentHlthBad, na.rm = T))
```

```
## # A tibble: 1 x 1
##   mean
##   <dbl>
## 1  3.46
```

Lets' view some distribution of DaysMentHlthBad between the two genders.

```
library(ggplot2)
NHANES |>
  mutate(LivingCond = as.numeric(HHIncome)/Poverty, .before = 2) |>
  filter(LivingCond <= 2) |>
  ggplot(aes(y = DaysMentHlthBad, fill = Gender)) + geom_boxplot()
```



```
library(moments)
NHANES|>mutate(LivingCond=as.numeric(HHIncome)/Poverty,.before = 2)|>
  filter(LivingCond<=2)|>group_by(Gender)|>
  summarise(mean=mean(DaysMentHlthBad,na.rm=T),
            median=median(DaysMentHlthBad,na.rm = T),
            mad=mad(DaysMentHlthBad,na.rm=T),

            sd=sd(DaysMentHlthBad,na.rm=T),
            kurtosis=kurtosis(DaysMentHlthBad,na.rm=T),

            skewness=skewness(DaysMentHlthBad,na.rm=T))
```

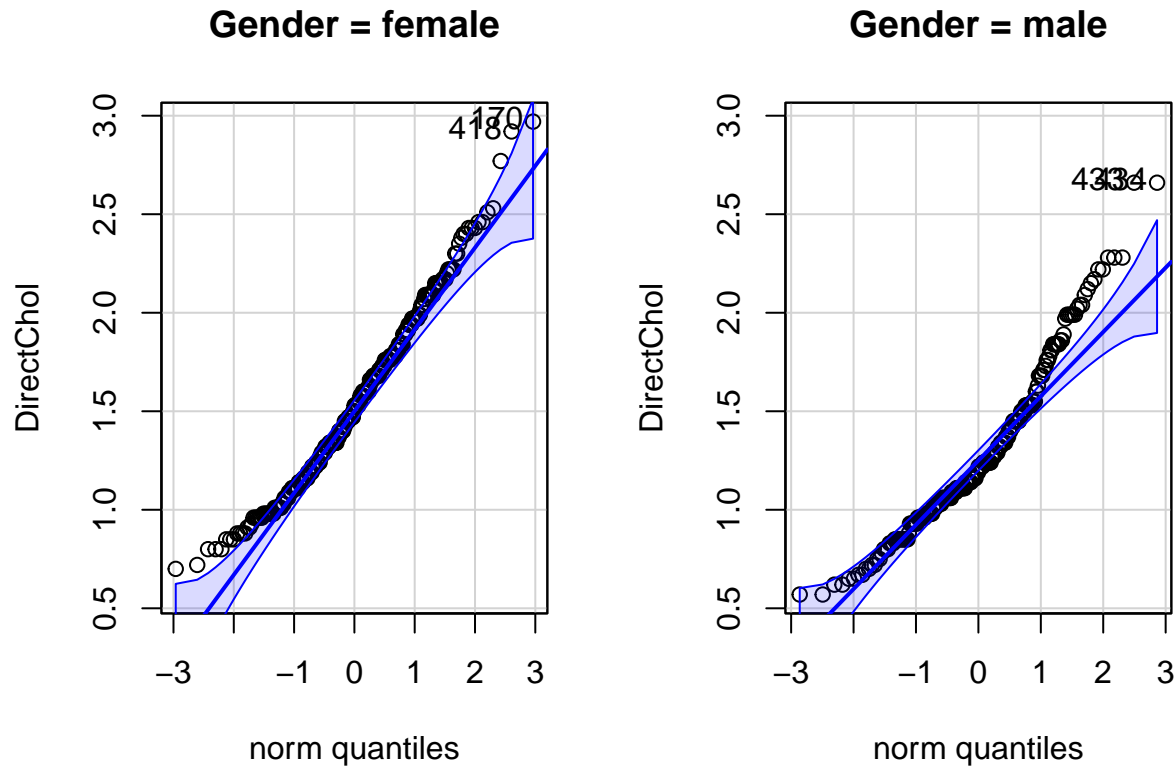
```
## # A tibble: 2 x 7
##   Gender mean median  mad   sd kurtosis skewness
##   <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 female  4.38     0     0  7.56   7.75   2.29
## 2 male    2.42     0     0  5.53  17.9   3.74
```

As the previous results showing, the distribution between the two gender's mentalhealth shows a difference.

- Both female and male have a median 0, which means more than half of them reported 0 days of bad mental health.
- Female has a higher mean, more variance
- Male reports more outliers, more right skewed and flatter distribution.

2. In the NHANES data create a dataframe of the 09-10 data, over 65 participants. Explore the distribution of DirectChol by Gender with QQ plots. Read the plots and describe the tails as heavy or light. Generate density plots to visualize those distributions.

```
new <- NHANES |>
  filter(SurveyYr == "2009_10", Age > 65)
library(car)
qqPlot(DirectChol ~ Gender, data = new) #check variable by groups
```



From the plot, we can see that the distribution of male Directchol has heavy right tails. Comparatively, female DirectChol distribution has light right tails. Male has a more skewed distribution than female.

3. Make a 2-way frequency table with the tabyl function. Adorn with appropriate elements to effectively communicate the joint distribution.

```
library(knitr)
library(kableExtra)
library(janitor)
library(dplyr)
library(tidyr)

tab <- NHANES %>%
  filter(SurveyYr == "2011_12" & Age > 18) %>%
  drop_na(Education) %>%
  drop_na(Gender) %>%
  tabyl(Education, Gender) %>%
  adorn_totals(c("row", "col")) %>%
```

Education/Gender	female	male	Total
8th Grade	3% (91)	3% (121)	6% (212)
9 - 11th Grade	5% (174)	6% (231)	11% (405)
High School	9% (338)	10% (341)	19% (679)
Some College	17% (615)	15% (545)	32% (1,160)
College Grad	16% (584)	15% (544)	31% (1,128)
Total	50% (1,802)	50% (1,782)	100% (3,584)

```

adorn_percentages("all") %>%
adorn_pct_formatting(rounding = "half up", digits = 0) %>%
adorn_ns() %>%
adorn_title("combined")
tab %>%
  kbl() %>%
  kable_styling()

```