STATS 500 HW1

Minxuan Chen

2023-09-09

Table of contents

Github repo: https://github.com/PKUniiiiice/STATS_500

1 Problem 1

The dataset `teengamb` concerns a study of teenage gambling in Britain.

1.1 Glimpse

```
1  data(teengamb)
2  head(teengamb)
```

```
  sex status income verbal gamble
1   1     51   2.00      8    0.0
2   1     28   2.50      8    0.0
3   1     37   2.00      6    0.0
4   1     28   7.00      4    7.3
5   1     65   2.00      8   19.6
6   1     61   3.47      6    0.1
```

1.2 Descriptive Statistics

Note that `sex` is a categorical variable so here we just show the description of the others.
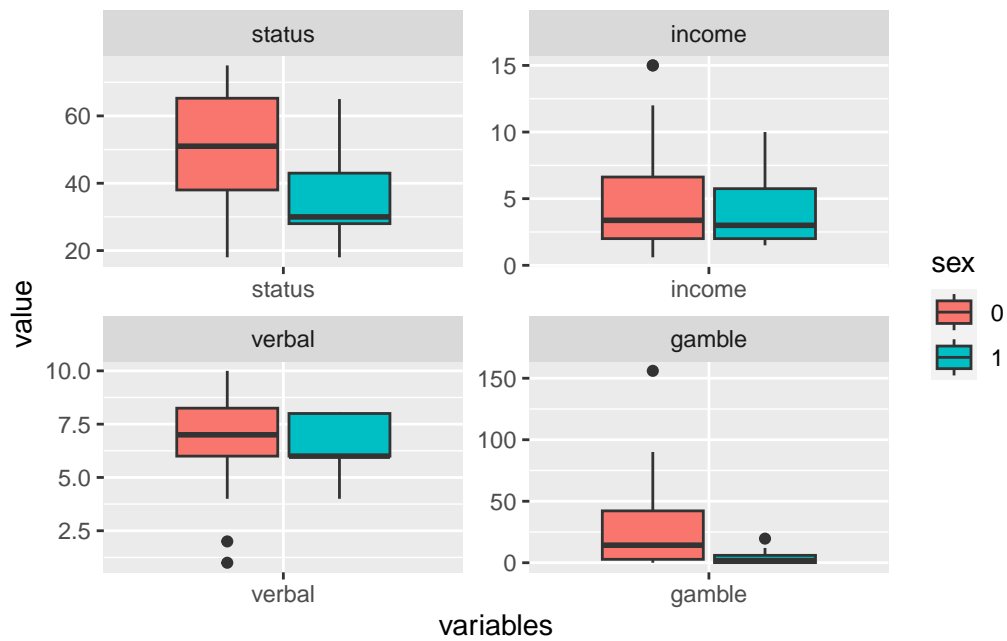
```
1  summary(teengamb[,2:5])
```

```
     status          income          verbal          gamble
 Min.   :18.00   Min.   : 0.600   Min.   : 1.00   Min.   :  0.0
 1st Qu.:28.00   1st Qu.: 2.000   1st Qu.: 6.00   1st Qu.:  1.1
 Median :43.00   Median : 3.250   Median : 7.00   Median :  6.0
 Mean   :45.23   Mean   : 4.642   Mean   : 6.66   Mean   : 19.3
 3rd Qu.:61.50   3rd Qu.: 6.210   3rd Qu.: 8.00   3rd Qu.: 19.4
 Max.   :75.00   Max.   :15.000   Max.   :10.00   Max.   :156.0
```

It seems that the variable `gample` is severely right-skewed.

1.3 Differences between gender and distributions

```
1  teengamb$sex <- as.factor(teengamb$sex)
2  md <- melt(teengamb, id.var=c("sex"), var="variables")
3  p <- ggplot(data=md, aes(x=variables, y=value, fill=sex))+
4      geom_boxplot()+
5      facet_wrap(~variables, scale="free")
6  show(p)
```

In the variables `status` and `gample`, there seems an obvious difference between male and famale.

```
1  ggpairs(teengamb, columns=2:5, ggplot2::aes(color=sex))
```