STAT 625 Diving data analysis

Yu Lu

Yale University

1 Explore Categorical Variables

We first look at the categorical variables to try to understand the game rules.

```
> x <- read.csv("http://www.stat.yale.edu/~jay/625/diving/Diving2000.csv", as.is=TRUE)
> str(x)
```

```
'data.frame':        10787 obs. of  10 variables:
 $ Event     : chr  "M3mSB" "M3mSB" "M3mSB" "M3mSB" ...
 $ Round     : chr  "Final" "Final" "Final" "Final" ...
 $ Diver     : chr  "XIONG Ni" "XIONG Ni" "XIONG Ni" "XIONG Ni" ...
 $ Country   : chr  "CHN" "CHN" "CHN" "CHN" ...
 $ Rank      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ DiveNo    : int  1 1 1 1 1 1 1 2 2 2 ...
 $ Difficulty: num  3.1 3.1 3.1 3.1 3.1 3.1 3.1 3 3 3 ...
 $ JScore    : num  8 9 8.5 8.5 8.5 8.5 8.5 8.5 8 8 ...
 $ Judge     : chr  "RUIZ-PEDREGUERA Rolando" "GEAR Dennis" "BOYS Beverley" "JOHNSON Ben
 $ JCountry  : chr  "CUB" "NZL" "CAN" "NOR" ...
```

```
> head(x)
```

```
  Event Round    Diver Country Rank DiveNo Difficulty JScore
1 M3mSB Final XIONG Ni     CHN    1      1        3.1    8.0
2 M3mSB Final XIONG Ni     CHN    1      1        3.1    9.0
3 M3mSB Final XIONG Ni     CHN    1      1        3.1    8.5
4 M3mSB Final XIONG Ni     CHN    1      1        3.1    8.5
5 M3mSB Final XIONG Ni     CHN    1      1        3.1    8.5
6 M3mSB Final XIONG Ni     CHN    1      1        3.1    8.5
                    Judge JCountry
1 RUIZ-PEDREGUERA Rolando      CUB
2             GEAR Dennis      NZL
3           BOYS Beverley      CAN
4           JOHNSON Bente      NOR
5         BOUSSARD Michel      FRA
6          CALDERON Felix      PUR
```

```
> attach(x)
```

```
> table(Event)

Event
M10mPF  M3mSB W10mPF  W3mSB
  2709   3192   2317   2569

> table(Round)

Round
 Final Prelim   Semi
  1848   6636   2303

> levels(factor(Judge))

 [1] "ALT Walter"                "BARNETT Madeleine"
 [3] "BOOTHROYD Sydney"          "BOUSSARD Michel"
 [5] "BOYS Beverley"             "BURK Hans-Peter"
 [7] "CALDERON Felix"            "CERMAKOVA Maria"
 [9] "CRUZ Julia"                "GEAR Dennis"
[11] "GEISSBUHLER Michael"       "HASSAN Mostafa"
[13] "HOOD Robin"                "HUBER Peter"
[15] "JOHNSON Bente"             "KELEMEN Ildiko"
[17] "LINDBERG Mathz"            "McFARLAND Steve"
[19] "MENA Jesus"                "RUIZ-PEDREGUERA Rolando"
[21] "SEAMAN Kathy"              "STEWART Anthea"
[23] "WANG Facheng"              "XU Yiming"
[25] "ZAITSEV Oleg"
```

It turns out that there are 4 types of game: M10mPF, M3mSB, W10mPF, W3mSB and there are 3 rounds: Final,Prelim and Semi. And look at a little bit about the Rank we can figure out how many players are there in each game and each round.

2 The biomodality of difficulty

Then we focus on the bimodality of the variable ”Difficulty”. After look at those small values of difficulty, it turns out the bimodel has something to do with the semi round. Perhaps the game requires every player to do a low difficulty dive in semifinals.

```
> small <- Difficulty <= 2.3
> head(x[small,])

     Event Round    Diver Country Rank DiveNo Difficulty JScore           Judge
2563 M3mSB  Semi XIONG Ni     CHN    1      1        1.6    8.0     GEAR Dennis
2564 M3mSB  Semi XIONG Ni     CHN    1      1        1.6    8.5    WANG Facheng
2565 M3mSB  Semi XIONG Ni     CHN    1      1        1.6    7.5      ALT Walter
2566 M3mSB  Semi XIONG Ni     CHN    1      1        1.6    7.5   JOHNSON Bente
```

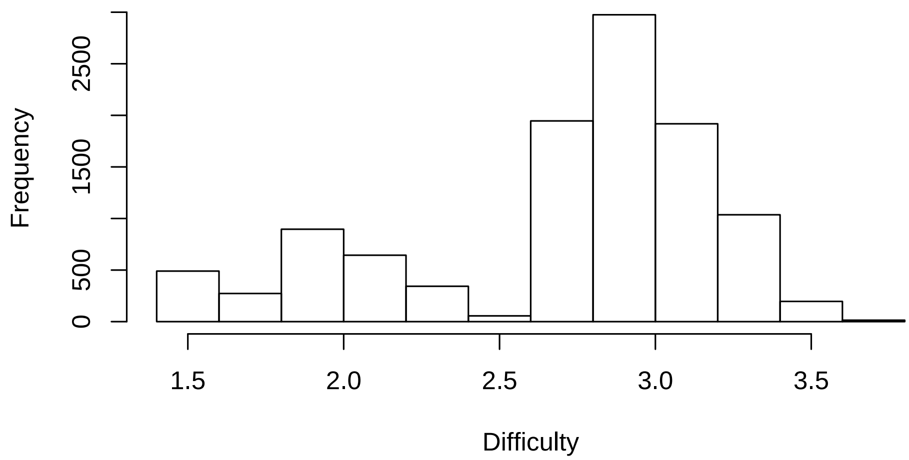

Figure 1: The histogram of difficulty.

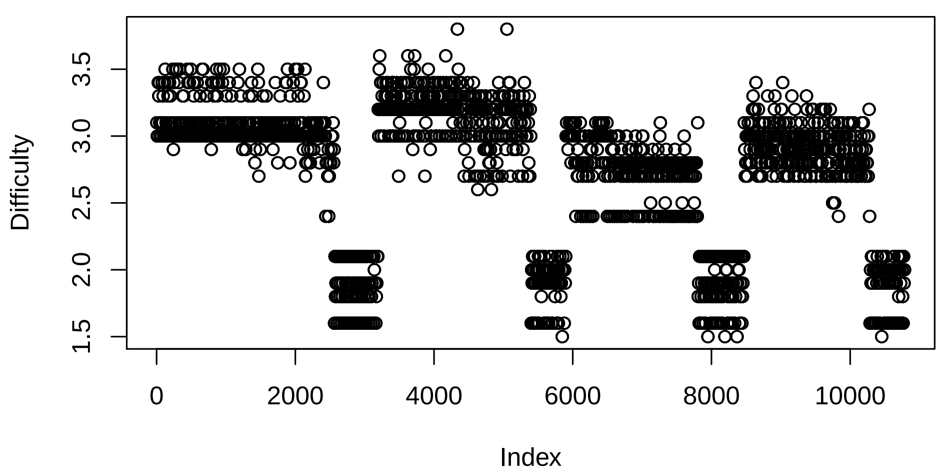

Figure 2: The plot of difficulty.

```
2567 M3mSB  Semi XIONG Ni     CHN    1      1          1.6    8.0 BOUSSARD Michel
2568 M3mSB  Semi XIONG Ni     CHN    1      1          1.6    7.5 McFARLAND Steve
     JCountry
2563      NZL
2564      CHN
2565      GER
2566      NOR
2567      FRA
2568      USA
```

```
> library(ggplot2)
```

3 The analysis of Score

Now we turn to the varible JScore. It seems there is no correlation between gender and difficulty or the judge's scores. And there is also no obivious correlation between difficulty and scores. Now we try to analysis whether the judges has bias in giving the score.

Then we look specifically at this judges bias of each country's diver. It is quite interesting that this chinese judge does not favor much to chinese divers.

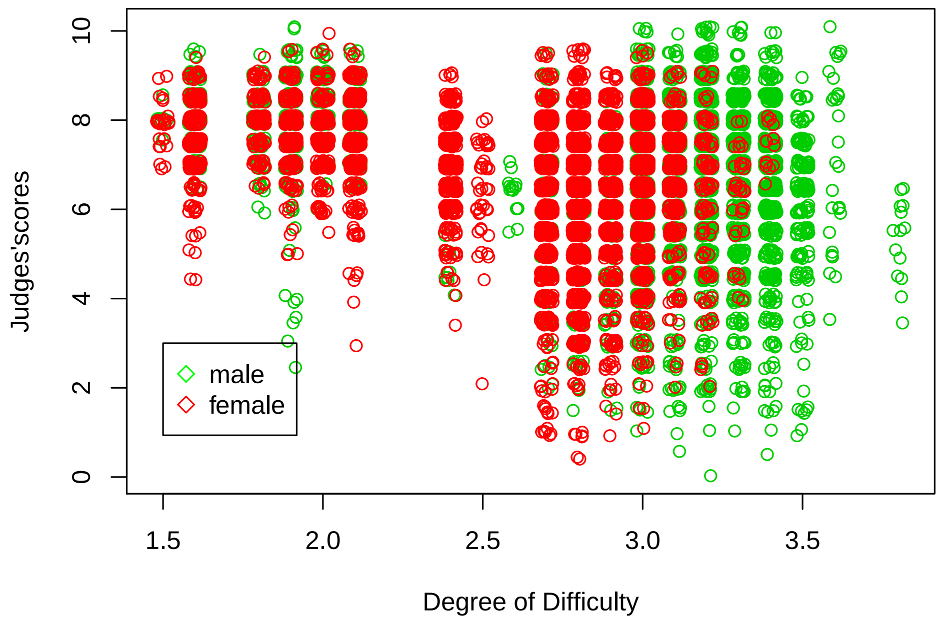

Figure 3: Judge's score and difficulty with respect to gender

We use the simplest additive model. Every player's score s is divided into three part: the true score t, judge's bias b and error e. that is

$$s = t + b + e$$

Here I just give a roughly estimation. First use 7 judges' score mean to esimate the player's true score. Then substract the mean score to esimate the judge's "bias" on a specific dive. By looking at the distribution of judge's bias (how close it to normal) we can figure out if whether he or she has bias. We just take a quick look at the "bias" mean. It seems number 23 judges has bias.

```
> mscore <- rep(0, length(JScore))
> for(i in 1: (length(JScore)/7)) {
+   mscore[(7*(i-1)+1):(7*(i-1)+7)] <-rep(mean(JScore[(7*(i-1)+1):(7*(i-1)+7)]), 7)
+ }
> bias <- list()
> for(judge in levels(factor(Judge))){
+         a <- Judge == judge
+         bias[[judge]] <- JScore[a]-mscore[a]
+ }
> sapply(1:length(levels(factor(Judge))), f <- function(x){mean(bias[[x]])})

 [1] -0.062535858 -0.089907067  0.049878345 -0.107855108  0.063086548
 [6] -0.062893082 -0.069834628 -0.028875380 -0.011757790  0.029336735
[11] -0.007659423  0.060030395  0.121031746  0.010284218  0.018576661
[16]  0.035383598  0.039057002  0.022613612 -0.047229640  0.013216513
[21]  0.004575496  0.088050314  0.173069228  0.054397560 -0.001440576
```