R Scripting

Lab for unit 1 - R basics and data structures (I)

Marcus Wurzer

21. September 2023

Please solve the following problem!

Biased Roulette

1. In the game of European roulette, a wheel is spun in one direction, then a ball is spun in the opposite direction. The ball eventually falls into one of 37 colored and numbered pockets on the wheel. Numbers 1–36 are classified into groups of numbers in three different ways:

    1. Low (1-18) vs. high (19-36)
    2. Odd (1, 3, ..., 35) vs. even (2, 4, ..., 36)
    3. Red (1, 3, 5, 7, 9, 12, 14, 16, 18, 19, 21, 23, 25, 27, 30, 32, 34, 36) vs. black (2, 4, 6, 8, 10, 11, 13, 15, 17, 20, 22, 24, 26, 28, 29, 31, 33, 35)

    Players may choose to place bets on one of these groups (e.g., put their money on low, red, uneven, etc.) and get double their money back if they win. Number 37 is the green zero.

    a. Create a vector that can be used to simulate a biased roulette wheel, where the ball is three time more likely to fall on the 0 than on any other number.
    b. "Spin" the wheel $n = 30, 300, 3000$ times and record the results. Make sure that the results are replicable!
    c. Determine the indices for all the (green) zero results in each of the three simulations and assign them to three different objects.
    d. Identify the lengths of the three vectors created in c. above.
    e. Compute the *simulated* occurrence probabilities for the (green) zero results and save them in a vector of length 3 with appropriately named elements. Compare them with the *expected* occurrence probabilities of a **fair** roulette wheel. To what extent do the probabilities differ?
    f. Based on the results, is there evidence to say that the roulette wheel is biased? Does the number of replications have an influence on your assessment? (Note: Usually, we wouldn't base our judgment solely on descriptive statistics, but would perform a statistical test and/or compute a confidence interval.)

    Hint:

    - Use `?sample` to find out how to sample in the case of unequal probability weights.

```r
# a.
roulette <- 0:36

# b.
set.seed(1)
res1 <- sample(roulette, 30, replace = TRUE, prob = c(3/39, rep(1/39, 36)))
res2 <- sample(roulette, 300, replace = TRUE, prob = c(3/39, rep(1/39, 36)))
res3 <- sample(roulette, 3000, replace = TRUE, prob = c(3/39, rep(1/39, 36)))
```

```r
# unequal probability weights -> biased

# c.
(ind1 <- which(res1 == 0))
```

```
## [1] 10 27
```

```r
(ind2 <- which(res2 == 0))
```

```
##  [1]  17  25  62  86 103 128 137 198 237 241 251 255 269 279
```

```r
(ind3 <- which(res3 == 0))
```

```
##   [1]   22   37   79  102  108  111  120  125  130  133  158  161  170  183  202
##  [16]  206  223  228  254  257  290  308  317  333  359  376  387  391  400  404
##  [31]  410  443  447  460  462  472  480  496  507  517  523  526  531  534  535
##  [46]  540  546  561  562  564  568  569  576  577  579  580  584  585  594  613
##  [61]  630  646  647  676  691  721  742  748  754  767  783  827  846  850  859
##  [76]  860  870  883  888  890  904  905  913  920  930  939  951  961  965  970
##  [91]  977  981  998  999 1017 1021 1034 1055 1063 1078 1083 1096 1101 1123 1128
## [106] 1151 1152 1171 1181 1204 1212 1213 1234 1267 1272 1273 1289 1297 1305 1307
## [121] 1310 1322 1353 1367 1370 1377 1402 1405 1410 1426 1448 1467 1471 1484 1495
## [136] 1510 1516 1532 1537 1568 1572 1586 1589 1595 1599 1605 1616 1619 1648 1667
## [151] 1678 1681 1685 1688 1694 1695 1697 1698 1714 1740 1746 1764 1791 1794 1798
## [166] 1801 1805 1823 1825 1858 1860 1861 1894 1909 1912 1925 1926 1954 1971 1996
## [181] 1999 2004 2032 2059 2067 2071 2072 2075 2077 2080 2091 2144 2149 2154 2155
## [196] 2164 2172 2178 2231 2238 2262 2294 2324 2327 2328 2351 2354 2371 2395 2424
## [211] 2430 2446 2485 2486 2487 2504 2524 2540 2553 2565 2608 2611 2615 2616 2630
## [226] 2640 2642 2656 2691 2708 2711 2728 2735 2742 2759 2772 2782 2786 2792 2801
## [241] 2813 2828 2835 2842 2843 2859 2867 2893 2902 2914 2921 2927 2932 2933 2936
## [256] 2942 2948 2974 2987 2998
```

```r
# d.
length(ind1)
```

```
## [1] 2
```

```r
length(ind2)
```

```
## [1] 14
```

```r
length(ind3)
```

```
## [1] 260
```

```r
# f.
(res.counts <- c(rep30 = length(ind1), rep300 = length(ind2), rep3000 = length(ind3)))
```

```
##   rep30  rep300 rep3000
##       2      14     260
```

```r
sim.prob <- res.counts / c(length(res1), length(res2), length(res3)) # simulated probabilities
3/39 # unknown probability for green zeros for the analyzed wheel
```

```
## [1] 0.07692308
```

```r
(exp.prob <- 1/37) # expected probability for green zeros when the wheel is fair
```

```
## [1] 0.02702703
```

```r
round(sim.prob - exp.prob, 4) # differences between fair wheel and analyzed wheel
```

```
##   rep30  rep300 rep3000
##  0.0396  0.0196  0.0596
```

```r
round(sim.prob - 3/39, 4) # differences between simulated probabilities and true parameter
```

```
##   rep30  rep300 rep3000
## -0.0103 -0.0303  0.0097
```

```r
# The highest numer of replications gives us a value that is closest to the
# true parameter and shows the highest discrepancy whe compared to a fair wheel, i.e.,
# there is indication that the wheel is biased.
# On the average, the number of replications will influence the precision of the
# estimation **on the average**, i.e., we could also expect that 300 replications
# give us a better indication than 30. Using the result at hand, this is not the case,
# but this can be considered as coincidental: If we would replicate this experiment
# over and over again (30 spins, then count zeroes, another 30 spins, again counting the zeroes
# and so on), we would see a larger variation in the results, compared to higher numbers
# of spins - and of course, this will also include results that are quite close to the
# true value by chance.
```