Final Exam

3190300985 LUIS LUZERN YUVEN

2022-07-04

1

```
tmp <- c(4,6,3)
(tmp1 <- rep(tmp, times = 10))
```

```
##  [1] 4 6 3 4 6 3 4 6 3 4 6 3 4 6 3 4 6 3 4 6 3 4 6 3 4 6 3 4 6 3
```

```
(tmp2 <- rep(tmp, times = c(10,20,30)))
```

```
##  [1] 4 4 4 4 4 4 4 4 4 4 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 3 3 3 3 3 3 3 3
## [39] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
```

2

```
xVec <- sample(0:999, 250, replace=T)
yVec <- sample(0:999, 250, replace=T)
```

(a)

```
vec <- vector(length = 249)
for(i in 1 : 249) vec[i] <- yVec[i+1] - xVec[i]
```

(b)

```
yVec[which(yVec >600)]
```

```
##   [1] 908 694 860 798 679 967 699 928 970 707 621 912 642 893 936 799 795 643
##  [19] 854 808 880 861 994 826 879 996 697 697 890 897 928 966 742 694 853 905
##  [37] 774 988 690 967 769 796 716 839 656 945 959 604 822 727 758 817 605 879
##  [55] 811 827 823 848 630 736 724 646 871 615 748 962 708 967 683 721 903 649
##  [73] 831 895 866 876 876 702 753 915 953 831 890 800 640 779 956 701 728 632
##  [91] 987 737 995 657 842 665 673 935 916 985 633 683 975 697 872 834 898 784
## [109] 682 776 899 846
```

(c)

```
which(yVec > 600)
```

```
##  [1]   3   5   7  10  11  14  17  18  19  20  23  25  26  27  31  32  34  35
## [19]  36  38  41  45  46  49  51  53  57  60  62  66  67  68  70  72  76  77
## [37]  81  83  84  89  95  97  98  99 100 101 103 106 107 112 114 116 117 121
## [55] 122 123 128 130 131 132 133 135 137 138 139 140 150 151 152 153 154 155
```

```
##  [73] 157 162 163 167 168 169 170 171 176 179 181 183 185 188 189 193 194 198
##  [91] 199 204 207 209 213 214 217 220 222 225 227 230 232 233 234 235 236 239
## [109] 241 243 244 250
```

(d)

```r
xVec[order(match(xVec,yVec))]
```

```
##   [1] 400 679 679 970 642 101 808 466 466  81 861 994 380 380 147 590 485 485
##  [19] 485 405 591 988 690 345 495 495 161  17 817 509 509 823 371 371 279 412
##  [37] 412 953  25 842 665 486 317 935 916 549 985 985 187 187  99  21 780 730
##  [55] 751 390 145  40 197 551 544 947 747 740 325 696 284  48 628 259 455 663
##  [73] 479 426 686 698 374 944 937 302 585 752 254  85 840 990 301 725 562 178
##  [91] 545 637 504 199 675 914 472 734  14 520 531 334 419 789  61 309 553 373
## [109] 195 124  39 599 171 983 512   5 413 143 533 955 169 528 476 801 327 339
## [127] 704 499 369  67 201 692 406 259 850  89 847 964 538 183 904 550  29 567
## [145] 177 894 293 718 492 926 583 759 585 993 681 513 370 867 958 536 327 437
## [163] 166 789 828 744 525 593 674 775 644 252 327 984 923 307  27 676 731 599
## [181] 533 499   4  56 335 767 645 498 734 886 113 841 625  94 704 261 206 375
## [199]  36 634  48 456 401 700 943 788 245 478 709 458 700  54 620 220 169 375
## [217] 692 586 547 999 810 142 550 937  98 954 997  13 703 268   2 666 175 336
## [235] 398  49 227 269 318 914 142 426 219 907 511 534 768 593 331 511
```

(e)

```r
yVec[seq(1,250,3)]
```

```
##  [1]  74 400 860 798 106 128 970 368 912 100 936 795 358 466 555 994 826  44 537
## [20] 380 133 510 928 742   9 853 485 591 468 411 204 162 796 656 959 604 506 727
## [39] 129 493 879 449 186 848 724 535 748 359 386 279 967 903 831 133 866 482 702
## [58] 468 450 363 890 125 330 111 701 573 987 340 146  12 539 665 673 935 549 154
## [77] 218 975 834 187 682 899 249 846
```

3

```r
prod <- seq(2,38,2) / seq(3,39,2)
1 + sum(cumprod(prod))
```

```
## [1] 6.976346
```

4

(a)

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
data(state)
state.x77 <- as.data.frame(state.x77)
dim(state.x77 %>% filter(Income < 4300))
```

```
## [1] 20  8
```

There are 20 states with income less than 4300

(b)

```
state.x77 %>% filter(Income == max(Income))
```

```
##        Population Income Illiteracy Life Exp Murder HS Grad Frost   Area
## Alaska        365   6315        1.5    69.31   11.3    66.7   152 566432
```

Alaska is the state with the highest income

(c)

```
state.x77 <- state.x77 %>% mutate(Illit.level = ifelse(Illiteracy < 2,
                                          ifelse(Illiteracy < 1, "low", "some"),
                                          "high"))
```

(d)

```
state.x77 %>% filter(Illit.level == "low") %>% filter(Income == max(Income))
```

```
##          Population Income Illiteracy Life Exp Murder HS Grad Frost Area
## Maryland       4122   5299        0.9    70.22    8.5    52.3   101 9891
##          Illit.level
## Maryland         low
```

Maryland is the state with the highest income among states with low illiteracy. Its income is 5299

5

(a)

```
x1 <- runif(1000)
x2 <- runif(1000)
dist.edge <- vector(length=1000)
for(i in 1:1000){
  dist1 <- sqrt(x1[i]^2 + min(x2[i]-seq(0,1,0.01))^2)       # left edge (x1 = 0)
  dist2 <- sqrt((x1[i]-1)^2 + min(x2[i]-seq(0,1,0.01))^2)  # right edge (x1 = 1)
  dist3 <- sqrt(min(x1[i]-seq(0,1,0.01))^2 + x2[i]^2)       # bottom edge (x2 = 0)
  dist4 <- sqrt(min(x1[i]-seq(0,1,0.01))^2+ (x2[i]-1)^2)   # top edge (x2 = 1)
  dist.edge[i] <- min(c(dist1,dist2,dist3,dist4))
}
(prob <- length(which(dist.edge < 0.25)) / 1000)
```

```
## [1] 0.157
```

(b)

```
dist.vertex <- vector(length=1000)
for(i in 1:1000){
  dist1 <- sqrt(x1[i]^2 + (x2[i])^2)       # (0,0)
  dist2 <- sqrt((x1[i]-1)^2 + (x2[i])^2)   # (1,0)
  dist3 <- sqrt((x1[i])^2 + (x2[i]-1)^2)   # (0,1)
  dist4 <- sqrt((x1[i]-1)^2+ (x2[i]-1)^2)  # (1,1)
  dist.vertex[i] <- min(c(dist1,dist2,dist3,dist4))
}
(prob <- length(which(dist.vertex < 0.25)) / 1000)
```

```
## [1] 0.204
```

6

```
val <- 0:4
pmf <- c(1,2,2,2,3)
draw <- sample(val, 1000, pmf, replace=T)
table(draw)
```

```
## draw
##   0   1   2   3   4
##  92 221 213 200 274
```

We see that the results are close to the theoretical probabilities, where the frequencies are 100,200,200,200,300 for `x=0, x=1, x=2, x=3, x=4` respectively.

7

(a)

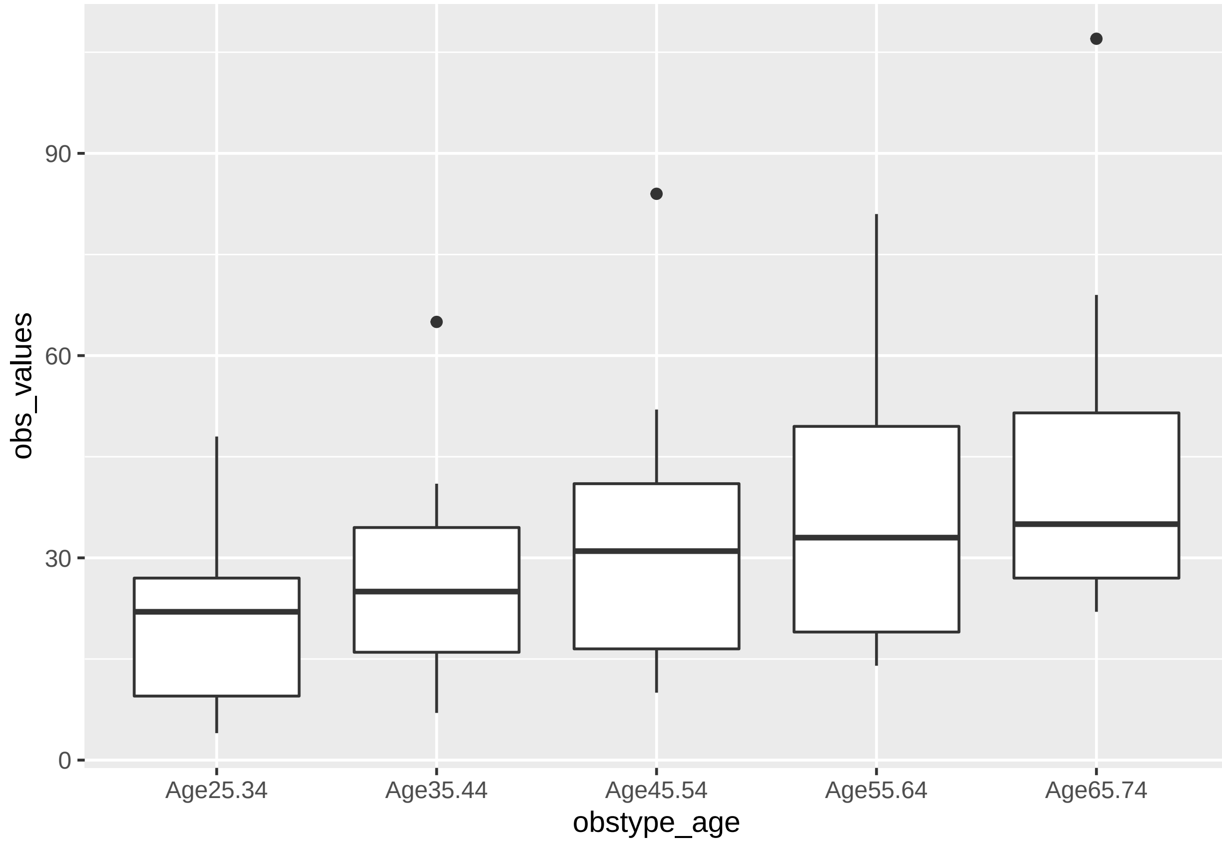
```
suicrates <- tibble(Country = c('Canada', 'Israel', 'Japan', 'Austria', 'France',
                                'Germany', 'Hungary', 'Italy', 'Netherlands',
                                'Poland', 'Spain', 'Sweden', 'Switzerland', 'UK',
                                'USA'),
Age25.34 = c(22, 9, 22, 29, 16, 28, 48, 7, 8, 26, 4, 28, 22, 10, 20),
Age35.44 = c(27, 19, 19, 40, 25, 35, 65, 8, 11, 29, 7, 41, 34, 13, 22),
Age45.54 = c(31, 10, 21, 52, 36, 41, 84, 11, 18, 36, 10, 46, 41, 15, 28),
Age55.64 = c(34, 14, 31, 53, 47, 49, 81, 18, 20, 32, 16, 51, 50, 17, 33),
Age65.74 = c(24, 27, 49, 69, 56, 52, 107, 27, 28, 28, 22, 35, 51, 22, 37))

(suicrates_long <- suicrates %>%
  gather(key = obstype_age, value = obs_values, -Country))
```

```
## # A tibble: 75 x 3
##    Country     obstype_age obs_values
##    <chr>       <chr>            <dbl>
##  1 Canada      Age25.34            22
##  2 Israel      Age25.34             9
##  3 Japan       Age25.34            22
##  4 Austria     Age25.34            29
##  5 France      Age25.34            16
##  6 Germany     Age25.34            28
##  7 Hungary     Age25.34            48
```

```
##   8 Italy       Age25.34             7
##   9 Netherlands Age25.34             8
##  10 Poland      Age25.34            26
## # ... with 65 more rows
```

(b)

```
suicrates_long %>% ggplot() + geom_boxplot(aes(x = obstype_age, y = obs_values))
```



The median of mortality rates for men aged 25-34, 35-44, 45-54, 55-64, and 65-74 are around 20, 25, 30, 33, and 35 respectively. The rates in `Age35.44` is the most balanced. There are outliers in `Age35.44`, `Age45.54`, and `Age65.74`

8

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```
data(steam)
# set \eps = 0.001
fit <- nls(Press ~ a*exp(b*Temp/(g+Temp))+0.001, data=steam, start=list(a=5,b=20,g=200))
pred <- predict(fit)
```

```
steam %>% ggplot(aes(x = Temp, y = pred)) + geom_line() +
  geom_line(aes(x = Temp, y = Press), col = "red") + labs(x = "Temp", y = "Press")
```