P2

Jeremiah Theisen

2024-09-10

```
loans = read.csv("https://github.com/TienChih/tbil-stats/raw/main/data/loans_full_schema.csv")

names(loans)
```

```
##  [1] "emp_title"                    "emp_length"
##  [3] "state"                        "homeownership"
##  [5] "annual_income"                "verified_income"
##  [7] "debt_to_income"               "annual_income_joint"
##  [9] "verification_income_joint"    "debt_to_income_joint"
## [11] "delinq_2y"                    "months_since_last_delinq"
## [13] "earliest_credit_line"         "inquiries_last_12m"
## [15] "total_credit_lines"           "open_credit_lines"
## [17] "total_credit_limit"           "total_credit_utilized"
## [19] "num_collections_last_12m"     "num_historical_failed_to_pay"
## [21] "months_since_90d_late"        "current_accounts_delinq"
## [23] "total_collection_amount_ever" "current_installment_accounts"
## [25] "accounts_opened_24m"          "months_since_last_credit_inquiry"
## [27] "num_satisfactory_accounts"    "num_accounts_120d_past_due"
## [29] "num_accounts_30d_past_due"    "num_active_debit_accounts"
## [31] "total_debit_limit"            "num_total_cc_accounts"
## [33] "num_open_cc_accounts"         "num_cc_carrying_balance"
## [35] "num_mort_accounts"            "account_never_delinq_percent"
## [37] "tax_liens"                    "public_record_bankrupt"
## [39] "loan_purpose"                 "application_type"
## [41] "loan_amount"                  "term"
## [43] "interest_rate"                "installment"
## [45] "grade"                        "sub_grade"
## [47] "issue_month"                  "loan_status"
## [49] "initial_listing_status"       "disbursement_method"
## [51] "balance"                      "paid_total"
## [53] "paid_principal"               "paid_interest"
## [55] "paid_late_fees"
```

2.2.1

a.

```
roll=sample(1:6,20,replace=TRUE)

roll
```

```
##  [1] 3 5 3 6 6 2 4 1 4 6 1 5 3 6 4 1 4 1 2 3
```

3 4 4 3 4 3 3 4 3 2 6 2 4 5 2 6 5 1 4 5 b. A = {4, 4, 4, 4, 2, 6, 2, 4, 2, 6, 4} = 11 c. {4, 4, 4, 4, 6, 4, 6, 4} = 8 d. B = {4, 4, 4, 4, 6, 4, 5, 6, 5, 4, 5} = 11 e. {4, 4, 4, 4, 6, 4, 6, 4} = 8 f. 0.55 g. 0.727 h. 0.55 i. 0.727 j. {4, 4, 4, 4, 6, 4, 6, 4} = 8

2.2.2

a. B = {4, 5, 6} = 3 A = {2, 4, 6} = 3

b. {4, 6} = 2

c. 2/3

d. P(A) = 1/2. 1/3 chance for one that satisfies A and B. The chances of getting one that does not satify both is the same as the chance of getting one that satisfies both.

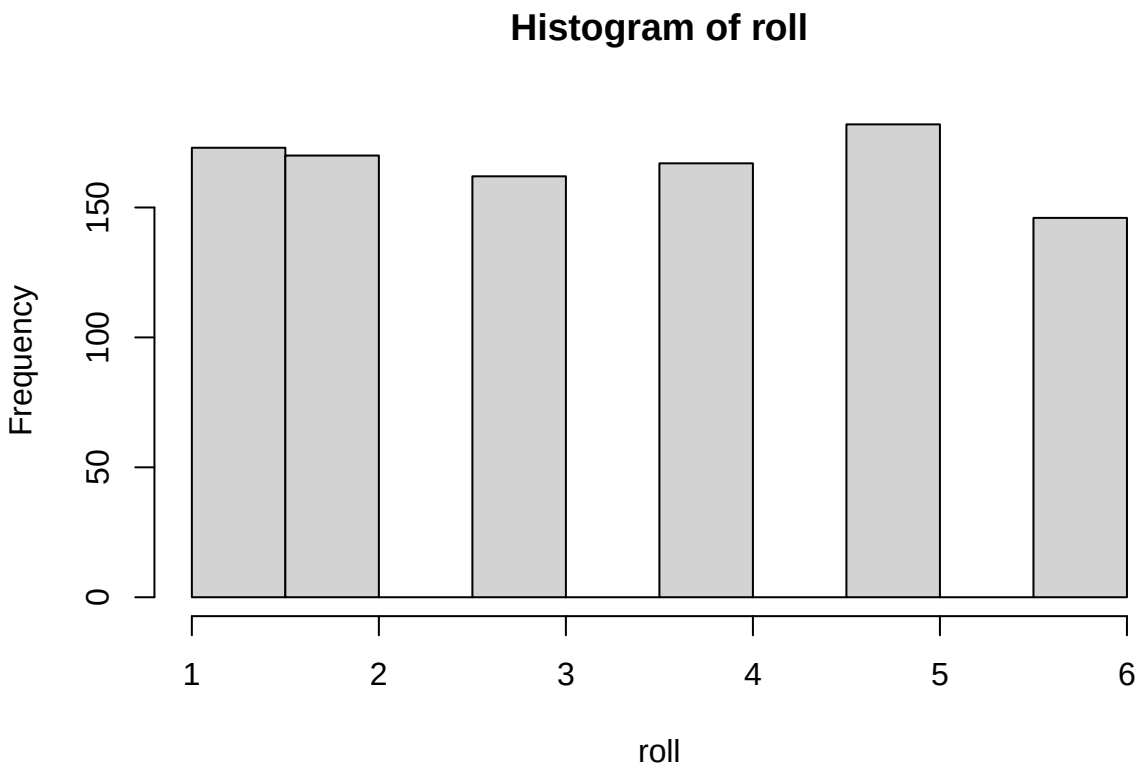
2.2.3

a. P(A) = 1/2, P(B) = 1/2, P(A and B) = 1/3
b. P(A|B) = 2/3 Calculation: (1/3 / 1/2)
c. 2/3 is the same probability I got in 2.2.2.c

2.2.4

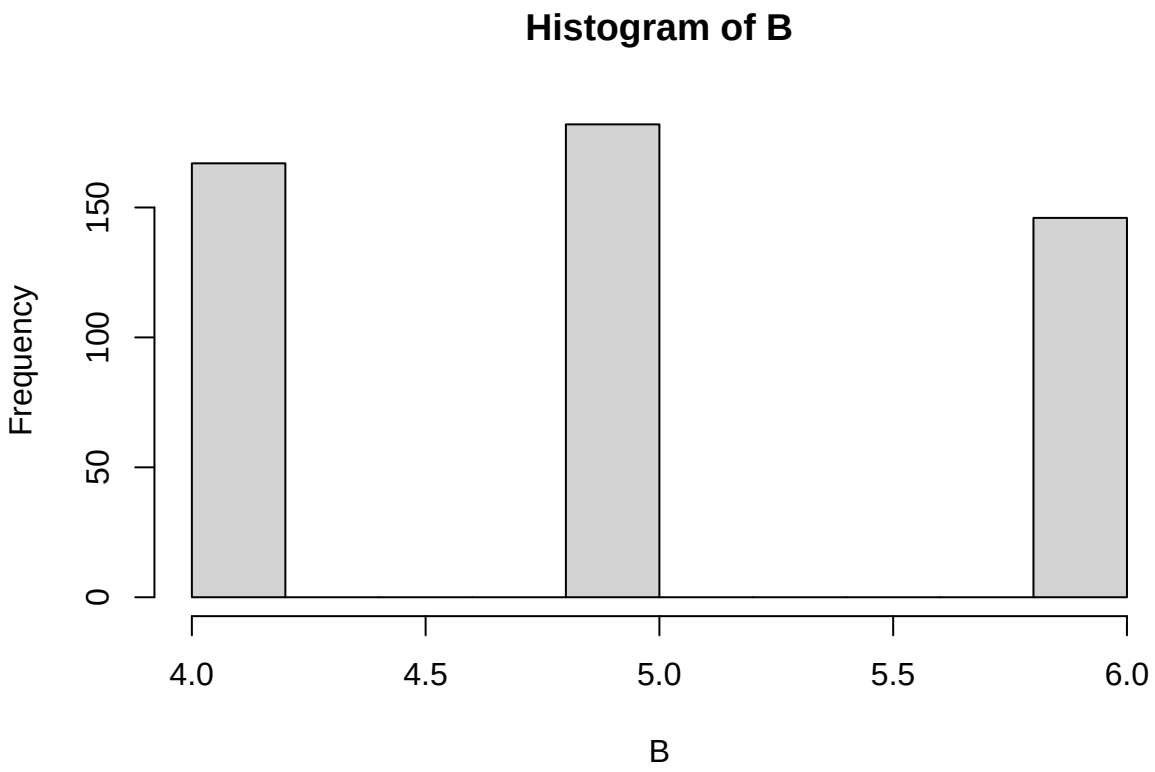
a.

```
roll=sample(1:6,1000,replace=TRUE)

hist(roll)
```



b.

```
B=subset(roll, roll>3)

hist(B)
```



c.

```
length(B)
```

```
## [1] 495
```

d.

```
length(which(B%%2==0))
```

```
## [1] 313
```

e. 0.668
f. It is very close

2.2.5

a. R = 20%, C = 25%, R and C = 30%. Since P(R and/or C) = 30%, there must be a 15% overlap. R but not C = 5% and C without R = 10%.
b. 10%
c. 5%

2.2.6

a.

```
length(which(loans$homeownership=="OWN"))
```

```
## [1] 1353
```

b. 0.1353
c.

```
gradeAloans=subset(loans, loans$grade=="A")
nrow(gradeAloans)
```

```
## [1] 2459
```

d.

```
length(which(gradeAloans$homeownership=="OWN"))
```

```
## [1] 347
```

e. 0.1411
f.

```
index = sample(1:nrow(loans), 1000)
samp=loans[index,]

table(samp$grade,samp$homeownership)
```

```
##
##     MORTGAGE OWN RENT
##   A      127  24   78
##   B      152  43  123
##   C      118  37  124
##   D       54  18   65
##   E       12   6   12
##   F        4   1    2
```

$35/140 = 0.25$

2.2.7

a. (5, x), where x does not matter because it is the green dice
b. P(A) = 6/36 = 1/6
c. (x, 2|4|6)
d. P(B) = 18/36 = 1/2
e. (5, 2), (5, 4), (5, 6), P(A and B) = 3/36 = 1/12
f. P(A|B) = 1/12 / 1/2 = 2/12 = 1/6
g. The probability does not appear to be affected

2.2.8

a.

```
bluedie = sample(1:6, 1000, replace = TRUE)
greendie = sample(1:6, 1000, replace = TRUE)

doubledie=data.frame(bluedie, greendie)

head(doubledie, n=10)
```

```
##    bluedie greendie
## 1        5        3
## 2        6        2
## 3        5        4
## 4        2        3
## 5        1        6
## 6        3        4
## 7        1        5
## 8        4        1
## 9        6        3
## 10       4        2
```

b.

```
length(which(doubledie$bluedie==5))
```

```
## [1] 171
```

$\mathrm{P(A)} = 0.162$

c.

```
length(which(doubledie$greendie%%2==0))
```

```
## [1] 506
```

$\mathrm{P(B)} = 0.481$ d.

```
greendieeven=subset(doubledie, doubledie$greendie%%2==0)

head(greendieeven, n=10)
```

```
##    bluedie greendie
## 2        6        2
## 3        5        4
## 5        1        6
## 6        3        4
## 10       4        2
## 11       4        2
## 15       6        2
## 16       3        6
## 17       2        4
## 19       3        4
```

e.

```
length(which(greendieeven$bluedie==5))
```

```
## [1] 79
```

P(A and B) = 0.073, P(A|B) = 0.152

2.2.9

a. Yes, because P(A|B) = P(A and B), P(A and B) = P(A)P(B)
b. No. P(A) and P(B) overlap too much.
c. Maybe. The proportion of homeowners with the loan, vs the proportion of people who got grade A loans total are very close (0.141 vs 0.135), but is not be close enough to state definitely.
d. Yes, P(A|B) = P(A), P(A and B) = P(A)P(B)

2.2.10

a. Suppose what?
b. P(S) = 0.25
c. P(A) = 0.077
d. P(S and A) = 0.019
e. Yes, P(S and A) = P(A)P(S) and P(A) = 1/13 and P(A|S) = 1/13
f. It would be dependent.If you drew an ace, the chance would be 3/51, if not, 4/51

2.2.11

a. It would be independent because the provious draw has no way of effecting the outcome of the second draw, since the card has been placed back in the deck and reshuffled.
b. 0.25
c. 0.25
d. 0.0625
e. 0.0059