Homework 7

Miles Tweed

October 26, 2020

Problem 1

5.30

a

```r
total.lowest.inc <- 0.9556 + 0.0085
prob.audit.lowest <- 0.0085 / total.lowest.inc
print(paste('Probability of being audited in the lowest income category:',
            round(prob.audit.lowest,4)))
```

```
## [1] "Probability of being audited in the lowest income category: 0.0088"
```

b

```r
total.audit <- 0.0085 + 0.0009 + 0.003
prob.lowest.audit <- 0.0085 / total.audit
print(paste('Probability of being in the lowest income category if audited:',
            round(prob.lowest.audit,3)))
```

```
## [1] "Probability of being in the lowest income category if audited: 0.685"
```

5.31

a

```r
total.christian <- 57199 + 36148 + 16834 + 11366 + 51855
total.population <- 228182
prob.christian <- total.christian/total.population
print(paste("The probability that a random individual is Christian:",
            round(prob.christian, 3)))
```

```
## [1] "The probability that a random individual is Christian: 0.76"
```

b

```r
prob.catholic <- 57199 / total.christian
print(paste("Prabability of being Catholic given that they are Christian:",
            round(prob.catholic,3)))
```

```
## [1] "Prabability of being Catholic given that they are Christian: 0.33"
```

c

```r
total.answ <- total.population - 11815
prob.no.rel <- 34169 / total.answ
print(paste("Probability of no religion given that they answered",
            round(prob.no.rel, 3)))
```

```
## [1] "Probability of no religion given that they answered 0.158"
```

5.39

a

```r
tot.pop <- 317
tot.very.happy <- 147
prob.very.happy <- tot.very.happy / tot.pop
print(paste("The prodability or being very happy is:",
            round(prob.very.happy, 3)))
```

```
## [1] "The prodability or being very happy is: 0.464"
```

b

```r
tot.male <- 146
tot.female <- 171
very.happy.m <- 69
very.happy.f <- 78

# (i)
prob.vh.m <- very.happy.m / tot.male

# (ii)
prob.vh.f <- very.happy.f / tot.female

print(paste("(i) Probability of being very happy for males:",
            round(prob.vh.m, 3)))
```

```
## [1] "(i) Probability of being very happy for males: 0.473"
```

```r
print(paste("(ii) Probability of being very happy for females",
            round(prob.vh.f, 3)))
```

```
## [1] "(ii) Probability of being very happy for females 0.456"
```

c

Yes, the events of being very happy and being male are independent because the total probability $\approx$ probability for males $\approx$ probability for females ≈ 0.46.

5.48

a

The probability will be less than in exercise 13 because every student will have the same probability of having a different birth date. This means that the probability of every student being different is greater which means that the compliment probability of one student being the same will be less.

b

```r
prob.diff <- 364/365
prob.all.diff <- prob.diff^24
prob.same <- 1-prob.all.diff
print(paste("The probability that at least one student will have that birthday:",
            round(prob.same,4)))
```

```
## [1] "The probability that at least one student will have that birthday: 0.0637"
```

5.50

a

Since there are only four holes with a non-zero probability of a hole-in-one, the probability that a golfer will not get any holes in one is the compliment of the probability of getting a hole in one for every hole.

```
prob.all.HIO <- 0.0005 * 0.0015 * 0.0005 * 0.0025
prob.no.HIO <- 1-prob.all.HIO

print(paste("The probability of having no holes in one is:",
            prob.no.HIO))
```

```
## [1] "The probability of having no holes in one is: 0.999999999999063"
```

b

This would follow the multiplicative rule because all of the rounds are independent.

```
prob.no.HIO.20round <- prob.no.HIO^20
print(paste("The probability of not having a hole in one in 20 rounds:",
            prob.no.HIO.20round))
```

```
## [1] "The probability of not having a hole in one in 20 rounds: 0.999999999981251"
```

c

The probability of at least one hole in one will be the compliment of the probability of not holes in one in 20 round.

```
prob.oneormore.HIO <- 1- prob.no.HIO.20round

print(paste("The probability of at least one hole in one in twenty rounds:",
            prob.oneormore.HIO))
```

```
## [1] "The probability of at least one hole in one in twenty rounds: 1.87494464398696e-11"
```

5.57

a

```
prob.s.and.pos <- 0.01*0.86
prob.s.and.neg <- 0.01*0.14
prob.sc.and.pos <- 0.99*0.12
prob.sc.and.neg <- 0.99*0.88

print(paste("The intersection probability P(S and POS):",
            prob.s.and.pos))
```

```
## [1] "The intersection probability P(S and POS): 0.0086"
```

```
print(paste("The intersection probability P(S and NEG):",
            prob.s.and.neg))
```

```
## [1] "The intersection probability P(S and NEG): 0.0014"
```

```
print(paste("The intersection probability P(Sc and POS):",
            prob.sc.and.pos))
```

```
## [1] "The intersection probability P(Sc and POS): 0.1188"
```

```r
print(paste("The intersection probability P(Sc and NEG):",
            prob.sc.and.neg))
```

```
## [1] "The intersection probability P(Sc and NEG): 0.8712"
```

b

```r
prob.pos <- prob.s.and.pos + prob.sc.and.pos

print(paste("The probability of having a positive test result:",
            prob.pos))
```

```
## [1] "The probability of having a positive test result: 0.1274"
```

c

```r
prob.s.pos <- prob.s.and.pos / prob.pos

print(paste("The probability of having breast cancer given a positive test:",
            round(prob.s.pos,4)))
```

```
## [1] "The probability of having breast cancer given a positive test: 0.0675"
```

d

In the first branch only one percent of the 100 people would have the condition, therefore the frequencies of having the condition or not is 1 and 99 respectively. Of those that have the condition, since the sensitivity of the test is 0.86 we would expect that it would be more likely than not that one person with the condition would test positive which would leave no one to have a false negative. Of those without the condition, the specificity is 0.88 which mean that 12% of the 99 would have false positives while the rest would have true negatives. This leaves the frequencies of false positives and true negatives at 12 and 87 respectively.

The probabilities are calculated by taking the number of occurrences of interest and dividing them by the total number of occurrences of that type. For example, in this case the number of positive test results in which the person had the condition is one and the total number of positive results is 13.

```r
prob.s.pos <- 1/13
print(paste("Probability of having the condition given a positive result",
            round(prob.s.pos,4)))
```

```
## [1] "Probability of having the condition given a positive result 0.0769"
```

5.62

a

```r
tot.down <- 54
tot.pop  <- 5282
prob.down <- tot.down / tot.pop

print(paste("The probability that Down syndrom occurs:",
            round(prob.down,4)))
```

```
## [1] "The probability that Down syndrom occurs: 0.0102"
```

b

```
down.pos <- 48
not.down.neg <- 3921
tot.not.down <- 5228

sens <- down.pos / tot.down
spec <- not.down.neg / tot.not.down

print(paste("(i) The sensitivity is:", round(sens,3),
            "and (ii) the specificity is:", round(spec,3)))
```

```
## [1] "(i) The sensitivity is: 0.889 and (ii) the specificity is: 0.75"
```

c

```
prob.no.down <- 1 - prob.down
prob.yes.and.pos <- prob.down * sens
prob.no.and.pos <- prob.no.down * (1 - spec)
prob.yes.and.neg <- prob.down * (1-sens)
prob.no.and.neg <- prob.no.down * spec
prob.test.pos <- prob.yes.and.pos + prob.no.and.pos
prob.test.neg <- prob.no.and.neg + prob.yes.and.neg

prob.yes.pos <- prob.yes.and.pos / prob.test.pos
prob.no.neg <- prob.no.and.neg / prob.test.neg

print(paste("The probability of Down syndrome given a positive test:",
            round(prob.yes.pos,4)))
```

```
## [1] "The probability of Down syndrome given a positive test: 0.0354"
```

```
print(paste("The probability of no Down syndrome given a negative test:",
            round(prob.no.neg,4)))
```

```
## [1] "The probability of no Down syndrome given a negative test: 0.9985"
```

d

The sensitivity describes how likely the test is to correctly identify someone with the condition and specificity describes how likely the test is to correctly identify someone without the condition. $P(S|POS)$ describes the probability of a true positive given a positive result while $P(S^c|NEG)$ describes the probability of a true negative given a negative result.

6.5

a

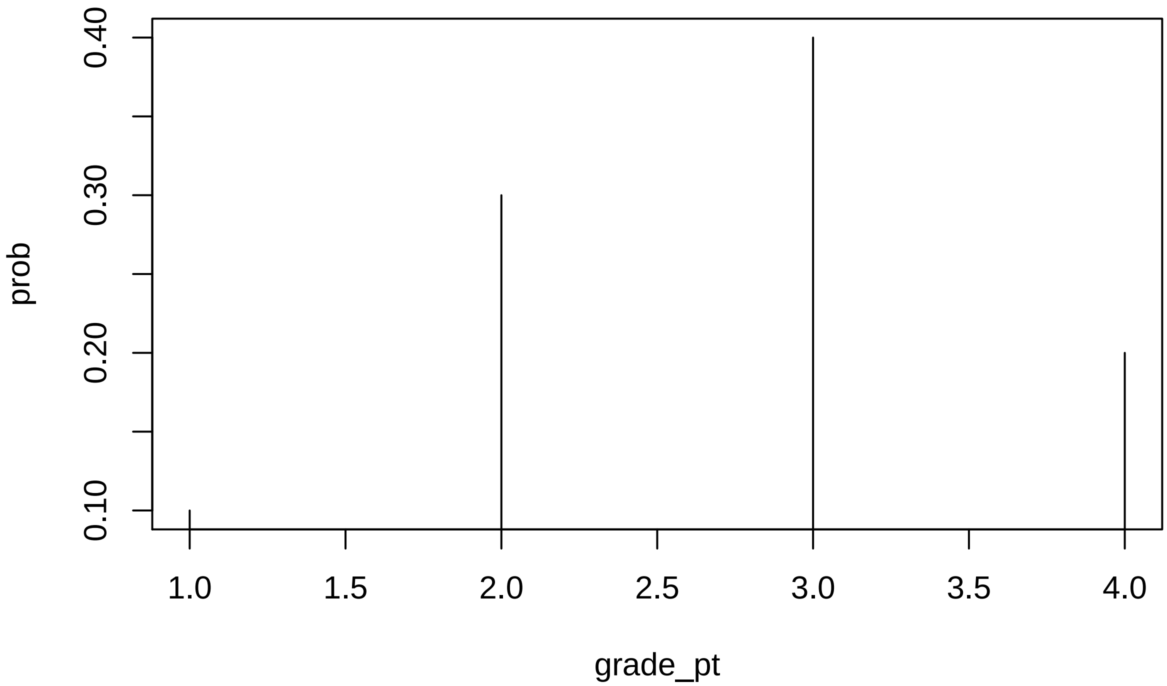
The Probability distribution is the potential values for grade point average and the probabilities of each.

```
prob.dist <- data.frame(c(4,3,2,1), c(0.2, 0.4, 0.3, 0.1))
names(prob.dist) <- c("grade_pt", "prob")
prob.dist
```

```
##   grade_pt prob
## 1        4  0.2
## 2        3  0.4
## 3        2  0.3
```

```
## 4        1  0.1
```

```
plot(prob ~ grade_pt, data = prob.dist, type = 'h')
```



b

The mean of the probability distribution is the grade point that a random student would expect to receive on average. The value (2.7) is between a C and a B.

```
prob.dist.mean <- (4*0.2 + 3*0.4 + 2*0.3 + 1*0.1)
print(paste("The mean of the probability distribution is:",
            prob.dist.mean))
```

```
## [1] "The mean of the probability distribution is: 2.7"
```

6.12

a

The sample space includes (i) winning both ($50), (ii) winning the first and losing the second ($30), (iii) losing the first and winning the second ($20), (iv) and losing both auctions ($0).

b

```
prob.first <- 0.1
prob.second <- 0.2

# (i)
prob.win.both <- prob.first * prob.second
print(paste("The probability of winning both:",
            prob.win.both))
```

```
## [1] "The probability of winning both: 0.02"
```

```r
# (ii)
prob.win.first <- prob.first * (1-prob.second)
print(paste("The probability of winning the first only:",
            prob.win.first))
```

```
## [1] "The probability of winning the first only: 0.08"
```

```r
# (iii)
prob.win.second <- (1 - prob.first) * prob.second
print(paste("The probability of winning the second only::",
            prob.win.second))
```

```
## [1] "The probability of winning the second only:: 0.18"
```

```r
# (iv)
prob.lose.both <- (1 - prob.first) * (1 - prob.second)
print(paste("The probability of losing both:",
            prob.lose.both))
```

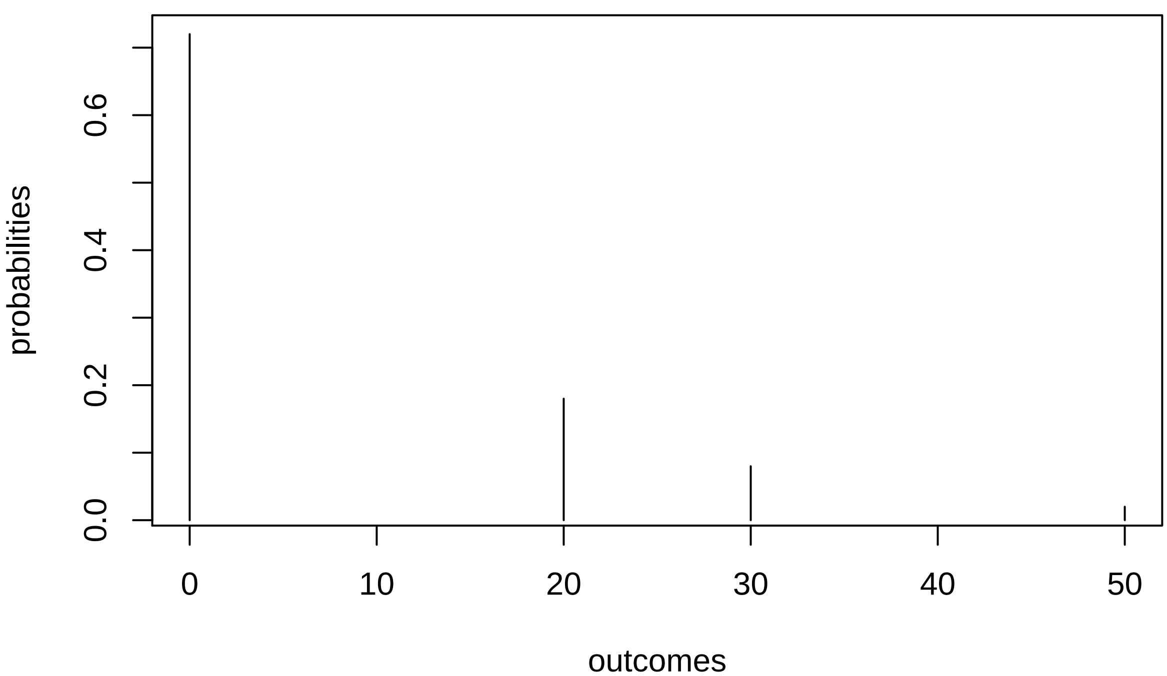
```
## [1] "The probability of losing both: 0.72"
```

c

```r
auction.prob.dist <- data.frame(c(50, 30, 20, 0), c(0.02, 0.08, 0.18, 0.72))
names(auction.prob.dist) <- c("outcomes", "probabilities")
auction.prob.dist
```

```
##   outcomes probabilities
## 1       50          0.02
## 2       30          0.08
## 3       20          0.18
## 4        0          0.72
```

```r
plot(probabilities ~ outcomes, data = auction.prob.dist, type = 'h')
```



d

```
auction.mean <- (50*0.02 + 30*0.08 + 20*0.18 + 0*0.72)
print(paste("The mean of X: $", auction.mean))
```

```
## [1] "The mean of X: $ 7"
```

Problem 2

Part 1

Using:

$\mu = E[X] = \sum_{i=1}^{k\ (or\ \infty)} x_i P(X = x_i)$

Prove:

$E[aX + b] = aE[X] + b$

Answer:

$E[aX + b] = \sum_{i=1}(ax_i + b)P(x_i)$

$= \sum_{i=1}(ax_i \cdot P(x_i) + b \cdot P(x_i))$

$= \sum_{i=1} ax_i \cdot P(x_i) + \sum_{i=1} b \cdot P(x_i)$

$= a\left(\sum_{i=1} x_i \cdot P(x_i)\right) + b\left(\sum_{i=1} P(x_i)\right)$

Since $\sum_{i=1} P(x_i) = 1$

$= a \cdot E(X) + b$

Part 2

Using:

$V[X] = E(X - E(X))^2 = \sum_i (x_i - E[X])^2 P(X = x_i)$

Prove:

$V[aX + b] = a^2 V[X]$

Answer:

$V[aX + b] = V[X'] = E[X'^2] - (E[X']^2) = E((aX + b)^2) - (E(aX + b))^2$

$= E(a^2X^2 + 2abX + b^2) - (aE(X) + b)(aE(X) + b)$

$= a^2E(X^2) + 2abE(X) + b^2 - a^2(E(X))^2 - 2abE(X) - b^2$

$= a^2E(X^2) - a^2(E(X))^2$

$= a^2(E(X^2) - (E(X))^2)$

$= a^2V[X]$