STA 2210 Homework 2 (Due on Wed 5/27 by 11:59pm)

The data set BM, a .csv file, contains data on percent body fat and other various measurements of body size, for a sample of 252 men. Write your R codes, in addition to your answer, to the following problems. (Don’t forget to refer to the R reference card to find helpful commands.)

If one of the 252 men is selected at random, find the probability that

1. **they have a BODYFAT that is less than 15.0.**

bodyfat\_under15 <- subset(BM, (BODYFAT < 15.0))

81/252

= 0.3214286 = 32.1%

1. **they do not have a BODYFAT that is less than 15.0.**

1 - bodyfat\_under15 <- subset(BM, (BODYFAT < 15.0))

= 0.6785714 = 67.9%

1. **they have an AGE that is greater than 45 and a WEIGHT that is greater than 160.**

age\_over45 <- subset(BM, (AGE > 45)) = 114/252 = 0.452

weight\_over160 <- subset(BM, (WEIGHT > 160)) = 184/252 = 0.730

ageweight <- subset(BM, (AGE > 45) & (WEIGHT > 160)) = 82/252 = 0.325

P(A & B) = P(A) \* P(B) =

0.452 \* 0.730 = 0.330 = 33.0%

1. **they have a CHEST that is less than 100 or an ABDOMEN that is greater than 100.**

chest\_under100 <- subset(BM, (CHEST < 100)) = 129/252 = 0.512

abdomen\_over100 <- subset(BM, (ABDOMEN > 100)) = 56/252 = 0.222

Chest\_abdomen <- subset(BM, (CHEST < 100) & (ABDOMEN > 100)) = 0/252

chestabdomen <- subset(BM, (CHEST < 100) | (ABDOMEN > 100)) = 185/252 = .734

P(A | B) = P(A) + P(B) – P(A & B) = 0.512 + 0.222 – 0.00 = 0.734 = 73.4%

**5. they have an ANKLE greater than 23.0 given that they have a HEIGHT greater than 70.00.**

ankle\_over\_23 <- subset(BM, (ANKLE > 23.0)) = 111/252 = 0.440

height\_over\_70 <- subset(BM, (HEIGHT > 70.0)) = 125/252 = 0.496

ankleheight <- subset(BM, (ANKLE > 23.0) & (HEIGHT > 70.0) )

= 74/252 = 0.294 = 29.4%

(ANKLE > 23.0 | HEIGHT > 70.0) = (P ANKLE & P HEIGHT) / (P HEIGHT):

0.294 / 0.496 = 0.593 = 59.3%

**6. Are “having an ANKLE that is greater than 23.0” and “a HEIGHT that is greater than 70.0” independent events? Justify your answer.**

Ankle > 23 with height < 70 inches:

ankle\_height\_low <- subset(BM, (ANKLE > 23.0) & (HEIGHT < 70.0))

= 31/252 instances = .123 = 12.3%

Ankle > 23 with height > 70 inches:

ankle\_height\_high <- subset(BM, (ANKLE > 23.0) & (HEIGHT > 70.0))

= 74/252 instances = .294 = 29.4%

They are not independent events because the probability of having an ankle over 23.0 cm increases with increased height, and decreases with decreased height, as the two formulas show. Having a height less than 70.0 inches lowers the probability of having an ankle over 23.0 cm by 17.2%, compared to a height greater than 70.0 inches%. Having larger ankles in this case is inclined to be more dependent to having taller height.

**7. Are “having a CHEST that is less than 100” and “an ABDOMEN that is greater than 100” disjoint (i.e. mutually exclusive) events? Justify your answer.**

Chest\_abdomen <- subset(BM, (CHEST < 100) & (ABDOMEN > 100)) = 0/252

= 0 instances, or 0% of men having both a chest under 100 cm and an abdomen greater than 100 cm. This indicates that the two cases occurring are disjoint, as they are not possible to happen at the same time.