

Math/Stat Project Final Report

Neha Ashtekar, Sfurti Ruge : team_10_final

December 9, 2016

Habitual Alcohol Consumption and Metabolic Syndrome in Patients with Sleep Disordered Breathing

```
glimpse(OA_HAC_Data) # names, types of columns
```

```
## Observations: 683
## Variables: 15
## $ Number      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,...
## $ Sex         <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "M", ...
## $ Age         <dbl> 66.7, 58.5, 48.8, 62.2, 51.1, 64.2, 61.8, 55.2, 58...
## $ Smoking     <chr> "N", "N", "N", "Y", "N", "Y", "N", "X", "N", "N", ...
## $ HAC/w       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ Alc_freq/w  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ Alc_glasses <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ SBP         <int> 142, 95, 121, 120, 100, 131, 95, 110, 101, 107, 14...
## $ DBP         <int> 76, 79, 88, 84, 63, 81, 71, 70, 65, 77, 97, 81, 64...
## $ Waist_C     <dbl> 87, 83, 84, 90, 97, 91, 90, 78, 83, 84, 75, 84, 89...
## $ BMI         <dbl> 24.51, 22.90, 23.50, 28.00, 27.21, 27.31, 23.59, 2...
## $ MetS        <int> 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1,...
## $ AHI         <dbl> 37.3, 32.8, 5.6, 11.5, 15.2, 33.6, 44.4, 5.5, 5.3,...
## $ AHI_group   <int> 3, 3, 1, 1, 2, 3, 3, 1, 1, 2, 1, 1, 1, 2, 1, 3,...
## $ HAC_group   <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
```

Summary and Definitions of few terms in the research paper :

- 1) Investigate the relationship between HAC(Habitual Alcohol Consumption) and MetS(Metabolic Syndrome) and SDB(sleep-disordered breathing)
 - compare the prevalence of MetS and sleep quality according to amount of HAC in male patients with SDB(Sleep Disordered Breathing)
 - perform analysis to answer the below questions :
 - Does alcohol consumption increase MetS?
 - Does habitual alcohol consumption aggravate SDB and sleep quality?
- 2) Participants included 683 untreated SDB male patients.
- 3) Average (age: 54.4 +/- 7.80 y, apnea-hypopnea index (AHI): 29.0 +/- 21.53/h)
- 4) HAC was assessed as the no. of drinks consumed per week during past 12 months.
- 5) HAC groups — no drinking, light drinking < 13, heavy drinking >= 13 drinks/week
- 6) MetS = Circumference > 90cm for obesity AND 2 of the four issues (low HDL, hypertriglyceridemia, hypertension, hyperglycemia)
- 7) SDB = AHI >= 5 / h
- 8) Hypertension = SBP > 130, DBP > 85
- 9) Other terms are provided in a tabular format, at the end of the report.

Exploratory Data Analysis

Image of Graph 1 :

[Click Here - Graph 1](#)

Text of Graph 1 :

Proportion of sleep-disordered breathing (SDB) severity according to habitual alcohol consumption (HAC). The distribution of severity of SDB was significantly different according to HAC amount ($X^2 = 16.20$, $p = 0.003$). The numbers of patients with severe SDB (apnea-hypopnea index greater than or equal to 30 per h) were the highest in heavy-drinkers. On the contrary, proportion of mild SDB was the greatest in non-drinkers.

Text describing how Graph 1 was created :

1. First, we calculate the total number of observations across AHI_group column based on the HAC_group.
2. Next, we calculate the chi-square and p-value by using the same matrix.
3. Next, we calculate the proportions and create new columns in our new data frame 'sdb_hac_proportion' using the mutate function.
4. Next, we plot a barplot using the data in sdb_hac_proportion data frame. Here we make use of the various barplot parameters to replicate the barplot form the research paper. For example, displaying the text on the barplot, displaying legends correctly, setting the ylim.

Code to reproduce Graph 1 :

```
chi_matrix <- as.data.frame.matrix(table(OSA_HAC_Data$HAC_group,
  OSA_HAC_Data$AHI_group))

chi_value <- chisq.test(chi_matrix)
chi_sq_value <- chi_value$statistic
chi_sq_p_value <- chi_value$p.value
chi_value

##
## Pearson's Chi-squared test
##
## data:  chi_matrix
## X-squared = 16.198, df = 4, p-value = 0.002765

sdb_hac_proportion <- as.data.frame.matrix(table(OSA_HAC_Data$HAC_group,
  OSA_HAC_Data$AHI_group))
sdb_hac_proportion["total"] <- sdb_hac_proportion$`1` + sdb_hac_proportion$`2` +
  sdb_hac_proportion$`3`
sdb_hac_proportion <- mutate(sdb_hac_proportion, mild_sdb = (sdb_hac_proportion$`1`/sdb_hac_proportion$
  100, moderate_sdb = (sdb_hac_proportion$`2`/sdb_hac_proportion$total) *
  100, severe_sdb = (sdb_hac_proportion$`3`/sdb_hac_proportion$total) *
  100)
sdb_hac_proportion

##    1  2   3 total mild_sdb moderate_sdb severe_sdb
## 1 61 47  36   144 42.36111      32.63889   25.00000
## 2 98 96 115   309 31.71521      31.06796   37.21683
```

```
## 3 56 80 94 230 24.34783 34.78261 40.86957

sdb_bind_data <- rbind(sdb_hac_proportion$mild_sdb, sdb_hac_proportion$moderate_sdb,
  sdb_hac_proportion$severe_sdb)

sdb_hac_bp <- barplot(height = sdb_bind_data, beside = TRUE,
  main = "Proportion of sleep-disordered breathing (SDB) severity
  according to habitual alcohol consumption (HAC).",
  cex.main = 0.85, ylab = "%", ylim = c(0, 50), names.arg = c("Non-drinkers",
    "light-drinkers", "heavy-drinkers"), legend.text = c("Mild SDB",
    "Moderate SDB", "Severe SDB"), args.legend = list(x = "topright"),
  col = c("white", "lightgrey", "darkgrey"))

roundAt <- round(sdb_bind_data, 1)
text(c(4, 4.75, 6, 6.75), c(45, 45, 45, 45), c("X^2=", round(chi_sq_value,
  2), "p=", round(chi_sq_p_value, 4)))
text(sdb_hac_bp, 20, roundAt, cex = 1, pos = 3)
```

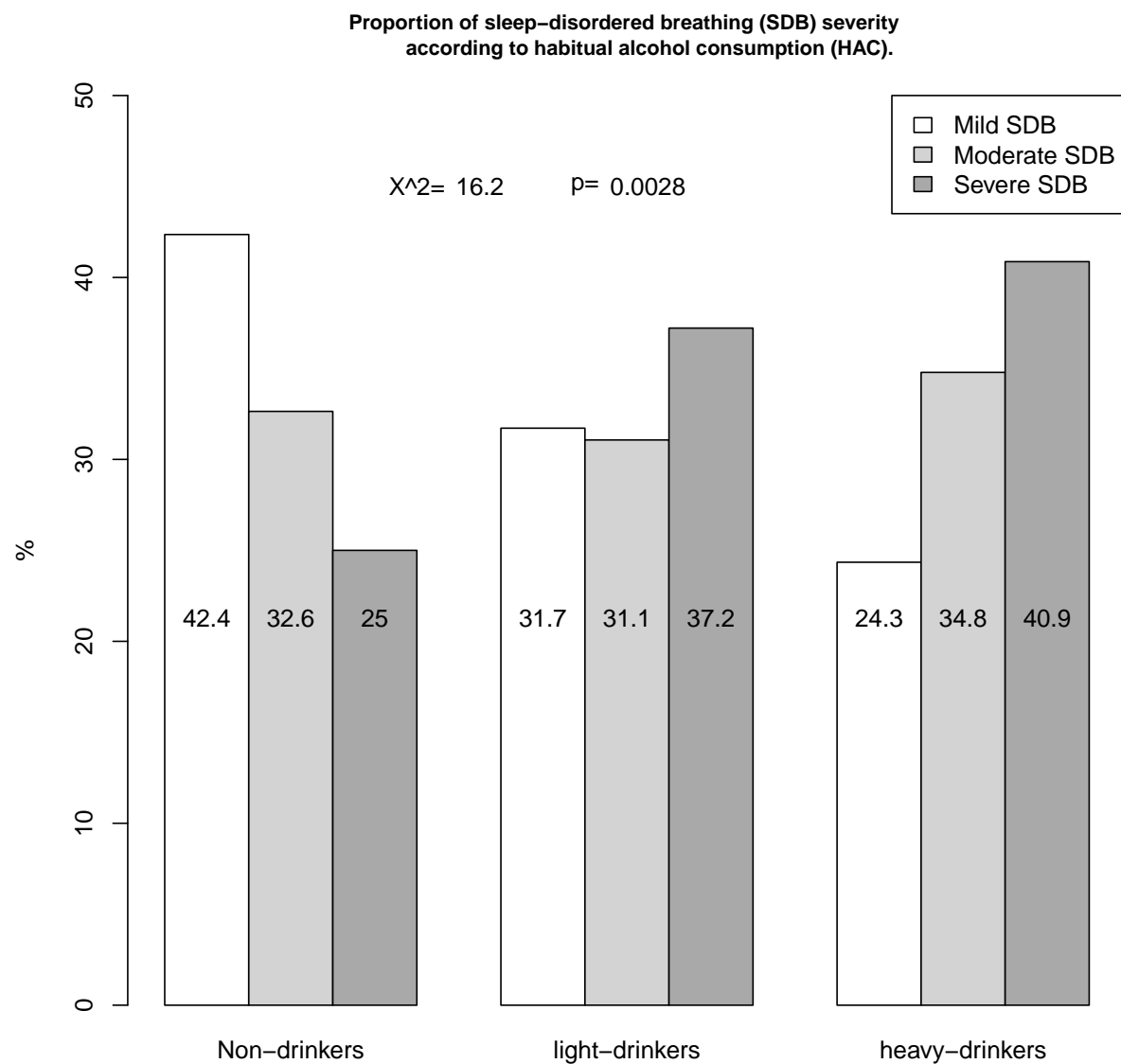


Image of Table 1 :

[Click Here - Table 1](#)

Text of Table 1 :

Characteristics of subjects with sleep disordered breathing according to habitual alcohol consumption.

Text describing how Table 1 was created :

1. First, we calculated the mean \pm sd for the Smoking parameter for all types of drinkers based on the HAC_group.

2. Similarly, we calculated the values for the rest of the table.
3. Initially, we calculated mean \pm sd values using the subset and select commands, the sd values were off by some numbers from the original sd values. However, Alison suggested us using dplyr group_by and summarize to calculate the values in a more efficient way. The sd values after the suggested changes, do match with the original values.
4. The R code chunks to create this particular table might be repetitive, so please be patient. We do understand the report becomes too huge, but we thought to include it as a part of our Final report.

Code to reproduce Table 1 :

	Non-drinkers	Light-drinkers	Heavy-drinkers	p
Age	56.48 \pm 8.19	54.16 \pm 7.95	53.41 \pm 7.11	0.001
Smoking				0.0337
Never smoking	75(52.08)	126(40.77)	84(36.52)	
Ex-smoker	43(29.86)	100(32.36)	78(33.91)	
Current smoking	26(18.05)	83(26.86)	68(29.56)	
Alcohol consumption/week	0	5.172 \pm 3.88	31.30 \pm 16.82	<0.001
Freq.Alcohol intake/week	0	1.437 \pm 1.009	3.294 \pm 1.47	<0.001
<1/month		6(1.94)	0	
1/month		18(5.82)	0	
2-4/month		217(70.22)	33(14.34)	
2-3/week		62(20.06)	131(56.95)	
>=4week		6(1.94)	66(28.69)	
Avg. intake per case	0	3.873 \pm 2.407	10.38 \pm 5.71	<0.001
<1 drink		13(4.20)	0	
1-2 drinks		96(31.06)	0	
3-4 drinks		90(29.12)	4(1.73)	
5-6deinks		15(4.85)	6(2.60)	
>=7drinks		68(29.56)	192(83.47)	
SBP,mmHg	118.6 \pm 11.2	119.4 \pm 14.38	121.3 \pm 14.67	0.184
DBP,mmHg	75.44 \pm 9.90	76.53 \pm 9.70	79.24 \pm 10.12	<0.001
Anthropometric measures				
Waist Circumference	90.55 \pm 9.09	89.95 \pm 7.18	91.7 \pm 7.92	0.038
Body mass index, kg/m ²	25.89 \pm 3.76	25.43 \pm 2.73	26.15 \pm 3.07	0.026

HAC(Habitual Alcohol Consumption) HAC_group = 0, 1 or 2, that is non-drinkers, light-drinkers and heavy drinkers

- Note : As the data for Biochemical markers and Metabolic syndrome are unavailable, we could not replicate the results in our analysis. This is the last part of Table 1. This was discussed with both the instructors.

```
# Age Non-drinkers : 56.48 +/- 8.19 Light-Drinkers : 54.16
# +/- 7.95 Heavy-Drinkers : 53.41 +/- 7.11

age_summary <- OSA_HAC_Data %>% group_by(HAC_group) %>% summarise(mean = mean(Age,
  na.rm = TRUE), sd = sd(Age, na.rm = TRUE))
age_summary

## # A tibble: 3 × 3
##   HAC_group    mean      sd
##   <int>    <dbl>    <dbl>
## 1         0 56.48125 8.186984
```

```
## 2      1 54.15825 7.949866
## 3      2 53.41391 7.111254

# Non-drinkers
total_smokers_hac_0 <- subset(OSA_HAC_Data, select = Smoking,
  subset = HAC_group == 0)
count(total_smokers_hac_0) #144

## # A tibble: 1 × 1
##       n
##   <int>
## 1   144

# Non-drinkers, Never Smoking
smoke_N_hac_0 <- subset(OSA_HAC_Data, select = Smoking, subset = (HAC_group ==
  0 & Smoking == "N"))
count(smoke_N_hac_0) #75 - displayed in Table 1

## # A tibble: 1 × 1
##       n
##   <int>
## 1    75

proportion_N_hac_0 = (count(smoke_N_hac_0)/count(total_smokers_hac_0)) *
  100
proportion_N_hac_0 # 52.08 - displayed in Table 1

##       n
## 1 52.08333

# Non-drinkers, Ex smoker
smoke_X_hac_0 <- subset(OSA_HAC_Data, select = Smoking, subset = (HAC_group ==
  0 & Smoking == "X"))
count(smoke_X_hac_0) # 43 - displayed in Table 1

## # A tibble: 1 × 1
##       n
##   <int>
## 1    43

proportion_X_hac_0 = (count(smoke_X_hac_0)/count(total_smokers_hac_0)) *
  100
proportion_X_hac_0 # 29.86 - displayed in Table 1

##       n
## 1 29.86111

# Non-drinkers, Current smoker
smoke_Y_hac_0 <- subset(OSA_HAC_Data, select = Smoking, subset = (HAC_group ==
  0 & Smoking == "Y"))
count(smoke_Y_hac_0) # 26 - displayed in Table 1

## # A tibble: 1 × 1
##       n
##   <int>
## 1    26

proportion_Y_hac_0 = (count(smoke_Y_hac_0)/count(total_smokers_hac_0)) *
  100
```

```

proportion_Y_hac_0 # 18.05 - displayed in Table 1

##           n
## 1 18.05556
# Light-drinkers
total_smokers_hac_1 <- subset(OSA_HAC_Data, select = Smoking,
                             subset = HAC_group == 1)
count(total_smokers_hac_1) #309

## # A tibble: 1 × 1
##           n
##   <int>
## 1   309
# Light-drinkers, Never smoking
smoke_N_hac_1 <- subset(OSA_HAC_Data, select = Smoking, subset = (HAC_group ==
  1 & Smoking == "N"))
count(smoke_N_hac_1) #126 - displayed in Table 1

## # A tibble: 1 × 1
##           n
##   <int>
## 1   126
proportion_N_hac_1 = (count(smoke_N_hac_1)/count(total_smokers_hac_1)) *
  100
proportion_N_hac_1 # 40.77 - displayed in Table 1

##           n
## 1 40.7767
# Light-drinkers, Ex smoker
smoke_X_hac_1 <- subset(OSA_HAC_Data, select = Smoking, subset = (HAC_group ==
  1 & Smoking == "X"))
count(smoke_X_hac_1) # 100 - displayed in Table 1

## # A tibble: 1 × 1
##           n
##   <int>
## 1   100
proportion_X_hac_1 = (count(smoke_X_hac_1)/count(total_smokers_hac_1)) *
  100
proportion_X_hac_1 # 32.36 - displayed in Table 1

##           n
## 1 32.36246
# Light-drinkers, Current smoker
smoke_Y_hac_1 <- subset(OSA_HAC_Data, select = Smoking, subset = (HAC_group ==
  1 & Smoking == "Y"))
count(smoke_Y_hac_1) # 83 - displayed in Table 1

## # A tibble: 1 × 1
##           n
##   <int>
## 1    83

```

```

proportion_Y_hac_1 = (count(smoke_Y_hac_1)/count(total_smokers_hac_1)) *
  100
proportion_Y_hac_1 # 26.86 - displayed in Table 1

##           n
## 1 26.86084

# Heavy-drinkers
total_smokers_hac_2 <- subset(OSA_HAC_Data, select = Smoking,
  subset = HAC_group == 2)
count(total_smokers_hac_2) #230

## # A tibble: 1 × 1
##           n
##   <int>
## 1    230

# Heavy-drinkers, Never smoking
smoke_N_hac_2 <- subset(OSA_HAC_Data, select = Smoking, subset = (HAC_group ==
  2 & Smoking == "N"))
count(smoke_N_hac_2) #84 - displayed in Table 1

## # A tibble: 1 × 1
##           n
##   <int>
## 1     84

proportion_N_hac_2 = (count(smoke_N_hac_2)/count(total_smokers_hac_2)) *
  100
proportion_N_hac_2 # 36.52 - displayed in Table 1

##           n
## 1 36.52174

# Heavy-drinkers, Ex smoker
smoke_X_hac_2 <- subset(OSA_HAC_Data, select = Smoking, subset = (HAC_group ==
  2 & Smoking == "X"))
count(smoke_X_hac_2) # 78 - displayed in Table 1

## # A tibble: 1 × 1
##           n
##   <int>
## 1     78

proportion_X_hac_2 = (count(smoke_X_hac_2)/count(total_smokers_hac_2)) *
  100
proportion_X_hac_2 # 33.91 - displayed in Table 1

##           n
## 1 33.91304

# Heavy-drinkers, Current smoker
smoke_Y_hac_2 <- subset(OSA_HAC_Data, select = Smoking, subset = (HAC_group ==
  2 & Smoking == "Y"))
count(smoke_Y_hac_2) # 68 - displayed in Table 1

## # A tibble: 1 × 1
##           n

```



```
##      <int>
## 1      68

proportion_Y_hac_2 = (count(smoke_Y_hac_2)/count(total_smokers_hac_2)) *
  100
proportion_Y_hac_2  # 29.56 - displayed in Table 1

##              n
## 1 29.56522

# alcohol consumption/week Non-drinkers : 0 Light-Drinkers :
# 5.172 +/- 3.88 Heavy-Drinkers : 31.29 +/- 16.82

alcohol_cons_summary <- OSA_HAC_Data %>% group_by(HAC_group) %>%
  summarise(mean = mean(`HAC/w`, na.rm = TRUE), sd = sd(`HAC/w`,
    na.rm = TRUE))
alcohol_cons_summary

## # A tibble: 3 × 3
##   HAC_group    mean      sd
##   <int>    <dbl>    <dbl>
## 1         0  0.000000  0.000000
## 2         1  5.172233  3.881157
## 3         2 31.295913 16.820004

# Frequency of alcohol intake/week Non-drinkers : 0
# Light-Drinkers : 1.437 +/- 1.009 Heavy-Drinkers : 3.294 +/-
# 1.47

alcohol_intake_summary <- OSA_HAC_Data %>% group_by(HAC_group) %>%
  summarise(mean = mean(`Alc_freq/w`, na.rm = TRUE), sd = sd(`Alc_freq/w`,
    na.rm = TRUE))
alcohol_intake_summary

## # A tibble: 3 × 3
##   HAC_group    mean      sd
##   <int>    <dbl>    <dbl>
## 1         0  0.000000  0.000000
## 2         1  1.436602  1.009077
## 3         2  3.294739  1.472801

total_hac_1 <- subset(OSA_HAC_Data, select = `Alc_freq/w`, subset = HAC_group ==
  1)
count(total_hac_1)  #309

## # A tibble: 1 × 1
##       n
##   <int>
## 1   309

total_hac_2 <- subset(OSA_HAC_Data, select = `Alc_freq/w`, subset = HAC_group ==
  2)
count(total_hac_2)  #309

## # A tibble: 1 × 1
##       n
##   <int>
## 1   230
```

```

# The range for Alc_freq/w in below code chunks is estimated
# by us as they were not provided in the research paper.
# Here, we consider 'a week' as (1/4th) part of a 30-day
# month. So, in few cases our values differ from the original
# values.

# <1/month-----
less_1_month <- subset(OSA_HAC_Data, select = `Alc_freq/w`, `Alc_freq/w` >
  0 & `Alc_freq/w` < 0.25 & HAC_group == 1)
count(less_1_month) #6

```

```

## # A tibble: 1 × 1
##       n
##   <int>
## 1     6

proportion_less_1_month_hac_1 = (count(less_1_month)/count(total_hac_1)) *
  100
proportion_less_1_month_hac_1 # 1.941748

```

```

##       n
## 1 1.941748

```

```

# 1/month-----
one_per_month <- subset(OSA_HAC_Data, select = `Alc_freq/w`,
  `Alc_freq/w` >= 0.25 & `Alc_freq/w` < 0.49 & HAC_group ==
  1)
count(one_per_month) #18

```

```

## # A tibble: 1 × 1
##       n
##   <int>
## 1    18

proportion_one_per_month_hac_1 = (count(one_per_month)/count(total_hac_1)) *
  100
proportion_one_per_month_hac_1 # 5.825243

```

```

##       n
## 1 5.825243

```

```

# 2-4/month-----
# hac=1
two_four_month1 <- subset(OSA_HAC_Data, select = `Alc_freq/w`,
  `Alc_freq/w` >= 0.5 & `Alc_freq/w` < 2 & HAC_group == 1)
count(two_four_month1) #217

```

```

## # A tibble: 1 × 1
##       n
##   <int>
## 1   217

proportion_two_four_per_month_hac_1 = (count(two_four_month1)/count(total_hac_1)) *
  100
proportion_two_four_per_month_hac_1 #70.22654

```

```

##       n

```

```
## 1 70.22654
# hac=2
two_four_month2 <- subset(OSA_HAC_Data, select = `Alc_freq/w`,
  `Alc_freq/w` >= 0.5 & `Alc_freq/w` < 2 & HAC_group == 2) #33
count(two_four_month2)

## # A tibble: 1 × 1
##       n
##   <int>
## 1     33

proportion_two_four_per_month_hac_2 = (count(two_four_month2)/count(total_hac_2)) *
  100
proportion_two_four_per_month_hac_2 #14.34783

##       n
## 1 14.34783

# 2-3/week----- hac =
# 1
two_three_week1 <- subset(OSA_HAC_Data, select = `Alc_freq/w`,
  `Alc_freq/w` >= 2 & `Alc_freq/w` < 4 & HAC_group == 1)
count(two_three_week1) #62

## # A tibble: 1 × 1
##       n
##   <int>
## 1     62

proportion_two_three_week1_hac_1 = (count(two_three_week1)/count(total_hac_1)) *
  100
proportion_two_three_week1_hac_1 #20.06472

##       n
## 1 20.06472

# hac = 2
two_three_week2 <- subset(OSA_HAC_Data, select = `Alc_freq/w`,
  `Alc_freq/w` >= 2 & `Alc_freq/w` < 4 & HAC_group == 2)
count(two_three_week2) #131

## # A tibble: 1 × 1
##       n
##   <int>
## 1    131

proportion_two_three_week1_hac_2 = (count(two_three_week2)/count(total_hac_2)) *
  100
proportion_two_three_week1_hac_2 #56.95652

##       n
## 1 56.95652

# >=4/week-----
# hac = 1
two_three_week1 <- subset(OSA_HAC_Data, select = `Alc_freq/w`,
  `Alc_freq/w` >= 4 & HAC_group == 1)
count(two_three_week1) #6
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1     6

proportion_two_three_week1_hac_1 = (count(two_three_week1)/count(total_hac_1)) *
  100
proportion_two_three_week1_hac_1  #1.941748
```

```
##       n
## 1 1.941748
```

```
# hac = 2
two_three_week2 <- subset(OSA_HAC_Data, select = `Alc_freq/w`,
  `Alc_freq/w` >= 4 & HAC_group == 2)
count(two_three_week2)  #66
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    66

proportion_two_three_week1_hac_2 = (count(two_three_week2)/count(total_hac_2)) *
  100
proportion_two_three_week1_hac_2  #28.69565
```

```
##       n
## 1 28.69565
```

```
# average intake per case Non-drinkers : 0 Light-Drinkers :
# 3.873528 +/- 2.407464 Heavy-Drinkers : 10.377348 +/-
# 5.716741
```

```
avg_intake_per_case <- OSA_HAC_Data %>% group_by(HAC_group) %>%
  summarise(mean = mean(`Alc glasses`, na.rm = TRUE), sd = sd(`Alc glasses`,
    na.rm = TRUE))
avg_intake_per_case
```

```
## # A tibble: 3 × 3
##   HAC_group    mean    sd
##   <int>    <dbl>  <dbl>
## 1      0  0.000000  0.000000
## 2      1  3.873528  2.407464
## 3      2 10.377348  5.716741
```

```
# The range for 'Alc glasses' in below code chunks is
# estimated by us as they were not provided in the research
# paper. So, in few cases our values differ from the
# original values.
```

```
#-----<1 drink for hac_group = 1
count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` <
  1 & HAC_group == 1))  #13
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    13
```

```

proportion_drink_less_hac_1 = (count(subset(OSA_HAC_Data, select = `Alc glasses`,
  `Alc glasses` < 1 & HAC_group == 1))/count(total_hac_1)) *
  100
proportion_drink_less_hac_1  #4.20712

```

```

##          n
## 1 4.20712

```

```

#-----1-2 drink for hac_group = 1
count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  1 & `Alc glasses` < 3 & HAC_group == 1)) # 96

```

```

## # A tibble: 1 × 1
##       n
##   <int>
## 1     96

```

```

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  1 & `Alc glasses` < 3 & HAC_group == 1))/count(total_hac_1) *
  100

```

```

##          n
## 1 31.06796

```

```

# 31.06796

```

```

#-----3-4 drink for hac_group = 1 and 2
count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  3 & `Alc glasses` <= 4 & HAC_group == 1)) #90

```

```

## # A tibble: 1 × 1
##       n
##   <int>
## 1     90

```

```

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  3 & `Alc glasses` <= 4 & HAC_group == 1))/count(total_hac_1) *
  100 #29.12621

```

```

##          n
## 1 29.12621

```

```

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  3 & `Alc glasses` <= 4 & HAC_group == 2)) #4

```

```

## # A tibble: 1 × 1
##       n
##   <int>
## 1      4

```

```

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  3 & `Alc glasses` <= 4 & HAC_group == 2))/count(total_hac_2) *
  100 #1.73913

```

```

##          n
## 1 1.73913

```

```

#-----5-6 drink for hac_group = 1 and 2
count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=

```

```

5 & `Alc glasses` <= 6 & HAC_group == 1)) #15

## # A tibble: 1 × 1
##       n
##   <int>
## 1    15

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  5 & `Alc glasses` <= 6 & HAC_group == 1))/count(total_hac_1) *
  100 #4.854369

##       n
## 1 4.854369

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  5 & `Alc glasses` <= 6 & HAC_group == 2)) #6

## # A tibble: 1 × 1
##       n
##   <int>
## 1     6

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  5 & `Alc glasses` <= 6 & HAC_group == 2))/count(total_hac_2) *
  100 #2.608696

##       n
## 1 2.608696

#----->=7 drink for hac_group = 1 and 2
count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  7 & HAC_group == 1)) #68

## # A tibble: 1 × 1
##       n
##   <int>
## 1    68

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  7 & HAC_group == 1))/count(total_hac_2) * 100 #29.56522

##       n
## 1 29.56522

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  7 & HAC_group == 2)) #192

## # A tibble: 1 × 1
##       n
##   <int>
## 1   192

count(subset(OSA_HAC_Data, select = `Alc glasses`, `Alc glasses` >=
  7 & HAC_group == 2))/count(total_hac_2) * 100 #83.47826

##       n
## 1 83.47826

```

```

# SBP, mmHg Non-drinkers : 118.6 +/- 15.53 Light-Drinkers :
# 119.4 +/- 14.38 Heavy-Drinkers : 121.3 +/- 14.67

sbp_summary <- OSA_HAC_Data %>% group_by(HAC_group) %>% summarise(mean = mean(SBP,
  na.rm = TRUE), sd = sd(SBP, na.rm = TRUE))
sbp_summary

## # A tibble: 3 × 3
##   HAC_group    mean      sd
##   <int>    <dbl>    <dbl>
## 1         0 118.6111 15.53346
## 2         1 119.4337 14.38568
## 3         2 121.2652 14.67030

# DBP, mmHg Non-drinkers : 75.44 +/- 9.90 Light-Drinkers :
# 76.53 +/- 9.70 Heavy-Drinkers : 79.24 +/- 10.12

dbp_summary <- OSA_HAC_Data %>% group_by(HAC_group) %>% summarise(mean = mean(DBP,
  na.rm = TRUE), sd = sd(DBP, na.rm = TRUE))
dbp_summary

## # A tibble: 3 × 3
##   HAC_group    mean      sd
##   <int>    <dbl>    <dbl>
## 1         0 75.44444  9.904282
## 2         1 76.52751  9.694190
## 3         2 79.24348 10.124199

# Anthropometric measures waist circumference, cm
# non-drinkers : 90.55 +/- 9.09 light-drinkers : 89.95 +/-
# 7.18 heavy-drinkers : 91.7 +/- 7.92

waist_hac_summary <- OSA_HAC_Data %>% group_by(HAC_group) %>%
  summarise(mean = mean(Waist_C, na.rm = TRUE), sd = sd(Waist_C,
    na.rm = TRUE))
waist_hac_summary

## # A tibble: 3 × 3
##   HAC_group    mean      sd
##   <int>    <dbl>    <dbl>
## 1         0 90.55208  9.086361
## 2         1 89.94822  7.182268
## 3         2 91.69783  7.916389

# BMI, kg/m2 non-drinkers : 25.89 +/- 3.76 light-drinkers :
# 25.43 +/- 2.73 heavy-drinkers : 26.15 +/- 3.07

bmi_hac_summary <- OSA_HAC_Data %>% group_by(HAC_group) %>% summarise(mean = mean(BMI,
  na.rm = TRUE), sd = sd(BMI, na.rm = TRUE))
bmi_hac_summary

## # A tibble: 3 × 3
##   HAC_group    mean      sd
##   <int>    <dbl>    <dbl>
## 1         0 25.88931  3.761520
## 2         1 25.43353  2.728021

```

```
## 3          2 26.14626 3.070079
```

Calculating the last column, that is the p-value in Table 1

- We did not cover this part during our EDA analysis report, and so we calculate the last column separately for each case in our Final report.
- As suggested in the Statistical Analyses part of the paper :
- The alpha level for statistical significance was set at $p < 0.05$. Descriptive statistics of mean \pm SD and number (percentage) are summarized in Table 1. Normally distributed continuous data were analyzed using one-way analysis of variance (ANOVA). Categorical variables were analyzed using the Chi-square test.
- [Click Here - Statistical Analyses](#)

```
# Age*, years , p = 0.001 (expected)
age_lm_data <- subset(OSA_HAC_Data, select = c(Age, HAC_group))
age_lm_data$HAC_group <- factor(age_lm_data$HAC_group)
age_lm_out <- lm(Age ~ HAC_group, data = age_lm_data)
anova(age_lm_out) # 0.0007724 ~ 0.001

## Analysis of Variance Table
##
## Response: Age
##           Df Sum Sq Mean Sq F value    Pr(>F)
## HAC_group   2     865   432.72   7.2421 0.0007724 ***
## Residuals 680  40631    59.75
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Never-Smoking, p = 0.034 (expected)
chi_smoking_table <- subset(OSA_HAC_Data, select = c(Smoking,
  HAC_group))
chi_smoking_matrix <- as.data.frame.matrix(table(chi_smoking_table$Smoking,
  chi_smoking_table$HAC_group))
chisq.test(chi_smoking_matrix) # 0.03374

##
## Pearson's Chi-squared test
##
## data:  chi_smoking_matrix
## X-squared = 10.432, df = 4, p-value = 0.03374

# Alcohol consumption/week, p = <0.001 (expected)
alc_per_week_lm_data <- subset(OSA_HAC_Data, select = c(`HAC/w`,
  HAC_group))
alc_per_week_lm_data$HAC_group <- factor(alc_per_week_lm_data$HAC_group)
alc_per_week_lm_out <- lm(`HAC/w` ~ HAC_group, data = alc_per_week_lm_data)
anova(alc_per_week_lm_out) # < 2.2e-16

## Analysis of Variance Table
##
## Response: HAC/w
##           Df Sum Sq Mean Sq F value    Pr(>F)
## HAC_group   2 120250    60125  588.9 < 2.2e-16 ***
## Residuals 680  69426     102
```



```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Frequency of alcohol intake per week, p = <0.001 (expected)
alc_freq_table <- subset(OSA_HAC_Data, select = c(`Alc_freq/w`,
  HAC_group))
alc_freq_table <- alc_freq_table %>% mutate(Freq = `Alc_freq/w`)
alc_freq_table$FreqCat <- cut(alc_freq_table$Freq, c(-1, 0.24,
  0.49, 1.99, 3.99, 7))
chi_alc_freq_matrix <- as.data.frame.matrix(table(alc_freq_table$FreqCat,
  alc_freq_table$HAC_group))
chisq.test(chi_alc_freq_matrix) # < 2.2e-16

##
## Pearson's Chi-squared test
##
## data:  chi_alc_freq_matrix
## X-squared = 931.1, df = 8, p-value < 2.2e-16

# Avg intake case per week, <1/drink , p = <0.001 (expected)
alc_intake_table <- subset(OSA_HAC_Data, select = c(`Alc glasses`,
  HAC_group))
alc_intake_table <- alc_intake_table %>% mutate(Freq = `Alc glasses`)
alc_intake_table$FreqCat <- cut(alc_intake_table$Freq, c(-1,
  0.99, 2.99, 4.99, 6, 100))
chi_alc_intake_matrix <- as.data.frame.matrix(table(alc_intake_table$FreqCat,
  alc_intake_table$HAC_group))
chisq.test(chi_alc_intake_matrix) # < 2.2e-16

##
## Pearson's Chi-squared test
##
## data:  chi_alc_intake_matrix
## X-squared = 901.69, df = 8, p-value < 2.2e-16

# SBP mmHg, p = 0.185 (expected)
sbp_dbp_lm_data <- subset(OSA_HAC_Data, select = c(SBP, DBP,
  HAC_group))
sbp_dbp_lm_data$HAC_group <- factor(sbp_dbp_lm_data$HAC_group)

sbp_anova_data <- lm(SBP ~ HAC_group, data = sbp_dbp_lm_data)
anova(sbp_anova_data) # 0.1847

## Analysis of Variance Table
##
## Response: SBP
##           Df Sum Sq Mean Sq F value Pr(>F)
## HAC_group   2    735   367.37   1.6933 0.1847
## Residuals 680 147529   216.95

# DBP mmHg, p < 0.001 (expected)
dbp_anova_data <- lm(DBP ~ HAC_group, data = sbp_dbp_lm_data)
anova(dbp_anova_data) # 0.0004058

## Analysis of Variance Table
##
## Response: DBP
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## HAC_group  2    1544   771.93   7.8999 0.0004058 ***
## Residuals 680   66445    97.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Anthropometric measures
waist_bmi_lm_data <- subset(OSA_HAC_Data, select = c(Waist_C,
  BMI, HAC_group))
waist_bmi_lm_data$HAC_group <- factor(waist_bmi_lm_data$HAC_group)

# waist circumference, cm, p = 0.038 (expected)
waist_anova_data <- lm(Waist_C ~ HAC_group, data = waist_bmi_lm_data)
anova(waist_anova_data) # 0.03813

## Analysis of Variance Table
##
## Response: Waist_C
##           Df Sum Sq Mean Sq F value    Pr(>F)
## HAC_group  2     406   202.970   3.2826 0.03813 *
## Residuals 680   42046    61.832
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Body Mass Index, p = 0.026
bmi_anova_data <- lm(BMI ~ HAC_group, data = waist_bmi_lm_data)
anova(bmi_anova_data) # 0.02637

## Analysis of Variance Table
##
## Response: BMI
##           Df Sum Sq Mean Sq F value    Pr(>F)
## HAC_group  2     69.6    34.797    3.655 0.02637 *
## Residuals 680  6473.9     9.520
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Final report : Replication of analyses/results in the paper that are based on the general linear model (t-tests, linear regression, Analysis of Variance)

Odds ratio (95% Confidence Intervals) for metabolic syndrome according to habitual alcohol consumption.

```
# create a new data frame by selecting the required columns
# MetS, HAC_group, Age, AHI and BMI. make sure the columns
# are factor variables before we perform glm
data_table3 <- subset(OSA_HAC_Data, select = c(MetS, HAC_group,
  Age, AHI, AHI_group, BMI))
data_table3$MetS <- factor(data_table3$MetS)
data_table3$HAC_group <- factor(data_table3$HAC_group)
glimpse(data_table3)

## Observations: 683
```

```

## Variables: 6
## $ MetS      <fctr> 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, ...
## $ HAC_group <fctr> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ Age       <dbl> 66.7, 58.5, 48.8, 62.2, 51.1, 64.2, 61.8, 55.2, 58.0...
## $ AHI       <dbl> 37.3, 32.8, 5.6, 11.5, 15.2, 33.6, 44.4, 5.5, 5.3, 1...
## $ AHI_group <int> 3, 3, 1, 1, 2, 3, 3, 1, 1, 2, 1, 1, 1, 2, 1, 3, 2, 2...
## $ BMI       <dbl> 24.51, 22.90, 23.50, 28.00, 27.21, 27.31, 23.59, 24....

## *****Start glm
## (Unadjusted)***** Fit a logistic
## regression model to the data_table3 data set, using
## HAC_group as predictor.
metS_glm_out <- glm(MetS ~ HAC_group, data = data_table3, family = "binomial")
summary(metS_glm_out)

##
## Call:
## glm(formula = MetS ~ HAC_group, family = "binomial", data = data_table3)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0826  -0.9236  -0.8843   1.2752   1.5021
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.6313     0.1750  -3.606  0.00031 ***
## HAC_group1   -0.1059     0.2131  -0.497  0.61929
## HAC_group2    0.4042     0.2197   1.840  0.06575 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 899.37  on 682  degrees of freedom
## Residual deviance: 890.94  on 680  degrees of freedom
## AIC: 896.94
##
## Number of Fisher Scoring iterations: 4

# broom function to breakdown the output.
tidy(metS_glm_out)

##      term      estimate std.error statistic    p.value
## 1 (Intercept) -0.6312718 0.1750380 -3.6064848 0.0003103731
## 2 HAC_group1  -0.1058923 0.2131267 -0.4968514 0.6192938117
## 3 HAC_group2   0.4042143 0.2196696  1.8401017 0.0657533032

## *****Start Odds ratio
## (Unadjusted)***** odds ratios
## only
round(exp(coef(metS_glm_out)), 2)

## (Intercept) HAC_group1 HAC_group2
##          0.53         0.90         1.50

## odds ratios and 95% CI
round(exp(cbind(OR = coef(metS_glm_out), confint(metS_glm_out))),

```

2)

```
##               OR 2.5 % 97.5 %
## (Intercept) 0.53  0.37   0.75
## HAC_group1  0.90  0.59   1.37
## HAC_group2  1.50  0.98   2.31

# broom one line output
tidy(metS_glm_out, exponentiate = TRUE, conf.int = TRUE)

##           term estimate std.error statistic    p.value  conf.low
## 1 (Intercept) 0.5319149 0.1750380 -3.6064848 0.0003103731 0.3748494
## 2 HAC_group1  0.8995215 0.2131267 -0.4968514 0.6192938117 0.5937715
## 3 HAC_group2  1.4981250 0.2196696  1.8401017 0.0657533032 0.9767271
##      conf.high
## 1 0.7457593
## 2 1.3708652
## 3 2.3133629

# So, we can say that : The odds of being a light
# drinker(HAC_group1) and having metabolic syndrome is 0.90
# The odds of being a heavy drinker(HAC_group2) and having
# metabolic syndrome is 1.50 Unadjusted : The results
# revealed that heavy drinkers were more likely to have MetS
# than non-drinkers [odds ratio (OR) = 1.50 (0.98 - 2.31)]
# *****End Odds ratio
# (Unadjusted)*****

## *****Start glm (Adjusted for
## Age)***** Fit a logistic
## regression model to the data_table3 data set, using
## HAC_group + Age as predictors.
metS_age_glm_out <- glm(MetS ~ HAC_group + Age, data = data_table3,
  family = "binomial")
summary(metS_age_glm_out)

##
## Call:
## glm(formula = MetS ~ HAC_group + Age, family = "binomial", data = data_table3)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1179  -0.9238  -0.8825   1.2847   1.5436
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.86582    0.61337  -1.412  0.1581
## HAC_group1  -0.09627    0.21451  -0.449  0.6536
## HAC_group2   0.41707    0.22209   1.878  0.0604 .
## Age          0.00415    0.01039   0.399  0.6897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 899.37 on 682 degrees of freedom
## Residual deviance: 890.78 on 679 degrees of freedom
## AIC: 898.78
##
## Number of Fisher Scoring iterations: 4
## *****Start Odds ratio (Adjusted for
## Age)***** odds ratios only
round(exp(coef(metS_age_glm_out)), 2)

## (Intercept) HAC_group1 HAC_group2 Age
## 0.42 0.91 1.52 1.00
## odds ratios and 95% CI
round(exp(cbind(OR = coef(metS_age_glm_out), confint(metS_age_glm_out))),
2)

## OR 2.5 % 97.5 %
## (Intercept) 0.42 0.13 1.39
## HAC_group1 0.91 0.60 1.39
## HAC_group2 1.52 0.98 2.35
## Age 1.00 0.98 1.02
# So, we can say that : The odds of being a light
# drinker(HAC_group1) and having metabolic syndrome is 0.91
# The odds of being a heavy drinker(HAC_group2) and having
# metabolic syndrome is 1.52 Adjusted(Age) : The results
# revealed that heavy drinkers were more likely to have MetS
# than non-drinkers [odds ratio (OR) = 1.52 (0.98 - 2.35)]
# *****End Odds ratio (Adjusted for
# Age)*****

## *****Start glm (Adjusted for Age and
## AHI)***** Fit a logistic regression
## model to the data_table3 data set, using HAC_group + Age +
## AHI as predictors.
metS_age_ahi_glm_out <- glm(MetS ~ HAC_group + Age + AHI, data = data_table3,
family = "binomial")
summary(metS_age_ahi_glm_out)

##
## Call:
## glm(formula = MetS ~ HAC_group + Age + AHI, family = "binomial",
## data = data_table3)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.0558 -0.9454 -0.8018 1.3045 1.7008
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.521891 0.642585 -2.368 0.0179 *
## HAC_group1 -0.139337 0.218645 -0.637 0.5239
## HAC_group2 0.321021 0.227116 1.413 0.1575
```

```
## Age          0.007840    0.010611    0.739    0.4600
## AHI          0.017075    0.003836    4.451 8.54e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 899.37 on 682 degrees of freedom
## Residual deviance: 870.06 on 678 degrees of freedom
## AIC: 880.06
##
## Number of Fisher Scoring iterations: 4
## *****Start Odds ratio (Adjusted for Age
## and AHI)***** odds ratios only
round(exp(coef(metS_age_ahi_glm_out)), 2)

## (Intercept) HAC_group1 HAC_group2      Age      AHI
##          0.22          0.87          1.38          1.01          1.02

## odds ratios and 95% CI
round(exp(cbind(OR = coef(metS_age_ahi_glm_out), confint(metS_age_ahi_glm_out))),
      2)

##          OR 2.5 % 97.5 %
## (Intercept) 0.22 0.06 0.76
## HAC_group1 0.87 0.57 1.34
## HAC_group2 1.38 0.89 2.16
## Age        1.01 0.99 1.03
## AHI        1.02 1.01 1.03

# So, we can say that : The odds of being a light
# drinker(HAC_group1) and having metabolic syndrome is 0.87
# The odds of being a heavy drinker(HAC_group2) and having
# metabolic syndrome is 1.38 Adjusted(Age) : The results
# revealed that heavy drinkers were more likely to have MetS
# than non-drinkers [odds ratio (OR) = 1.38 (0.89 - 2.16)]
# *****End Odds ratio (Adjusted for Age and
# AHI)*****

## *****Start glm (Adjusted for Age, AHI and
## BMI)***** Fit a logistic regression model
## to the data_table3 data set, using HAC_group + Age + AHI +
## BMI as predictors.
metS_age_ahi_bmi_glm_out <- glm(MetS ~ HAC_group + Age + AHI +
  BMI, data = data_table3, family = "binomial")
summary(metS_age_ahi_bmi_glm_out)

##
## Call:
## glm(formula = MetS ~ HAC_group + Age + AHI + BMI, family = "binomial",
##      data = data_table3)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -3.0413 -0.8138 -0.4390 0.8578 2.3485
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -16.770928  1.585779 -10.576 < 2e-16 ***
## HAC_group1   0.123173  0.258740  0.476 0.63404
## HAC_group2   0.532990  0.267094  1.996 0.04599 *
## Age          0.036743  0.012684  2.897 0.00377 **
## AHI          -0.000698  0.004662 -0.150 0.88100
## BMI          0.539075  0.049487 10.893 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 899.37  on 682  degrees of freedom
## Residual deviance: 690.18  on 677  degrees of freedom
## AIC: 702.18
##
## Number of Fisher Scoring iterations: 5
## *****Start Odds ratio (Adjusted for Age,
## AHI and BMI)***** odds ratios only
round(exp(coef(metS_age_ahi_bmi_glm_out)), 2)

## (Intercept)  HAC_group1  HAC_group2      Age      AHI      BMI
##           0.00         1.13         1.70      1.04      1.00      1.71

## odds ratios and 95% CI
round(exp(cbind(OR = coef(metS_age_ahi_bmi_glm_out), confint(metS_age_ahi_bmi_glm_out))),
      2)

##           OR 2.5 % 97.5 %
## (Intercept) 0.00 0.00 0.00
## HAC_group1  1.13 0.68 1.89
## HAC_group2  1.70 1.01 2.89
## Age         1.04 1.01 1.06
## AHI         1.00 0.99 1.01
## BMI         1.71 1.56 1.90

# So, we can say that : The odds of being a light
# drinker(HAC_group1) and having metabolic syndrome is 1.13
# The odds of being a heavy drinker(HAC_group2) and having
# metabolic syndrome is 1.71 Adjusted(Age) : The results
# revealed that heavy drinkers were more likely to have MetS
# than non-drinkers [odds ratio (OR) = 1.71 (1.01 - 2.89)]

## *****End Odds ratio (Adjusted for Age, AHI
## and BMI)*****
```

Table 3. Odds ratio (95% Confidence Intervals) for metabolic syndrome according to habitual alcohol consumption.

	Unadjusted	Adjusted for Age	Adjusted for Age and AHI	Adjusted for Age, AHI and BMI
Non-drinkers	1.0(reference)	1.0(reference)	1.0(reference)	1.0(reference)
Light-drinkers	0.90(0.59 - 1.37)	0.91(0.60 - 1.39)	0.87(0.57 - 1.34)	1.13(0.68 - 1.89)
Heavy-drinkers	1.50(0.97 - 2.31)	1.52(0.98 - 2.35)	1.38(0.89 - 2.16)	1.70(1.01 - 2.89)
AHI: apnea-hypopnea index	BMI : body-mass index			

Odds ratio (95% Confidence Intervals) for metabolic syndrome according to habitual alcohol consumption.

```
#-----
# create a new data frame by selecting the required columns
# AHI, MetS, HAC_group, Age, AHI_group and BMI. make sure the
# columns are factor variables before we perform glm
data_table4 <- subset(OSA_HAC_Data, select = c(MetS, HAC_group,
  Age, AHI, AHI_group, BMI))
data_table4$HAC_group <- factor(data_table4$HAC_group)
glimpse(data_table4)

## Observations: 683
## Variables: 6
## $ MetS      <int> 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0...
## $ HAC_group <fctr> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ Age       <dbl> 66.7, 58.5, 48.8, 62.2, 51.1, 64.2, 61.8, 55.2, 58.0...
## $ AHI       <dbl> 37.3, 32.8, 5.6, 11.5, 15.2, 33.6, 44.4, 5.5, 5.3, 1...
## $ AHI_group <int> 3, 3, 1, 1, 2, 3, 3, 1, 1, 2, 1, 1, 1, 2, 1, 3, 2, 2...
## $ BMI       <dbl> 24.51, 22.90, 23.50, 28.00, 27.21, 27.31, 23.59, 24....

# Fit a logistic regression model to the data_table4 data
# set, using HAC_group as predictor.

data_table4 <- data_table4 %>% mutate(SevereAHI = ifelse(AHI >
  30, 1, 0))
data_table4$SevereAHI <- factor(data_table4$SevereAHI)

## *****Start glm
## (Unadjusted)***** Fit a logistic
## regression model to the data_table4 data set, using
## HAC_group as predictor.
AHI_glm_out <- glm(SevereAHI ~ HAC_group, data = data_table4,
  family = "binomial")
summary(AHI_glm_out)

##
```



```
## Call:
## glm(formula = SevereAHI ~ HAC_group, family = "binomial", data = data_table4)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0251  -0.9649  -0.7585   1.3377   1.6651
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0986     0.1924  -5.709 1.14e-08 ***
## HAC_group1     0.5757     0.2256   2.552 0.01071 *
## HAC_group2     0.7293     0.2346   3.109 0.00188 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 891.55  on 682  degrees of freedom
## Residual deviance: 881.03  on 680  degrees of freedom
## AIC: 887.03
##
## Number of Fisher Scoring iterations: 4
```

```
# broom function to breakdown the output.
tidy(AHI_glm_out)
```

```
##      term      estimate std.error statistic      p.value
## 1 (Intercept) -1.0986123 0.1924499 -5.708562 1.139348e-08
## 2 HAC_group1   0.5756863 0.2255821  2.552004 1.071054e-02
## 3 HAC_group2   0.7292522 0.2345809  3.108745 1.878841e-03
```

```
## *****Start Odds ratio
## (Unadjusted)***** odds ratios
## only
round(exp(coef(AHI_glm_out)), 2)
```

```
## (Intercept) HAC_group1 HAC_group2
##      0.33      1.78      2.07
```

```
## odds ratios and 95% CI
round(exp(cbind(OR = coef(AHI_glm_out), confint(AHI_glm_out))),
      2)
```

```
##      OR 2.5 % 97.5 %
## (Intercept) 0.33 0.23 0.48
## HAC_group1  1.78 1.15 2.79
## HAC_group2  2.07 1.32 3.31
```

```
# broom one line output
tidy(AHI_glm_out, exponentiate = TRUE, conf.int = TRUE)
```

```
##      term      estimate std.error statistic      p.value conf.low
## 1 (Intercept) 0.3333333 0.1924499 -5.708562 1.139348e-08 0.225614
## 2 HAC_group1  1.7783505 0.2255821  2.552004 1.071054e-02 1.151026
## 3 HAC_group2  2.0735294 0.2345809  3.108745 1.878841e-03 1.317713
##      conf.high
## 1 0.4808921
```

```
## 2 2.7923054
## 3 3.3111796

# So, we can say that : The odds of being a light
# drinker(HAC_group1) and having metabolic syndrome is 1.78
# The odds of being a heavy drinker(HAC_group2) and having
# metabolic syndrome is 2.07 Unadjusted : The results
# revealed that heavy drinkers were more likely to have AHI
# than non-drinkers [odds ratio (OR) = 2.07 (1.32 - 3.31)]
# *****End Odds ratio
# (Unadjusted)*****

## *****Start glm (Adjusted for
## Age)***** Fit a logistic
## regression model to the data_table4 data set, using
## HAC_group + Age as predictors.

AHI_age_glm_out <- glm(SevereAHI ~ HAC_group + Age, data = data_table4,
  family = "binomial")
summary(AHI_age_glm_out)

##
## Call:
## glm(formula = SevereAHI ~ HAC_group + Age, family = "binomial",
## data = data_table4)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2140  -0.9791  -0.8100   1.3405   1.7770
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.06295    0.61708  -0.102  0.91875
## HAC_group1   0.53584    0.22706   2.360  0.01828 *
## HAC_group2   0.67677    0.23669   2.859  0.00425 **
## Age          -0.01844    0.01049  -1.757  0.07885 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 891.55  on 682  degrees of freedom
## Residual deviance: 877.93  on 679  degrees of freedom
## AIC: 885.93
##
## Number of Fisher Scoring iterations: 4

## *****Start Odds ratio (Adjusted for
## Age)***** odds ratios only
round(exp(coef(AHI_age_glm_out)), 2)

## (Intercept)  HAC_group1  HAC_group2      Age
##           0.94         1.71         1.97     0.98
```

```
## odds ratios and 95% CI
round(exp(cbind(OR = coef(AHI_age_glm_out), confint(AHI_age_glm_out))),
2)
```

```
##           OR 2.5 % 97.5 %
## (Intercept) 0.94 0.28 3.15
## HAC_group1 1.71 1.10 2.69
## HAC_group2 1.97 1.24 3.15
## Age        0.98 0.96 1.00
```

```
# So, we can say that : The odds of being a light
# drinker(HAC_group1) and having metabolic syndrome is 1.71
# The odds of being a heavy drinker(HAC_group2) and having
# metabolic syndrome is 1.97 Adjusted(Age) : The results
# revealed that heavy drinkers were more likely to have AHI
# than non-drinkers [odds ratio (OR) = 1.97 (1.24 - 3.15)]
# *****End Odds ratio (Adjusted for
# Age)*****
```

```
## *****Start glm (Adjusted for Age and
## BMI)***** Fit a logistic regression
## model to the data_table4 data set, using HAC_group + Age +
## BMI as predictors.
AHI_age_mets_glm_out <- glm(SevereAHI ~ HAC_group + Age + BMI,
data = data_table4, family = "binomial")
summary(AHI_age_mets_glm_out)
```

```
##
## Call:
## glm(formula = SevereAHI ~ HAC_group + Age + BMI, family = "binomial",
## data = data_table4)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5637  -0.9463  -0.6952   1.1992   2.0821
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.857297   1.176542  -5.828 5.60e-09 ***
## HAC_group1    0.722881   0.241031   2.999 0.00271 **
## HAC_group2    0.747106   0.249295   2.997 0.00273 **
## Age         -0.004729   0.011172  -0.423 0.67209
## BMI          0.229206   0.033987   6.744 1.54e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 891.55  on 682  degrees of freedom
## Residual deviance: 822.62  on 678  degrees of freedom
## AIC: 832.62
##
## Number of Fisher Scoring iterations: 4
```

```
# HAC_group + Age +BMI

## *****Start Odds ratio (Adjusted for Age
## and BMI)***** odds ratios only
round(exp(coef(AHI_age_mets_glm_out)), 2)

## (Intercept) HAC_group1 HAC_group2 Age BMI
## 0.00 2.06 2.11 1.00 1.26

## odds ratios and 95% CI
round(exp(cbind(OR = coef(AHI_age_mets_glm_out), confint(AHI_age_mets_glm_out))),
2)

## OR 2.5 % 97.5 %
## (Intercept) 0.00 0.00 0.01
## HAC_group1 2.06 1.30 3.34
## HAC_group2 2.11 1.30 3.47
## Age 1.00 0.97 1.02
## BMI 1.26 1.18 1.35

# So, we can say that : The odds of being a light
# drinker(HAC_group1) and having metabolic syndrome is 2.06
# The odds of being a heavy drinker(HAC_group2) and having
# metabolic syndrome is 2.11 Adjusted(Age) : The results
# revealed that heavy drinkers were more likely to have AHI
# than non-drinkers [odds ratio (OR) = 2.11 (1.30 - 3.47)]
# *****End Odds ratio (Adjusted for Age and
# BMI)*****
```

Table 4. Odds ratio (95% Confidence Intervals) for Severe apnea-hypopnea index according to habitual alcohol consumption.

	Unadjusted	Adjusted for Age	Adjusted for Age and BMI
Non-drinkers	1.0(reference)	1.0(reference)	1.0(reference)
Light-drinkers	1.78(1.15 - 2.79)	1.71(1.10 - 2.69)	2.06(1.30 - 3.34)
Heavy-drinkers	2.07(1.32 - 3.31)	1.97(1.24 - 3.15)	2.11(1.30 - 3.47)
AHI: apnea-hypopnea index BMI : body-mass index			

Produce ‘Results’ section from research paper (such as F-statistic, p-value) :

Click Here - Results Section

```
result_data_1 <- subset(OSA_HAC_Data, select = c(MetS, HAC_group,
Age, BMI, Waist_C))
result_data_1$MetS <- factor(result_data_1$MetS)
result_data_1$HAC_group <- factor(result_data_1$HAC_group)
glimpse(result_data_1)

## Observations: 683
## Variables: 5
```

```
## $ MetS      <fctr> 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, ...
## $ HAC_group <fctr> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ Age       <dbl> 66.7, 58.5, 48.8, 62.2, 51.1, 64.2, 61.8, 55.2, 58.0...
## $ BMI       <dbl> 24.51, 22.90, 23.50, 28.00, 27.21, 27.31, 23.59, 24....
## $ Waist_C   <dbl> 87, 83, 84, 90, 97, 91, 90, 78, 83, 84, 75, 84, 89, ...
```

```
# Heavy drinkers consumed a weekly average of 30.7±17.29
# drinks. Heavy drinkers were also the youngest (F = 7.24, p
# = 0.001)
```

```
hdrinkers_youngest_lm_out <- lm(Age ~ HAC_group, data = result_data_1)
summary(hdrinkers_youngest_lm_out)
```

```
##
## Call:
## lm(formula = Age ~ HAC_group, data = result_data_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -33.158  -4.720  -0.158   5.202  20.586
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   56.4812     0.6442  87.682 < 2e-16 ***
## HAC_group1    -2.3230     0.7799  -2.978 0.003001 **
## HAC_group2    -3.0673     0.8214  -3.734 0.000204 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.73 on 680 degrees of freedom
## Multiple R-squared:  0.02086,    Adjusted R-squared:  0.01798
## F-statistic: 7.242 on 2 and 680 DF,  p-value: 0.0007724
```

```
# Heavy drinkers also had the highest BMI (F = 3.65, p =
# 0.026)
hdrinkers_BMI_lm_out <- lm(BMI ~ HAC_group, data = result_data_1)
summary(hdrinkers_BMI_lm_out)
```

```
##
## Call:
## lm(formula = BMI ~ HAC_group, data = result_data_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.0393 -1.8314 -0.2593  1.4722 27.2807
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   25.8893     0.2571 100.687 <2e-16 ***
## HAC_group1    -0.4558     0.3113  -1.464  0.144
## HAC_group2     0.2570     0.3279   0.784  0.434
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.086 on 680 degrees of freedom
## Multiple R-squared:  0.01064,    Adjusted R-squared:  0.007726
```

```
## F-statistic: 3.655 on 2 and 680 DF, p-value: 0.02637
# Heavy drinkers also had the largest waist circumference (F
# = 3.28, p = 0.038)
hdrinkers_WC_lm_out <- lm(Waist_C ~ HAC_group, data = result_data_1)
summary(hdrinkers_WC_lm_out)

##
## Call:
## lm(formula = Waist_C ~ HAC_group, data = result_data_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.948  -3.948  -0.698   4.052  55.448
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  90.5521     0.6553  138.189  <2e-16 ***
## HAC_group1   -0.6039     0.7934   -0.761    0.447
## HAC_group2    1.1457     0.8356    1.371    0.171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.863 on 680 degrees of freedom
## Multiple R-squared:  0.009562, Adjusted R-squared:  0.006649
## F-statistic: 3.283 on 2 and 680 DF, p-value: 0.03813
# The prevalence of MetS in participants with SDB was 36.9%
# and was high in heavy drinkers (44.3%, X2 = 8.51, p =
# 0.014).

# First, lets calculate the prevalence of MetS in
# participants with SDB, which is 36.9%
chi_MetS_SDB_table <- subset(OSA_HAC_Data, select = c(MetS, AHI_group))
chi_MetS_SDB_matrix <- as.data.frame.matrix(table(chi_MetS_SDB_table$MetS,
  chi_MetS_SDB_table$AHI_group))
colnames(chi_MetS_SDB_matrix) <- c("mild_SDB", "moderate_SDB",
  "severe_SDB")
rownames(chi_MetS_SDB_matrix) <- c("MetS_NO", "MetS_YES")
chi_MetS_SDB_matrix$Total <- rowSums(chi_MetS_SDB_matrix)
chi_MetS_SDB_matrix

##           mild_SDB moderate_SDB severe_SDB Total
## MetS_NO         157          142         132   431
## MetS_YES          58           81         113   252

prevalence_MetS_SDB_result <- chi_MetS_SDB_matrix["MetS_YES",
  "Total"]/(chi_MetS_SDB_matrix["MetS_NO", "Total"] + chi_MetS_SDB_matrix["MetS_YES",
  "Total"]) * 100
prevalence_MetS_SDB_result

## [1] 36.89605

# Now, lets calculate the second part, that is, prevalence of
# MetS was high in heavy drinkers (44.3%, X2 = 8.51, p =
# 0.014).
chi_MetS_HD_table <- subset(OSA_HAC_Data, select = c(MetS, HAC_group))
```

```

chi_MetS_HD_matrix <- as.data.frame.matrix(table(chi_MetS_HD_table$MetS,
  chi_MetS_HD_table$HAC_group))
colnames(chi_MetS_HD_matrix) <- c("non-drinkers", "light-drinkers",
  "heavy-drinkers")
rownames(chi_MetS_HD_matrix) <- c("MetS_NO", "MetS_YES")
chi_MetS_HD_matrix

##           non-drinkers light-drinkers heavy-drinkers
## MetS_NO           94           209           128
## MetS_YES           50           100           102

# prevalence of MetS was high in heavy drinkers ---- 44.3%,
prevalence_MetS_HD_result <- chi_MetS_HD_matrix["MetS_YES", "heavy-drinkers"]/(chi_MetS_HD_matrix["MetS",
  "heavy-drinkers"] + chi_MetS_HD_matrix["MetS_YES", "heavy-drinkers"]) *
  100
prevalence_MetS_HD_result

## [1] 44.34783

# prevalence of MetS was high in heavy drinkers (X2 = 8.51, p
# = 0.014).
chi_MetS_HD_value <- chisq.test(chi_MetS_HD_matrix)
chi_MetS_HD_value

##
## Pearson's Chi-squared test
##
## data:  chi_MetS_HD_matrix
## X-squared = 8.5055, df = 2, p-value = 0.01423

```

Final report - 3 extension pieces

An additional plot

- Through the boxplot and violin plots below, we can say that the severity of SDB was the greatest in heavy drinkers than in those with non-drinkers or light-drinkers.

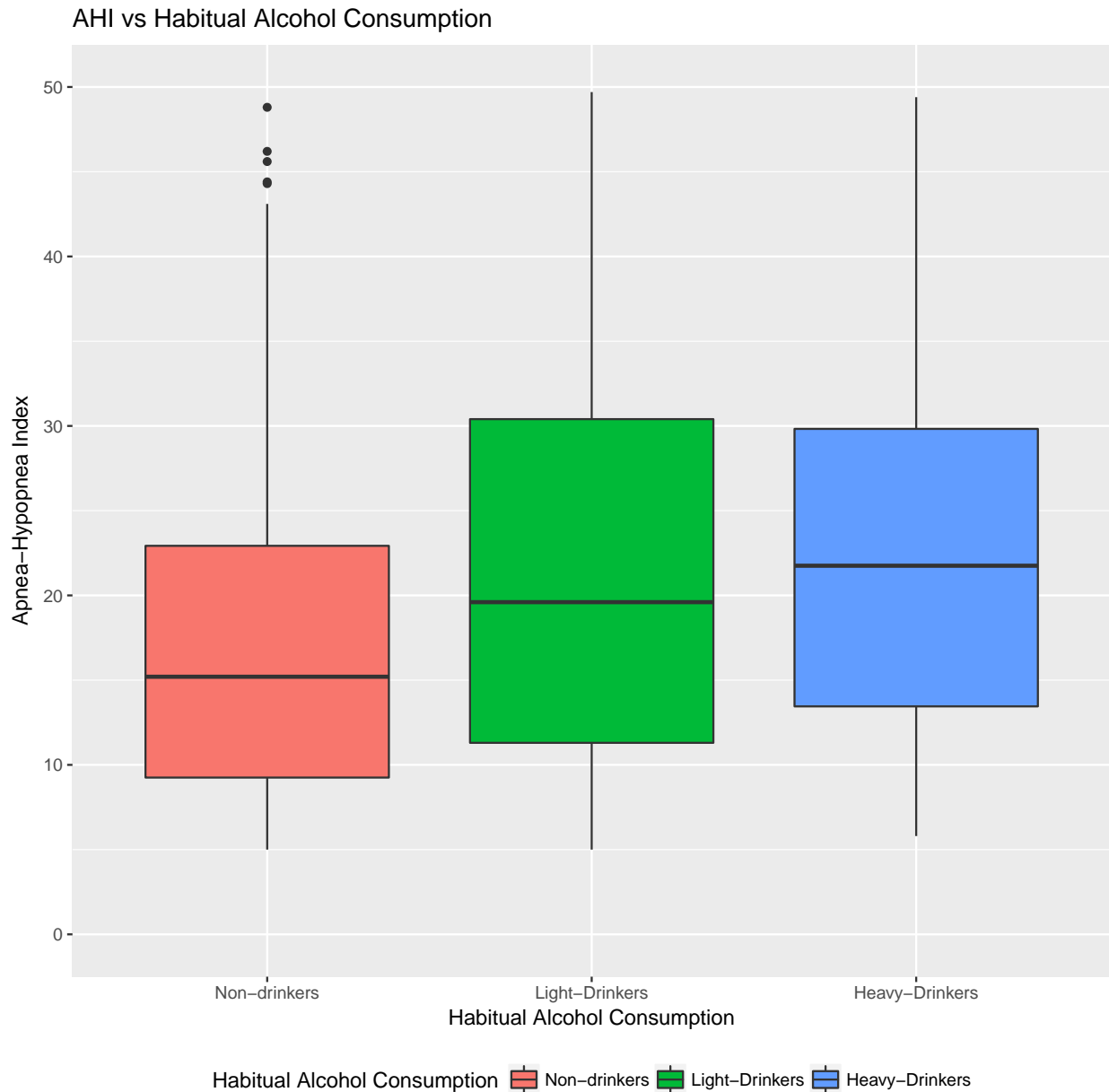
```

ahi_hac_data <- subset(OSA_HAC_Data, select = c(AHI, HAC_group))

ahi_hac_data$HAC_group <- factor(ahi_hac_data$HAC_group)

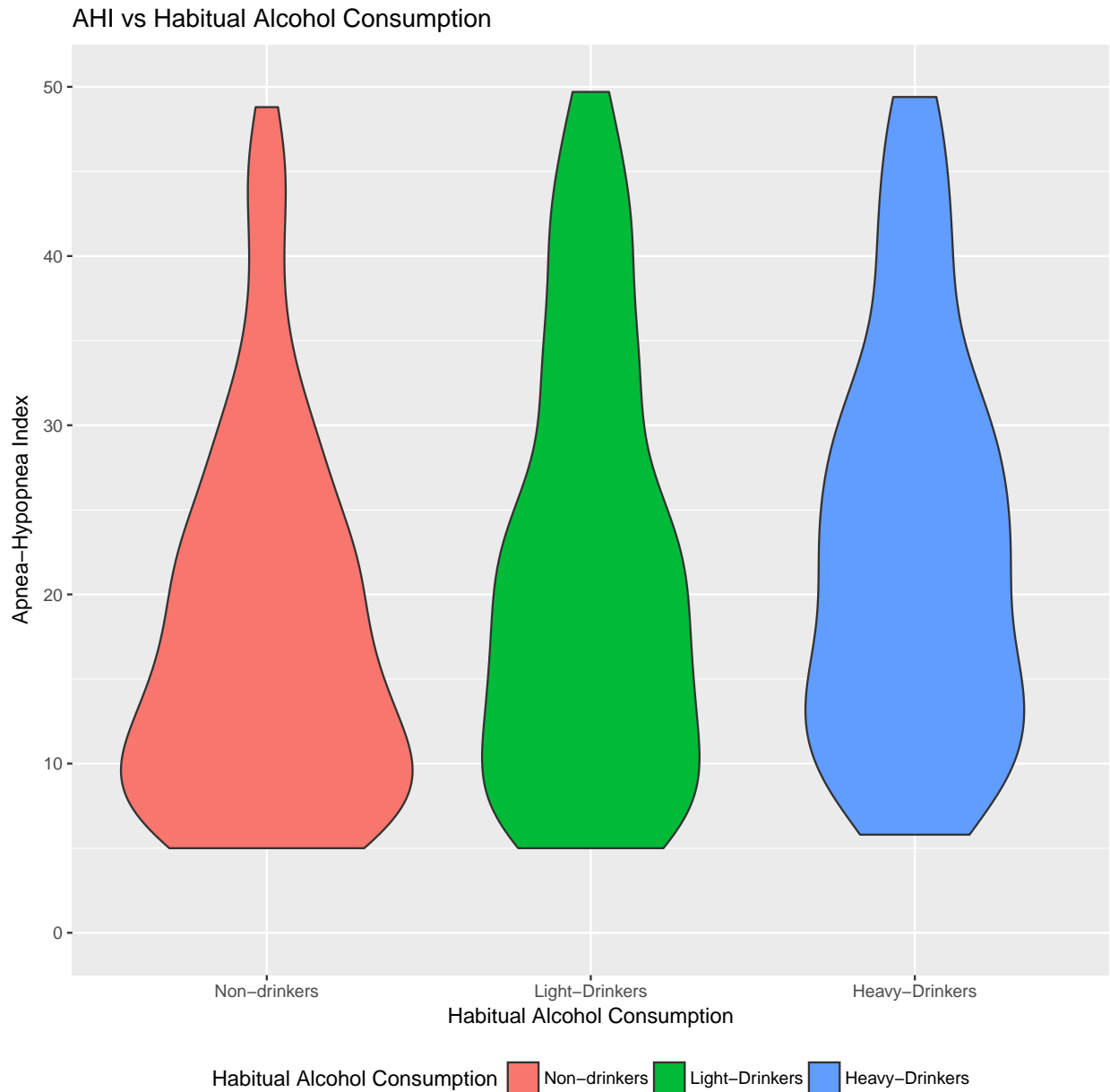
ggplot(ahi_hac_data, aes(y = AHI, x = HAC_group, fill = HAC_group)) +
  geom_boxplot() + ggtitle("AHI vs Habitual Alcohol Consumption") +
  labs(x = "Habitual Alcohol Consumption", y = "Apnea-Hypopnea Index") +
  scale_x_discrete(labels = c("Non-drinkers", "Light-Drinkers",
    "Heavy-Drinkers")) + scale_y_continuous(limits = c(0,
    50)) + scale_fill_discrete(name = "Habitual Alcohol Consumption",
    labels = c("Non-drinkers", "Light-Drinkers", "Heavy-Drinkers")) +
  theme(legend.position = "bottom")

```



```
ggplot(ahi_hac_data, aes(y = AHI, x = HAC_group, fill = HAC_group)) +
  geom_violin() + ggtitle("AHI vs Habitual Alcohol Consumption") +
  labs(x = "Habitual Alcohol Consumption", y = "Apnea-Hypopnea Index") +
  scale_x_discrete(labels = c("Non-drinkers", "Light-Drinkers",
    "Heavy-Drinkers")) + scale_y_continuous(limits = c(0,
    50)) + scale_fill_discrete(name = "Habitual Alcohol Consumption",
    labels = c("Non-drinkers", "Light-Drinkers", "Heavy-Drinkers")) +
  theme(legend.position = "bottom")
```

```
## Warning: Removed 118 rows containing non-finite values (stat_ydensity).
```

An additional analysis using a method we have covered this quarter in class that answers a new and different question or hypothesis you have come up with yourself

- After adjusting for age and BMI, we check if light-drinkers and heavy-drinkers are more habitual to smoking.
- So, to examine the association between HAC and Smoking after adjusting for possible confounders (age and BMI), we performed multiple logistic regression analyses.
- The results of the logistic regression analysis are reported as adjusted odds ratio (OR) with 95% confidence interval (CI).
- **Analyses result :** After adjusting for age and BMI, we found that light drinkers had a 1.88 times higher chances of being more habituated to smoking and heavy drinkers had a 2.14 times higher chances compared with non-drinkers.

```

smoking_glm_table <- subset(OSA_HAC_Data, select = c(Smoking,
  HAC_group, Age, BMI))

# Filter the Ex-smoker values, now smoking column consists of
# participants who are either non-smokers or current smokers.
# This is because we require the smoking column to be
# binomial to perform the regression analyses.

smoking_glm_table <- filter(smoking_glm_table, Smoking == "N" |
  Smoking == "Y")
smoking_glm_table$Smoking <- ifelse(smoking_glm_table$Smoking ==
  "N", 0, 1)
smoking_glm_table$Smoking <- factor(smoking_glm_table$Smoking)
smoking_glm_table$HAC_group <- factor(smoking_glm_table$HAC_group)
glimpse(smoking_glm_table)

## Observations: 462
## Variables: 4
## $ Smoking <fctr> 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, ...
## $ HAC_group <fctr> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ Age <dbl> 66.7, 58.5, 48.8, 62.2, 51.1, 64.2, 61.8, 58.0, 64.6...
## $ BMI <dbl> 24.51, 22.90, 23.50, 28.00, 27.21, 27.31, 23.59, 23....

## *****Start glm (Adjusted for Age and
## BMI)***** Fit a logistic regression
## model to the smoking_glm_table data set, using HAC_group +
## Age + BMI as predictors.

smoking_glm_out <- glm(Smoking ~ HAC_group + Age + BMI, data = smoking_glm_table,
  family = "binomial")
summary(smoking_glm_out)

##
## Call:
## glm(formula = Smoking ~ HAC_group + Age + BMI, family = "binomial",
## data = smoking_glm_table)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -1.7324 -1.0198 -0.7796 1.2911 1.7956
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.10431 1.20787 -0.914 0.36058
## HAC_group1 0.63165 0.27425 2.303 0.02127 *
## HAC_group2 0.76271 0.28544 2.672 0.00754 **
## Age -0.02331 0.01281 -1.820 0.06877 .
## BMI 0.05126 0.03134 1.635 0.10195
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 614.99 on 461 degrees of freedom

```

```
## Residual deviance: 597.27 on 457 degrees of freedom
## AIC: 607.27
##
## Number of Fisher Scoring iterations: 4
## odds ratios only
round(exp(coef(smoking_glm_out)), 2)

## (Intercept) HAC_group1 HAC_group2 Age BMI
## 0.33 1.88 2.14 0.98 1.05

## odds ratios and 95% CI
round(exp(cbind(OR = coef(smoking_glm_out), confint(smoking_glm_out))),
2)

## OR 2.5 % 97.5 %
## (Intercept) 0.33 0.03 3.50
## HAC_group1 1.88 1.11 3.26
## HAC_group2 2.14 1.24 3.79
## Age 0.98 0.95 1.00
## BMI 1.05 0.99 1.12

# So, we can say that : The odds of being a light
# drinker(HAC_group1) and habitual to smoking is 1.88 The
# odds of being a heavy drinker(HAC_group2) and habitual to
# smoking is 2.14 Adjusted(Age,BMI) : The results revealed
# that heavy drinkers were more likely habitual to smoking
# than non-drinkers [odds ratio (OR) = 2.14 (1.24 - 3.79)]

# After adjusting for age and BMI, heavy drinkers [OR =
# 2.14(95% CI = 1.24-3.79)] and light drinkers [OR = 1.88
# (95% CI = 1.11-3.26)] were more likely habitual to smoking
# than were the non-drinkers.
```

Table 5 : Odds ratio (95% Confidence Intervals) for Smoking according to habitual alcohol consumption.

	Adjusted for Age and BMI
Non-drinkers	1.0(reference)
Light-drinkers	1.88 (1.11 - 3.26)
Heavy-drinkers	2.14 (1.24 - 3.79)

An additional analysis using resampling based methods (bootstrapping or permutation tests) (again, this will come later; note that this does not have to be a new question or hypothesis- you can simply be testing the same hypothesis already tested in the paper using non-resampling based methods)

- We perform resampling methods to one of our previous analyses.
- Analyses : The prevalence of MetS in participants with SDB was 36.9% and was high in heavy drinkers (44.3%, $X^2 = 8.51$, $p = 0.014$).
- So, we perform resampling based permutation testing to compute the p-value(prevalence of MetS was high in heavy drinkers ($X^2 = 8.51$, $p = 0.014$)).
- So, the p-value of 0.014 tells us that the high prevalence of MetS in heavy drinkers is not due to chance.

```

# prevalence of MetS was high in heavy drinkers ( $X^2 = 8.51$ ,  $p = 0.014$ ).

# Earlier analysis chi_MetS_HD_table <- subset(OSA_HAC_Data,
# select = c(MetS,HAC_group)) chi_MetS_HD_table
# chi_MetS_HD_matrix <-
# as.data.frame.matrix(table(chi_MetS_HD_table$MetS,
# chi_MetS_HD_table$HAC_group))
# colnames(chi_MetS_HD_matrix)<-
# c('non-drinkers','light-drinkers','heavy-drinkers')
# rownames(chi_MetS_HD_matrix)<- c('MetS_NO', 'MetS_YES')
# chi_MetS_HD_matrix

# chisq.test(chi_MetS_HD_matrix) Pearson's Chi-squared test
# data: chi_MetS_HD_matrix X-squared = 8.5055, df = 2,
# p-value = 0.01423

# We perform resampling method (permutation test) to prove
# the same hypotheses that prevalence of MetS was high in
# heavy drinkers ( $X^2 = 8.51$ ,  $p = 0.014$ ).

chi_MetS_HD_table <- subset(OSA_HAC_Data, select = c(MetS, HAC_group))

chisq <- function(Obs) {
  Expected <- outer(rowSums(Obs), colSums(Obs))/sum(Obs)
  sum((Obs - Expected)^2/Expected)
}

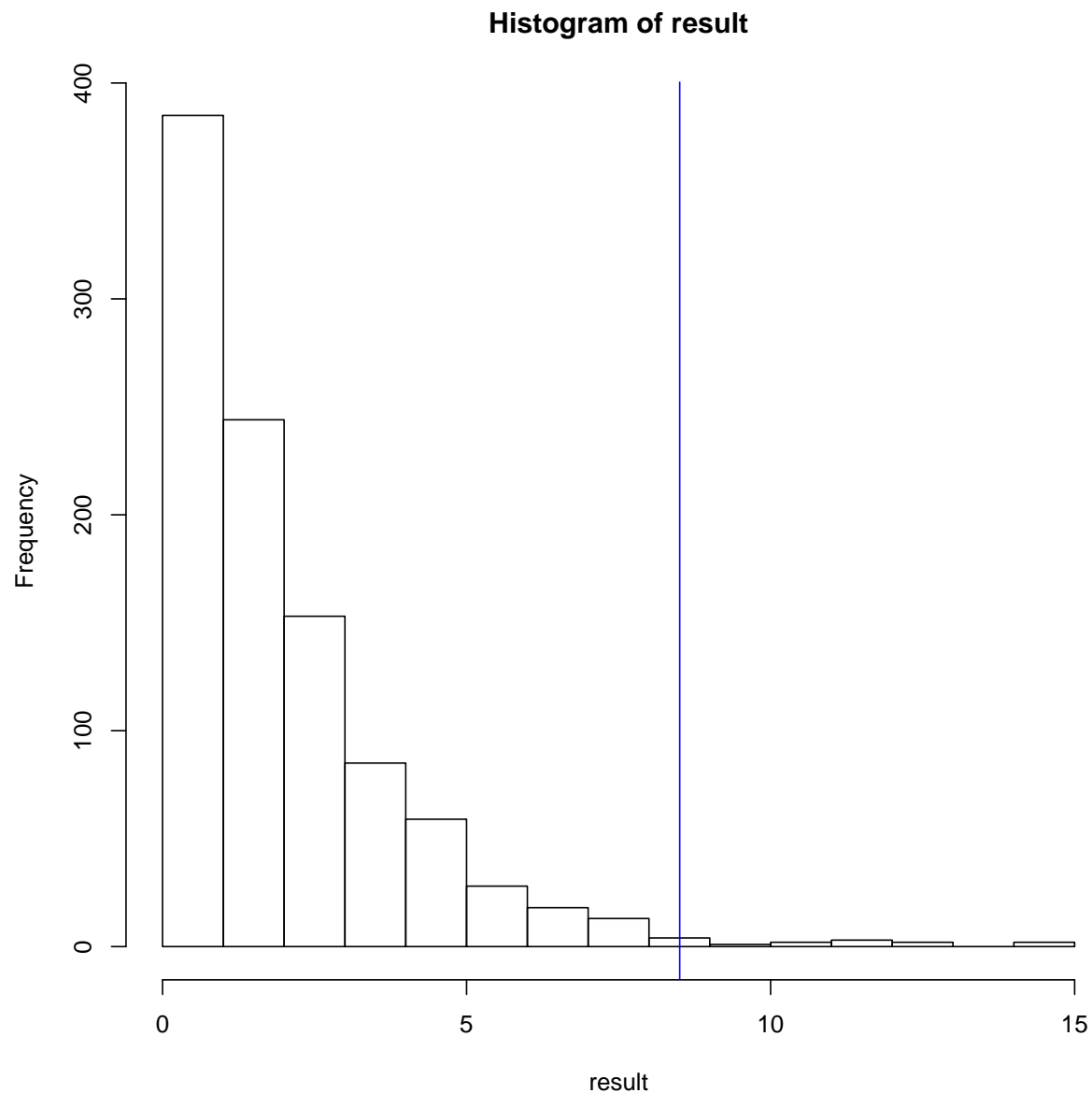
observed_value <- chisq(table(chi_MetS_HD_table$MetS, chi_MetS_HD_table$HAC_group))

B <- 10^3 - 1
result <- numeric(B)

for (i in 1:B) {
  Response.permuted <- sample(chi_MetS_HD_table$HAC_group)
  chi_MetS_HD.table <- table(chi_MetS_HD_table$MetS, Response.permuted)
  result[i] <- chisq(chi_MetS_HD.table)
}

hist(result)
abline(v = observed_value, col = "blue")

```



```
observed_value  # chi-square value  
## [1] 8.505462  
round((sum(result >= observed_value) + 1)/(B + 1), 5)  #P-value  
## [1] 0.013
```

Discussion :

1) Does alcohol consumption increase MetS?

- Our replication of the paper found that light, moderate, and heavy drinkers had increased ORs of 1.51,

- 1.71 and 2.11, respectively, for developing MetS compared with non-drinkers.
- Our replication of the analysis also suggested that heavy drinking increases waist and body weight as well as the prevalence of MetS.
- 2) Does habitual alcohol consumption aggravate SDB and sleep quality?
- Our replication of the analysis suggests that greater alcohol intake is associated with more severe SDB.
-

Other Terms :

1. Terms table

variables	abbreviation or subcategory
Sex	M=male
Age	
Smoking	N=never smoking; X=Ex-smoker; Y=current smoking
HAC/w	Habitual alcohol consumption, /week
Alc_freq/w	frequency of alcohol intake, /week
Alc glasses	Average intake, glassess/case
SBP	Systolic BP
DBP	Diastolic BP
Waist_C	Waist circumference, cm
BMI	Body mass index
MetS	Metabolic syndrome, 0=absent; 1=present
AHI	Apnea-hypopnea index
AHI_group	1= mild SDB; 2= moderate SDB; 3= severe SDB
HAC_group	0= non-drinker; 1= light-drinker; 2= heavy drinker

Figure 1:

2. Citation information for the paper you want to reproduce

Choi SJ, Lee SI, Joo EY (2016) Habitual Alcohol Consumption and Metabolic Syndrome in Patients with Sleep Disordered Breathing. PLoS ONE 11(8): e0161276. doi:10.1371/journal.pone.0161276

3. Link to the paper where we can access a PDF of the article

[Click Here](#) - Article Link

4. Direct link to the data repository

[Click Here](#) - Data repository link
