T tests and ANOVA

Your name and student ID today's date

Part 1: T tests and NHANES

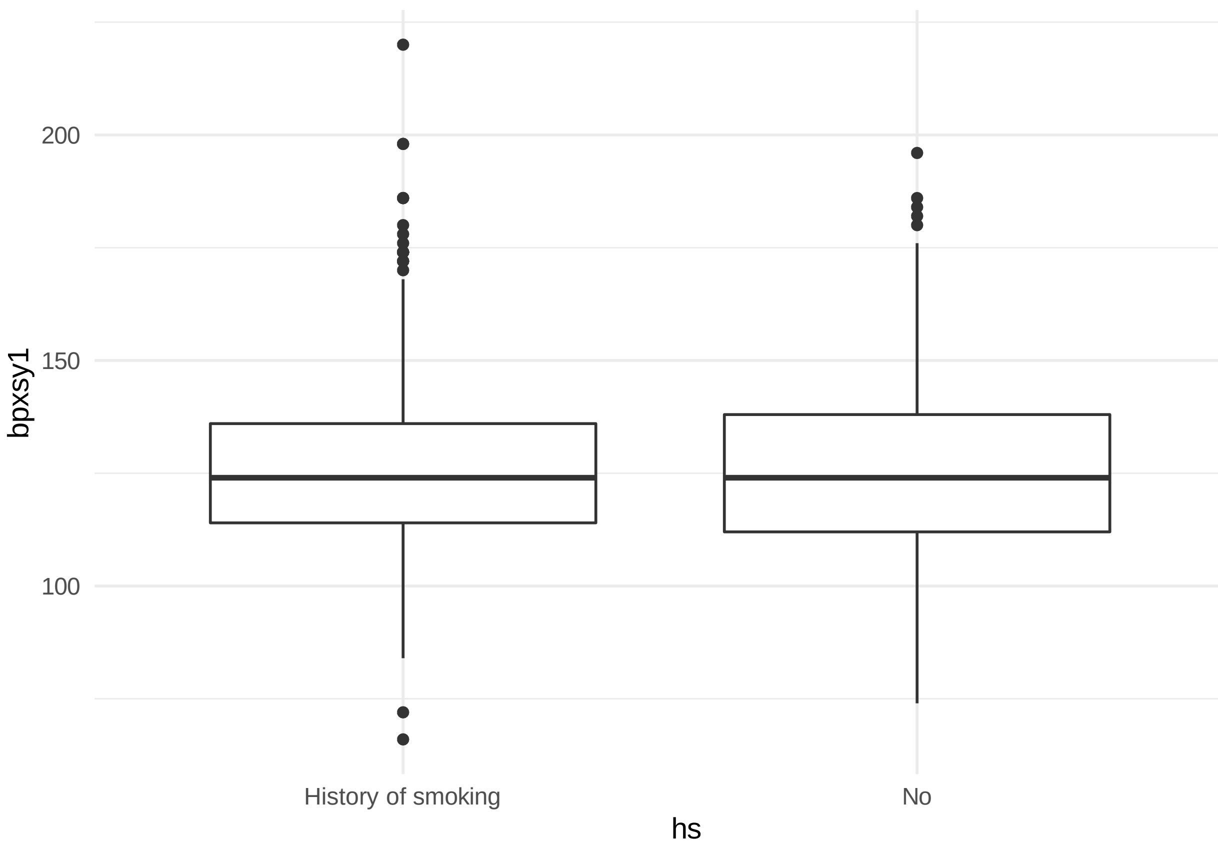
The NHANES is a large national survey conducted by the CDC. We will look at a reduced set of data from the NHANES for this lab.

```
##
## Attaching package: 'readr'

## The following objects are masked from 'package:testthat': ##
##     edition_get, local_edition

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:testthat': ##
##     matches

## The following objects are masked from 'package:stats': ##
##     filter, lag

## The following objects are masked from 'package:base': ##
##     intersect, setdiff, setequal, union
```

Rows: 2503 Columns: 40 ## Column specification ## Delimiter: ","
-chr (27): agegroup, ~~gender, military, born, citizen, drinkscat, bmicat, sys~~1... ## dbl (13): ridageyr, drinks, bmxwt, bmxht, bmxbmi, bpxpls, bpxsy1, bpxsy2, bp...
i Use 'spec()' to retrieve the full column specification for this data.
i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

1. [1 point] We are interested in looking at the systolic blood pressure, bpxsy, by history of smoking, hs. Start by generating an appropriate box plot to look at these data.

```
plot1 <- ggplot(nhanes, aes(x= hs, y = bpxsy1))+ geom_boxplot()+ theme_minimal()
plot1
```




```
nhanes$bpxdi1
```

```
#     [1]  90  86  80  80  72  84  78  58  72  56  66  70  74  46  98  70  66  62
#
#    [19]  68  68  52  66  80  76  62  70  88  68  82  58  68  74  72  64  60  64
#
#    [37]  52  64  72  58  74  52  94  72  90  66  76  78  66  68  64  68  68  54
#
#    [55]  70  48  86  74  58  80  66  52  52  64  66  70  64  70  66  66  72  90
#
#    [73]  84  58  68  70  74  52  76  80  94  68  78  86  62  72  68  82  58   0
#
#    [91]  78  48  72  68  76  70  66  80  82  72  76  80  76  48  86  58  48  62
#
#   [109]  60  84  68  74  56  66  88  52  92  56  72  52  72  70  68  68  76  62
```

```
#
#    [127]  66  56  74  70  76  68  52  78  76  58  52  64  74  76  66  80  72  86
#
#    [145]  74  78  90  66  66  92  72  72  62  66  64  56  66  62  72  78  68  68
#
#    [1     6 8   7 5   8   6   8   6   7   6   7   8   6   6 5   5   7   8
#     6     4 8   0 8   8   8   4   6   2   4   2   6   4   0 4   0   4   6
#     3]
#    [1     7 8   5 7   6   7   6   7   6   7   6   5   6   7 7   8   7   8
#     8     0 8   4 8   6   4   8   4   0   4   2   8   4   4 2   0   4   0
#     1]
#    [1     7 7   8 6   6   6   6   8   7   7   8   5   7   6 7   9   5   6
#     9     6 6   8 8   4   2   6   0   6   8   6   6   6   2 6   4   4   4
#     9]
#    [2     7 7   7 7   4   8   7   6   7   6   8   7   5   5 9   7   6   7
#     1     0 4   2 2   2   8   4   8   0   4   2   0   8   0 2   2   6   0
#     7]
#    [2     5 5   7 5   7   5   7   6   6   7   7   7   6   7 7   5   7   7
#     3     4 2   0 2   2   6   6   8   6   8   6   4   2   8 6   6   6   2
#     5]
#    [2     7 7   7 9   7   7   8   7   6   4   6   6   7   7 8   5   6   7
#     5     4 0   6 6   0   0   4   8   2   8   6   2   2   4 2   2   4   4
#     3]
#    [2     7 6   7 7   9   7   8   5   7   5   7   6   6   7 5   6   7   7
#     7     2 6   2 6   8   0   2   2   2   8   6   2   2   0 6   8   2   4
#     1]
#    [2     6 8   6 7   7   6   6   7   7   7   6   5   6   7 9   7   6   6
#     8     4 4   6 2   4   4   4   8   2   6   8   4   8   4 4   4   8   6
#     9]
#    [3     8 7   6 6   8   8   5   6   5   7   6   9   7   7 7   7   7   8
#     0     4 0   2 8   2   6   8   4   8   4   8   8   0   6 2   2   6   2
#     7]
#    [3     6 7   5 6   5   5   5   7   7   6   6   6   6   7 6   6   8   5
#     2     6 8   6 0   0   6   0   8   4   0   2   8   4   4 0   6   6   0
#     5]
#    [3     6 5   8 5   6   8   7   7   8   5   6   7   8   8 5   7   5   7
#     4     8 4   4 8   8   2   4   2   8   6   4   0   6   6 8   6   2   4
#     3]
#    [3     7 4   7 7   5   7   5   9   8   6   9   8   6   6 6   5   8   8
#     6     6 8   4 8   6   8   4   0   6   6   0   0   2   2 4   8   0   6
#     1]
#    [3     4 5   6 7   6   8   5   7   5   9   5   7   6   8 6   7   6   7
#     7     2 6   4 6   6   4   6   8   8   0   4   8   4   8 8   6   2   6
#     9]
#    [3     5 6   5 7   5   6   4   6   5   6   7   6   6   5 6   8   8   6
#     9     8 0   0 8   6   0   8   2   8   6   2   8   6   2 6   6   8   2
#     7]
#    [4     5 6   6 6   8   7   7   6   6   7   7   6   6   7 7   7   4   5
#     1     4 4   8 4   0   2   0   2   2   6   4   6   4   4 2   8   8   6
#     5]
#    [4     8 6   6 5   6   6   8   7   6   7   6   6   5   8 6   9   7   8
#     3     4 6   8 2   2   4   8   2   4   0   4   0   2   4 4   0   0   6
#     3]
#    [4     4 5   6 6   7   7   5   5   7   4   7   8   7   8 7   9   7   6
#     5     8 8   8 8   0   4   8   4   0   4   4   0   2   6 0   2   4   6
#     1]
#    [4     5 8   8 5   7   8   6   3   6   8   7   7   7   6 7   6   6   8
#     6     4 2   0 0   0   4   0   0   6   4   8   8   4   0 2   6   2   6
#     9]
```

```
#  [4  5 9  5 1  6  8 8 7 6 7 7 8 7  5 7  5 0  7
#  8   4 4  4 0  0  4 6 2 0 0 2 2 6  2 4  2    8
   7]         6
#  [5  8 3  5 9  0  6 7 7 8 7 7 5 7  8 5  7 7  8
#  0   2 4  6 8     6 2 6 2 4 2 6 8  4 6  6 4  4
   5]
#  [5  7 6  5 7  6  7 5 6 7 9 5 7 8  7 8  7 8  9
#  2   6 6  8 8  2  0 0 0 4 2 8 4 6  0 2  8 6  8
   3]
#  [5  7 5  7 6  9  5 7 7 6 7 6 5 6  6 7  7 7  7
#  4   6 8  4 8  2  6 0 6 8 4 4 8 4  8 2  0 6  6
   1]
#  [5  6 1  6 6  4  6 8 7 9 6 8 5 8  8 8  7 7  6
#  5   4 0  2 8  6  2 8 4 6 6 2 2 6  6 4  0 2  8
   9]    6
#  [5  7 7  8 7  6  7 7 8 5 6 7 4 5  6 1  7 7  6
#  7   8 2  2 4  0  8 0 0 4 8 8 4 4  4 1  6 6  6
   7]                                  2
#  [5  7 8  7 0  7  8 7 8 7 7 6 5 7  7 7  7 7  6
#  9   4 8  0    0  2 0 0 6 4 4 8 8  2 2  8 8  0
   5]
#  [6  5 7  6 8  5  7 7 7 6 7 7 7 7  0 6  7 7  6
#  1   8 4  4 8  8  4 6 4 8 2 6 6 8    4  6 2  6
   3]
#  [6  9 8  6 7  5  7 4 8 7 7 6 7 8  5 6  4 5  7
#  3   4 0  6 4  0  2 8 4 4 2 2 4 8  8 2  2 6  2
   1]
#  [6  5 7  8 6  4  7 3 7 6 0 5 8 7  0 6  7 6  5
#  4   6 6  6 6  8  4 6 2 4   6 0 2    4  0 2  0
   9]
#  [6  9 7  8 7  8  8 5 5 7 6 6 7 6  6 6  6 6  7
#  6   2 6  0 0  0  4 6 2 6 6 4 0 8  8 0  4 2  2
   7]
#  [6  6 6  6 6  7  1 6 7 7 6 9 5 8  7 1  7 6  7
#  8   4 4  8 4  0  0 8 8 2 6 0 6 2  2 2  0 2  6
   5]               4                  2
#  [7  6 7  6 1  6  5 7 7 6 5 6 5 7  7 6  7 5  7
#  0   6 2  6 0  8  8 2 8 6 6 8 8 8  2 4  6 8  2
   3]         0
#  [7  8 8  7 5  7  7 7 6 8 7 7 7 7  8 0  7 8  8
#  2   6 0  4 0  4  6 6 0 2 8 2 0 4  0    0 2  0
   1]
#  [7  5 7  8 7  7  7 5 8 6 7 5 7 4  7 7  8 7  9
#  3   0 2  4 6  4  4 4 4 8 8 4 2 0  6 6  0 4  0
   9]
#  [7  5 8  5 6  8  4 7 7 7 6 8 7 9  7 6  5 8  7
#  5   6 0  8 8  2  4 0 0 0 8 2 2 8  6 4  4 2  0
   7]
#  [7  7 6  7 7  6  6 8 7 8 6 7 5 6  8 7  7 7  4
#  7   6 8  4 6  6  2 0 6 2 0 2 8 6  4 6  2 2  8
   5]
#  [7  7 8  8 7  5  6 6 8 6 8 5 7 7  7 6  6 6  7
#  9   0 2  2 0  8  0 6 4 8 0 8 6 0  8 4  6 6  2
   3]
#  [8  4 7  6 6  7  6 7 6 6 9 5 6 6  6 7  6 2  6
#  1   4 0  4 6  2  8 2 2 8 6 8 6 2  2 8  8 8  6
   1]
#  [8  6 7  8 7  6  5 8 7 7 5 6 5 8  7 9  6 8  7
#  2   8 0  2 0  0  6 0 4 4 0 6 8 0  4 4  2 4  4
   9]
```

#	[8	5	6	7	7	7	6	4	7	7	8	7	6	8	7	6	5	7	6
#	4	8	8	4	2	6	8	8	2	6	0	2	4	6	2	0	2	6	6
	7]																		
#	[8	7	5	6	5	5	9	9	4	5	7	7	7	9	6	7	8	7	6
#	6	0	6	0	4	2	4	6	6	6	0	8	6	0	2	4	2	6	0
	5]																		
#	[8	7	4	7	6	6	1	6	8	6	6	6	6	7	8	4	6	6	5
#	8	4	4	8	6	0	0	6	0	0	2	0	4	8	6	4	2	6	8
	3]						0												
#	[9	8	7	7	7	5	7	5	8	7	7	5	8	7	7	6	6	6	8
#	0	0	4	4	8	8	8	8	0	2	6	6	4	2	8	0	4	6	2
	1]																		
#	[9	6	7	8	4	5	8	4	7	8	7	7	6	8	7	8	7	8	6
#	1	6	8	8	8	8	2	4	4	6	8	2	8	0	2	0	8	8	8
	9]																		
#	[9	6	8	7	6	6	7	7	6	8	6	7	8	8	8	8	7	7	8
#	3	8	2	0	8	2	2	2	6	4	8	4	0	6	6	0	0	4	4
	7]																		
#	[9	3	8	7	8	5	6	7	7	2	7	8	7	7	5	7	8	6	7
#	5	8	0	6	4	8	4	2	8	6	2	0	4	8	6	4	2	8	8
	5]																		
#	[9	7	6	6	6	7	7	7	6	7	7	6	6	9	7	7	8	7	6
#	7	2	4	8	4	8	8	4	4	2	8	4	6	0	4	0	0	0	6
	3]																		
#	[9	5	6	6	5	7	8	6	5	6	5	7	6	5	7	8	6	7	5
#	9	6	8	2	6	4	4	6	6	4	8	6	8	8	2	2	6	4	8
	1]																		
#	[10	6	7	7	7	1	7	6	4	7	6	6	8	6	8	8	8	7	5
#	09]	2	2	4	8	0	8	2	0	2	6	0	2	6	4	6	4	4	6
						0													
#	[10	7	8	7	6	7	6	6	7	6	5	7	8	6	7	5	5	0	7
#	27]	8	2	2	4	2	2	4	0	8	4	8	8	0	4	2	6		4
#	[10	8	5	3	7	6	7	5	9	7	6	8	6	7	6	6	6	0	7
#	45]	0	2	4	6	2	2	6	8	4	0	4	8	8	8	8	8		4
#	[10	1	7	7	5	7	8	7	7	5	8	6	7	6	6	6	7	7	6
#	63]	0	0	2	6	6	8	8	2	2	2	2	0	0	0	2	6	4	6
		0																	
#	[10	6	7	6	7	7	5	5	8	6	7	5	0	7	5	5	7	7	8
#	81]	4	6	8	0	0	8	2	0	0	0	0		4	8	4	4	4	4
#	[10	7	8	6	8	6	7	7	6	7	8	7	7	7	7	5	7	8	6
#	99]	8	4	0	2	4	6	4	8	2	2	2	8	6	4	4	8	2	6
#	[11	9	9	6	6	6	5	5	8	4	7	8	6	6	6	5	8	7	7
#	17]	0	4	8	8	6	8	8	8	6	0	2	6	8	4	6	2	0	2

```
## [1135]   62  88  78  86  66  52  66  72  90  66  72  82  74  70  76  70  72  84
## [1153]   70  48  60  88  92  68  64  76  86  64  74  80  78  70  82  58  76  46
## [1171]   86  82  66  78  78  60  86  70
```

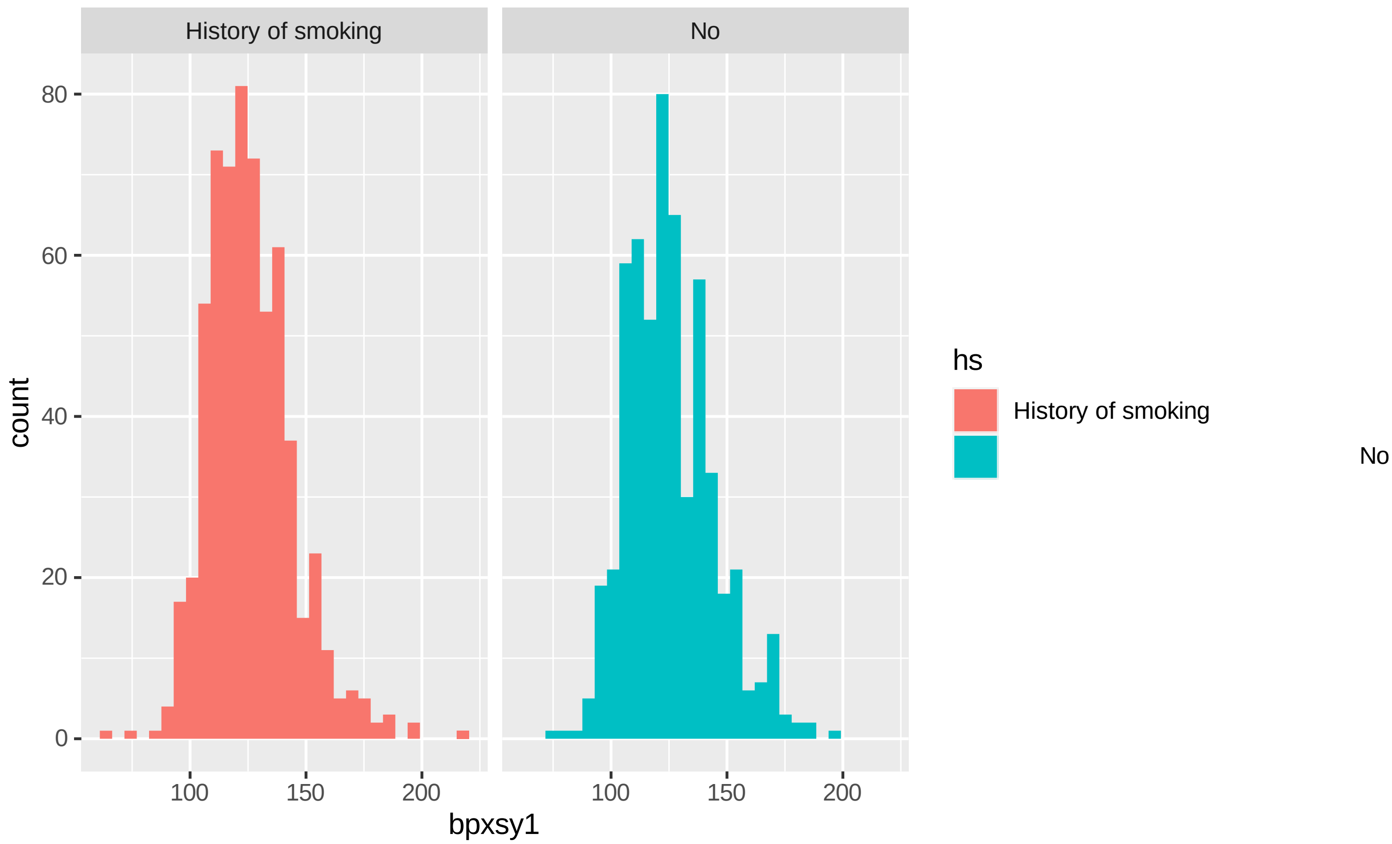
```r
. = ottr::check("tests/p1.R")
```

```
## 
## All tests passed!
```

2. [1 point] Now generate a set of faceted histograms that show the same data.

```
plot2 <- ggplot(nhanes, aes(x = bpxsy1)) + geom_histogram(aes(fill = hs)) +
         facet_wrap(~hs)
plot2
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
. = ottr::check("tests/p2.R")
```

```
##
## All tests passed!
```

3. [1 point] Summarize the means and standard deviations of the systolic blood pressurea for each category of hs. Assign p3 to a dataframe with the mean systolic blood pressures assigned to mean_bp and the standard deviations assigned to sd_bp.

```
p3 <- nhanes %>% group_by(hs)%>% summarize (mean_bp = mean(bpxsy1),
                                        sd_bp = sd(bpxsy1) )
p3
```

```
## # A tibble: 2 x 3
##   hs                 mean_bp sd_bp ##   <chr>              <dbl> <dbl> ## 1 History of smoking 126.   18.6
## 2 No                    126.   18.7
```

```
. = ottr::check("tests/p3.R")
```

```
##
## All tests passed!
```

4. Do we meet the all of the assumptions to run a two-sample t-test? Why or why not?

1. The observation are independent. 2, Mean (group) is normally distributed.
2. The sample(group) variance needs to be equal.

5. State the null and alternative hypotheses in the context of this question.

Ho : The mean SBP of smokers is equal to the mean SBP of non-smokers. Ha : The mean SBP of smokers is not equal to the mean SBP of non-smokers.

6. [1 point] Use an R function to test if the variability gives enough evidence to reject the null hypothesis of no difference between mean blood pressure by smoking history.

```
p6 <- t.test(bpxsy1 ~ hs, data = nhanes)
p6
```

```
##
##   Welch Two Sample t-test ##
## data:   bpxsy1 by hs
## t = 0.23094, df = 1161.9, p-value = 0.8174
## alternative hypothesis: true difference in means between group History of smoking and group No is no ## 95 percent confidence interval:
##   -1.883164   2.385630
## sample estimates:
## mean in group History of smoking            mean in group No
##                          126.1260                    125.8748
```

```
. = ottr::check("tests/p6.R")
```

```
##
## All tests passed!
```

7. Use these results to interpret your p-value in the context of this question. Do you reject or fail to reject the null hypothesis?

Under the null hypothesis, we have 81.74% of chance of seeing a difference between our two sample is 0.2512. There we would fail to reject the null hypothesis and not conclude that there is a significance difference between the SBP of smokers vs non-smokers.

Repeat your analysis above without using the `t.test()` function.

8. [1 point] First calculate the test statistic by hand. Do not round and assign this value to t_stat.

```r
# this code gives you the number of smokers in the dataset
n_s <- nrow(nhanes %>% filter(hs == 'History of smoking'))
n_s
```

```
## [1] 619
```

```r
# this code gives you the number of non-smokers in the dataset
n_ns <- nrow(nhanes %>% filter(hs == 'No'))
n_ns
```

```
## [1] 559
```

```r
# calculate your test statistic. You can make more objects if you wish.
t_stat <- 0.2512/ sqrt((18.56617^2 / n_s) + (18.71515^2/ n_ns))
t_stat
```

```
## [1] 0.2309112
```

```r
. = ottr::check("tests/p8.R")
```

```
##
## All tests passed!
```

9. [1 point] Now compare your test statistic to a t-distribution with df = 558 and calculate the p-value. This is an approximation using the smaller of the two sample sizes - 1.

```r
p_value <- pt(0.2309112, df = 558, lower.tail = FALSE) *2
p_value
```

```
## [1] 0.8174684
```

```r
. = ottr::check("tests/p9.R")
```

```
##
## All tests passed!
```

10. [1 point] Finally, construct a 99% confidence interval for these data. Interpret the interval in the context of this question and decide whether or not to reject the null hypothesis.

```
CV <- qt(0.005, df = 558, lower.tail = FALSE)

SE <- sqrt((18.56617^2 / n_s) + (18.71515^2/ n_ns))

lowerbound <- 0.2512 - CV * SE
upperbound <- 0.2512 + CV * SE
conf_int <- c(lowerbound, upperbound)
conf_int
```

```
## [1] -2.560568  3.062968
```

Our 99% conifdence interval for mean difference of SBP in smoker and non-smokers is (-2.560568 3.062968)

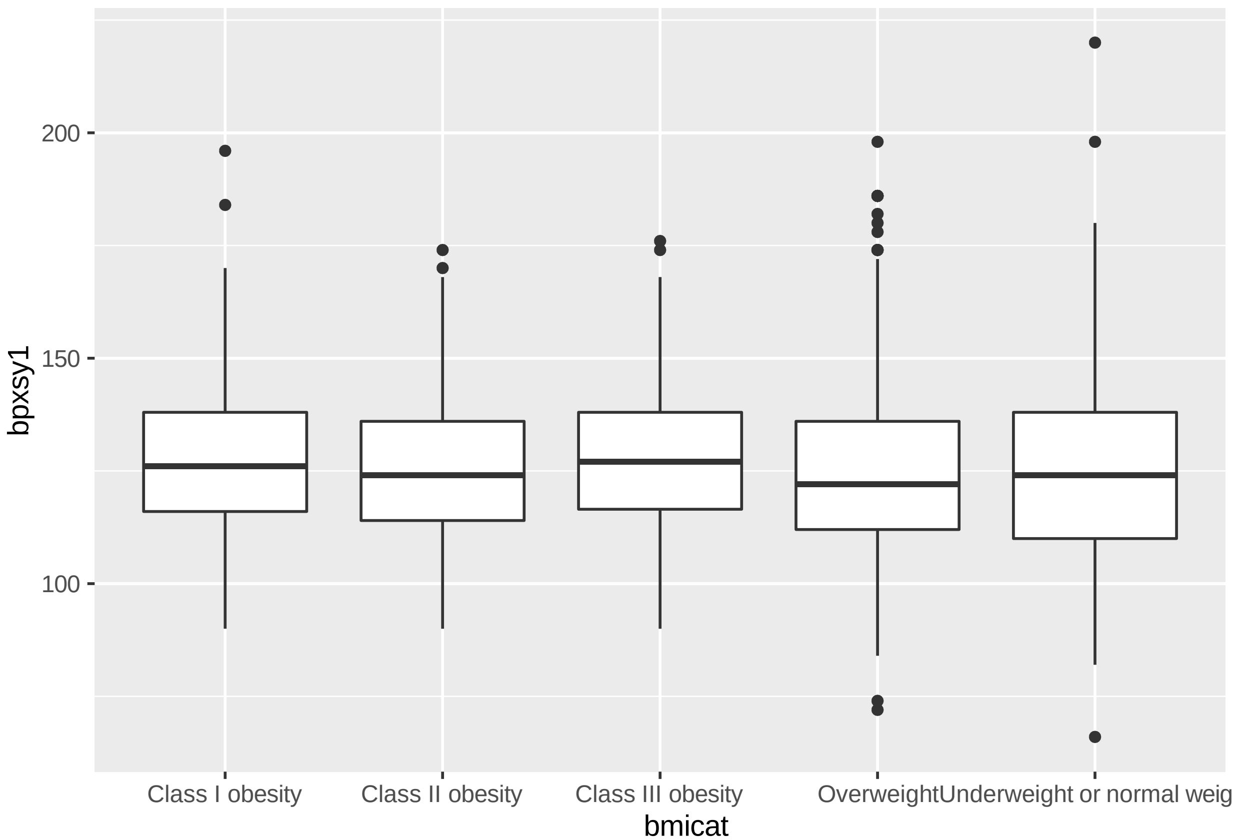
```
. = ottr::check("tests/p10.R")
```

```
##
## All tests passed!
```

Part 2: ANOVA

11. [1 point] We are interested in looking at the systolic blood pressure, bpxsy1, by BMI category, bmicat. Generate an appropriate box plot to visualize these data.

```
plot11 <- ggplot(nhanes, aes(x= bmicat, y = bpxsy1)) + geom_boxplot()
plot11
```



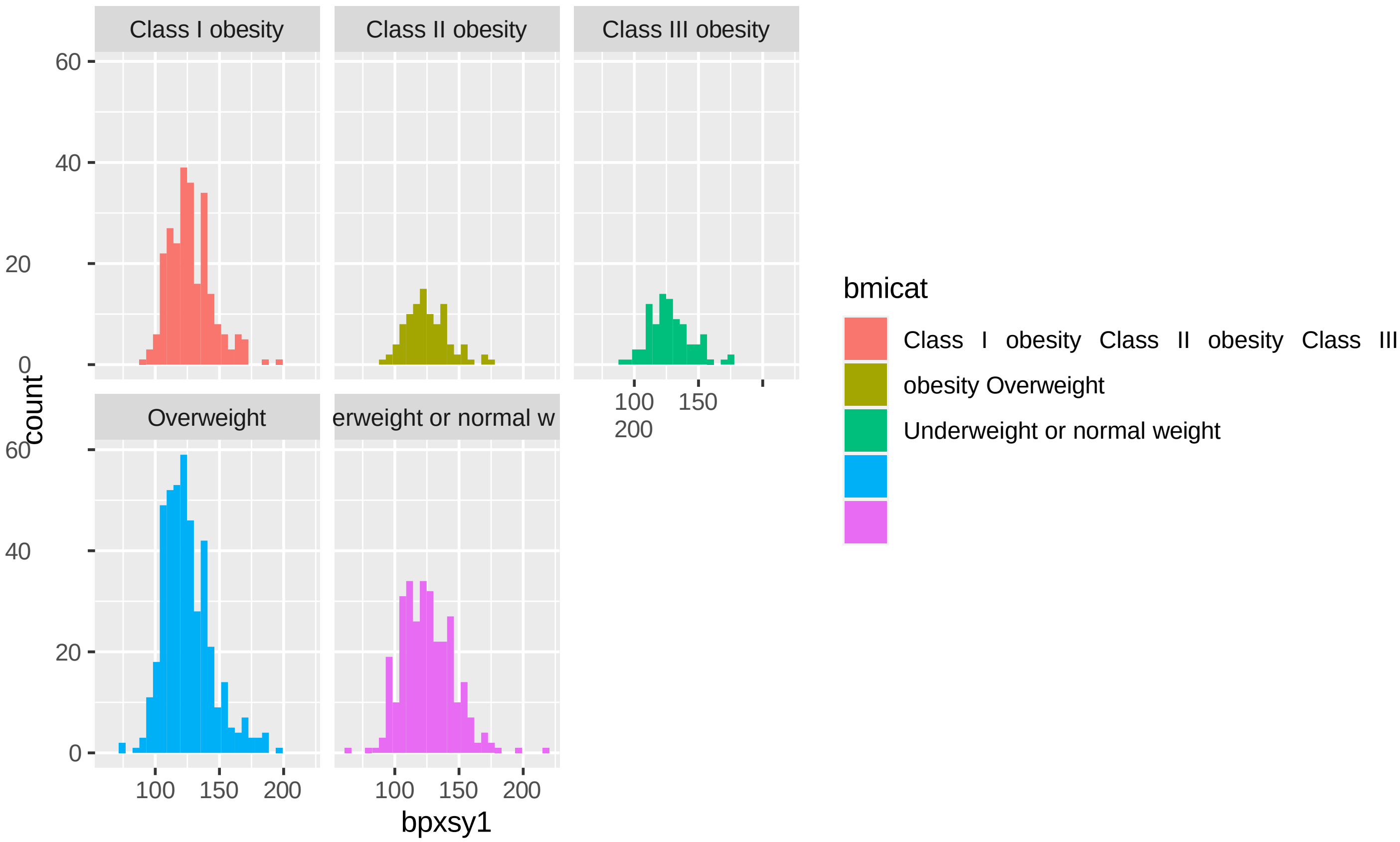

```
. = ottr::check("tests/p11.R")
```

```
##
## All tests passed!
```

12. [1 point] Now generate a set of faceted histograms that show the same data. It might be useful to assign a fill color to each category.

```
plot12 <- ggplot(nhanes, aes(x = bpxsy1)) + geom_histogram(aes(fill = bmicat))+
                 facet_wrap(~bmicat)
plot12
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```




```
. = ottr::check("tests/p12.R")
```

```
##
## All tests passed!
```

13. [1 point] Summarize the means and standard deviations of the outcome for each BMI category. Assign p13 to a dataframe with the mean systolic blood pressure assigned to mean_bp and the standard deviation assigned to sd_bp.

```
p13 <- nhanes %>% group_by(bmicat) %>% summarize(mean_bp = mean(bpxsy1),
                                                 sd_bp = sd(bpxsy1))
p13
```

```
## # A tibble: 5 x 3
##   bmicat                         mean_bp sd_bp
##   <chr>                            <dbl> <dbl>
## 1 Class I obesity                   128.  17.0
## 2 Class II obesity                  126.  16.9
## 3 Class III obesity                 128.  17.0
## 4 Overweight                        125.  19.0
## 5 Underweight or normal weight      125.  20.3
```

```
. = ottr::check("tests/p13.R")
```

```
## 
## All tests passed!
```

14. [1 point] Use an R function to test whether there is evidence to reject the null hypothesis of no difference between mean blood pressure by BMI category.

```
p14 <- aov(bpxsy1 ~ bmicat, data = nhanes)
tidy(p14) # tidy displays your output. It lives in the `broom` package
```

```
## # A tibble: 2 x 6
##   term         df   sumsq meansq statistic p.value
##   <chr>     <dbl>   <dbl>  <dbl>     <dbl>   <dbl>
## 1 bmicat        4   1651.   413.      1.19   0.314
## 2 Residuals  1173 406837.   347.     NA     NA
```

```
. = ottr::check("tests/p14.R")
```

```
## 
## All tests passed!
```

15. [1 point] Conduct a Tukey's HSD test using these data. What can you conclude assuming a standard error rate of 5%?

```r
p15 <- TukeyHSD(p14)
tidy(p15)
```

A tibble: 10 x 7

##	term	contrast	null.value	estimate	conf.low	conf.high	adj.p.value
##	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	bmicat	Class II obesity-C~	0	-2.09	-8.19	4.01	0.883
## 2	bmicat	Class III obesity-~	0	0.638	-5.61	6.89	0.999
## 3	bmicat	Overweight-Class I~	0	-2.60	-6.63	1.43	0.396
## 4	bmicat	Underweight or nor~	0	-2.18	-6.51	2.16	0.646
## 5	bmicat	Class III obesity-~	0	2.73	-4.74	10.2	0.856
## 6	bmicat	Overweight-Class I~	0	-0.510	-6.25	5.23	0.999
## 7	bmicat	Underweight or nor~	0	-0.0871	-6.04	5.87	1.00
## 8	bmicat	Overweight-Class I~	0	-3.24	-9.13	2.66	0.562
## 9	bmicat	Underweight or nor~	0	-2.81	-8.92	3.29	0.716
## 10	bmicat	Underweight or nor~	0	0.423	-3.38	4.22	0.998

Based on the Tukey's HSD test results and a standard error rate of 5%, we can conclude that there is statistically significant differences between the group means we have compared. Therefore, we fail to reject the null hypothesis of no difference between mean blood pressure by BMI category.

```r
. = ottr::check("tests/p15.R")
```

```
##
## All tests passed!
```