all_analysis

Anh Nguyen

2022-11-28

```
knitr::opts_knit$set(root.dir = getwd())
```

Introduction

Data Description

```
rm(list = ls())
source("utility/setup.r")
```

```
## Warning: package 'dplyr' was built under R version 4.2.2

## Warning: package 'ggpattern' was built under R version 4.2.2

## Warning: package 'tikzDevice' was built under R version 4.2.2

## Warning: package 'tidyverse' was built under R version 4.2.2

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v tibble  3.1.8     v purrr   0.3.5
## v tidyr   1.2.1     v stringr 1.4.1
## v readr   2.1.3     v forcats 0.5.2

## Warning: package 'tibble' was built under R version 4.2.2

## Warning: package 'tidyr' was built under R version 4.2.2

## Warning: package 'readr' was built under R version 4.2.2

## Warning: package 'purrr' was built under R version 4.2.2

## Warning: package 'stringr' was built under R version 4.2.2

## Warning: package 'forcats' was built under R version 4.2.2
```

```
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Warning: package 'mosaic' was built under R version 4.2.2
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2
##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.
```

```
## Warning: package 'table1' was built under R version 4.2.2
```

```
## Warning: package 'ggpubr' was built under R version 4.2.2
```

```
## Warning: package 'ggstar' was built under R version 4.2.2
```

```
## Warning: package 'lemon' was built under R version 4.2.2
```

```
## Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
```

```
head(included.data, 10)
```

```
##     PID site race age education gender married smoker drinker wgt_base month wgt_final tx height com
## 1  1003    1    2  53         1      1       1      0       0    106.9    15     106.8  0    158
## 2  1008    1    2  63         0      1       0      0       0    108.2    14     110.7  0    172
## 3  1012    1    2  18         0      1       0      0       0    127.5    13     130.1  0    169
## 4  1016    1    2  48         0      1       1      0       0    100.0    13      97.4  1    158
## 5  1017    1    2  56         1      0       1      0       0    140.8    13     143.8  0    181
## 6  1018    1    2  58         0      1       0      0       0     83.6    14      79.9  1    158
## 7  1021    1    2  53         0      1       1      0       0    100.3    13     102.7  0    159
## 8  1023    1    2  50         0      1       1      0       0    119.2    14     114.9  0    160
## 9  1025    1    2  40         1      1       0      0       1    125.7    14     120.9  1    170
## 10 1026    1    2  54         1      1       1      0       0    141.3    13     144.4  0    182
##                                    Race Gender     Marital Tobacco Alcohol   Education    Condition
## 1  African American (non-Hispanic black) Female     Married      No      No  > 12 years      Control
## 2  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 3  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 4  African American (non-Hispanic black) Female     Married      No      No <= 12 years Intervention
## 5  African American (non-Hispanic black)   Male     Married      No      No  > 12 years      Control
## 6  African American (non-Hispanic black) Female Not married      No      No <= 12 years Intervention
## 7  African American (non-Hispanic black) Female     Married      No      No <= 12 years      Control
## 8  African American (non-Hispanic black) Female     Married      No      No <= 12 years      Control
## 9  African American (non-Hispanic black) Female Not married      No     Yes  > 12 years Intervention
## 10 African American (non-Hispanic black) Female     Married      No      No  > 12 years      Control
```

Data Description (Limit to no more than 1 page, excluding tables and figures)

- Describe the data collection and study design.
- Choose and describe the primary and secondary outcomes (at least one continuous and one binary/categorical) and other relevant variables.
- Consider the relationship among the selected variables.
- Evaluate the potential distribution of your outcome variables, using graphical methods and summary statistics. If there is any missing data, they should be described in this section.
- Comment on normal and binomial/multinomial distribution assumptions.

Relationship - hypotheses:

- baseline weight: higher at female, african and hispanic american,
- weight loss:
  - positive correlation between months and weight loss
  - more weight loss are seen at: intervention group
  - no difference between sites

References:

https://cran.r-project.org/web/packages/table1/vignettes/table1-examples.html

Describe demographics: Summary statistics

Count the number of records for each categories and show percentages: e.g.: # of ppl of each gender, education, marital status, tobacco use, alcohol use (binary, use a same table with 3 columns), # of ppl of each site (5), race (4)

```
table1(~ Site + Gender + age + height + Race + comorbid + Education + Marital + Tobacco + Alcohol | Cond
```

```
## Warning in chisq.test(table(y, g)): Chi-squared approximation may be incorrect
```

```
## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	Control	Intervention	P-value
	(N=90)	(N=76)	
Clinical site			
Site 1	23 (25.6%)	15 (19.7%)	0.642
Site 2	23 (25.6%)	21 (27.6%)	
Site 3	19 (21.1%)	21 (27.6%)	
Site 4	13 (14.4%)	7 (9.2%)	
Site 5	12 (13.3%)	12 (15.8%)	
Gender			
Male	14 (15.6%)	12 (15.8%)	1

	Control	Intervention	P-value
Female	76 (84.4%)	64 (84.2%)	
Age (years)			
Mean (SD)	48.8 (11.1)	49.1 (11.2)	0.85
Median [Min, Max]	51.0 [18.0, 68.0]	51.0 [21.0, 68.0]	
Height (cm)			
Mean (SD)	165 (8.75)	164 (7.97)	0.937
Median [Min, Max]	164 [147, 187]	164 [149, 183]	
Race/ethnicity			
Asian	0 (0%)	2 (2.6%)	0.323
African American (non-Hispanic black)	59 (65.6%)	54 (71.1%)	
Hispanic/Latino	13 (14.4%)	9 (11.8%)	
Non-Hispanic white	18 (20.0%)	11 (14.5%)	
Number of obesity-related comorbid conditions			
Mean (SD)	1.09 (1.16)	1.29 (1.08)	0.251
Median [Min, Max]	1.00 [0, 5.00]	1.00 [0, 4.00]	
Education			
<= 12 years	33 (36.7%)	18 (23.7%)	0.102
> 12 years	57 (63.3%)	58 (76.3%)	
Marital status			
Not married	47 (52.2%)	45 (59.2%)	0.456
Married	43 (47.8%)	31 (40.8%)	
Current tobacco use			
No	74 (82.2%)	63 (82.9%)	1
Yes	16 (17.8%)	13 (17.1%)	
Current alcohol use			
No	63 (70.0%)	53 (69.7%)	1
Yes	27 (30.0%)	23 (30.3%)	

No significant differences in demographics between 2 groups.

Describe intervention outcomes: summary statisics and graphical methods

baseline wgt, final wgt, months histogram/bar chart/scatter , mean, sd

No significant differences in outcome variables between 2 groups

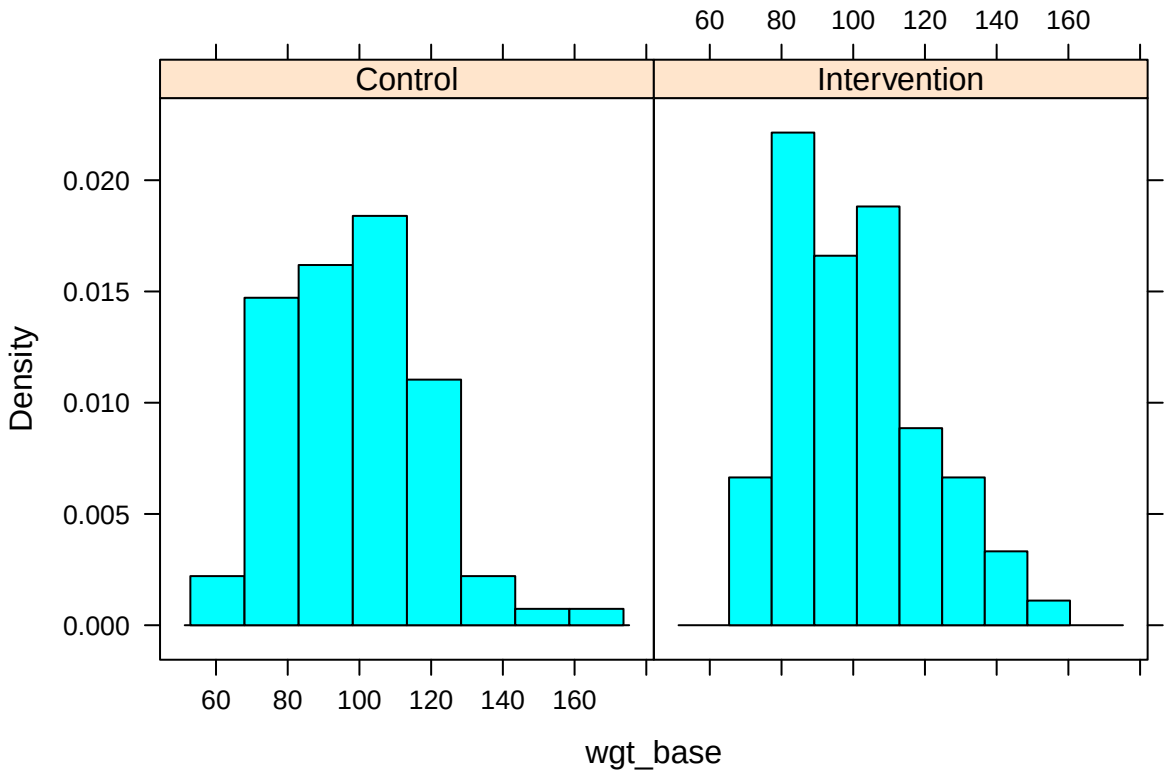
```
table1(~ wgt_base + wgt_final + month | Condition, data = included.data, topclass=table1.styles, overall
```

```
## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

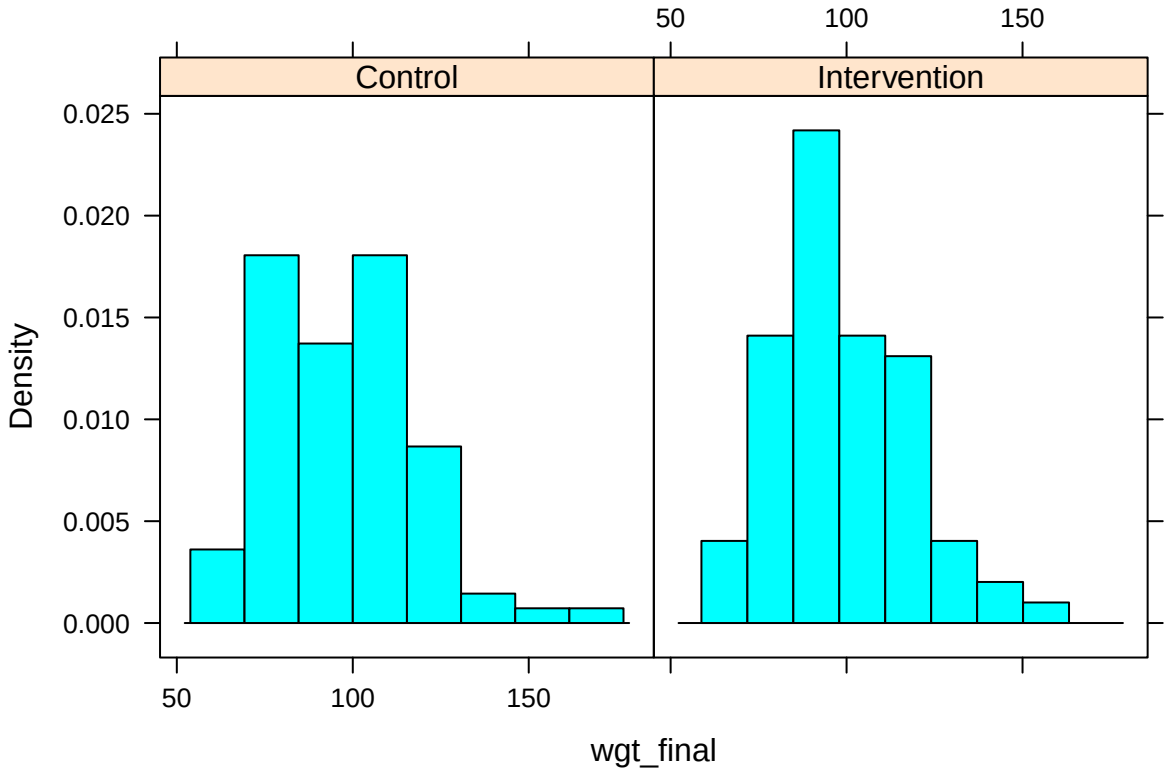
	Control	Intervention	P-value
	(N=90)	(N=76)	
Baseline weight (kg)			
Mean (SD)	99.2 (20.1)	101 (19.3)	0.536
Median [Min, Max]	98.5 [57.4, 163]	100 [68.2, 151]	
Final weight (kg)			
Mean (SD)	98.4 (20.7)	98.6 (19.1)	0.945
Median [Min, Max]	97.3 [54.8, 163]	96.7 [59.4, 151]	
Time since randomization to final weight (months)			

	Control	Intervention	P-value
Mean (SD)	11.8 (1.21)	11.7 (1.48)	0.517
Median [Min, Max]	11.0 [10.0, 15.0]	11.0 [10.0, 17.0]	

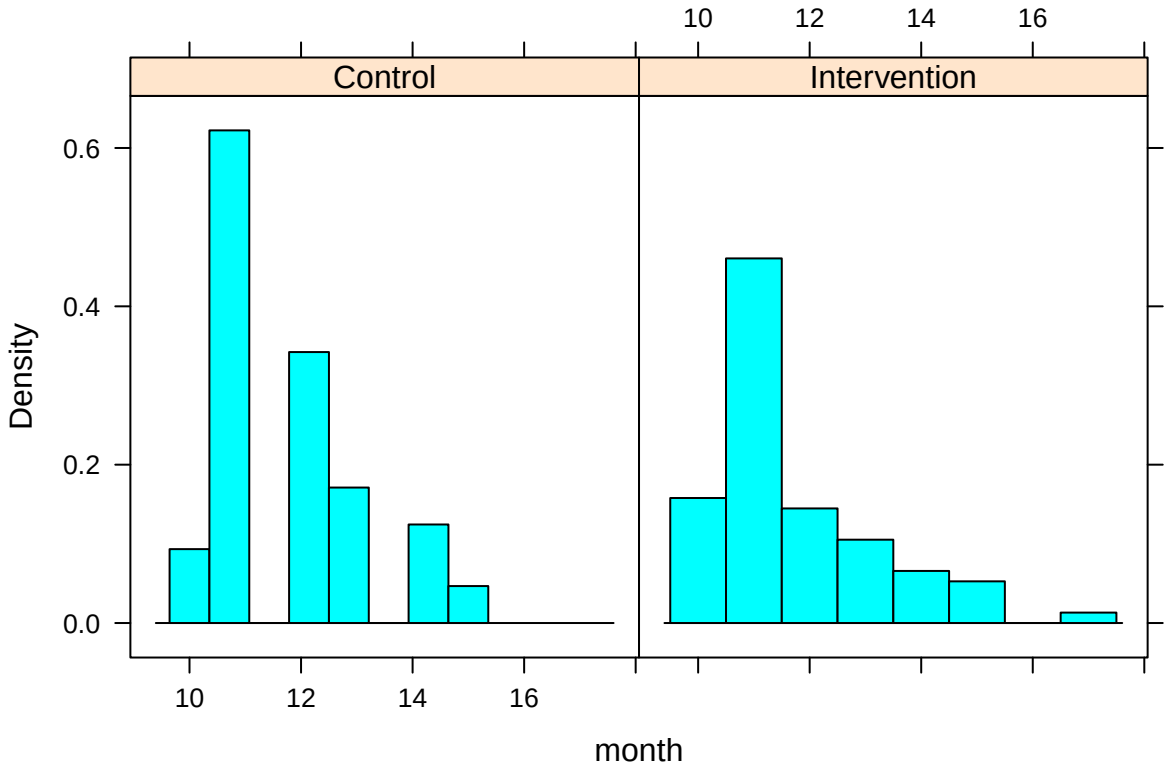
```
histogram(~wgt_base | Condition, data = included.data)
```



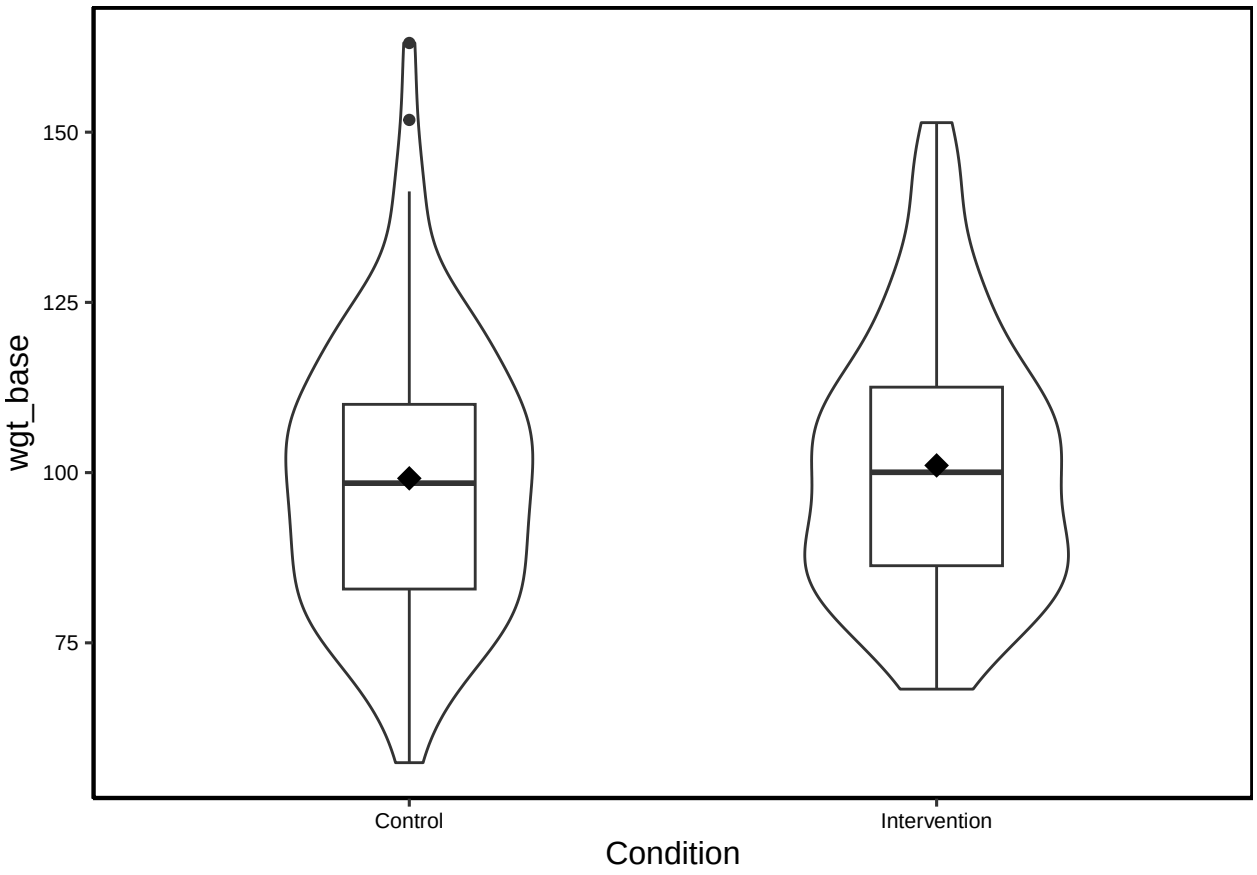

```
histogram(~wgt_final | Condition, data = included.data)
```




```
histogram(~month | Condition, data = included.data)
```




```
ggplot(data = included.data, mapping = aes(x = Condition, y = wgt_base)) +
  plt_theme +
  geom_violin(width = .5) +
  geom_boxplot(
    width = 0.25,
    outlier.size = 1.5) +
  stat_summary(
  fun = "mean",
  geom = "point",
  size = 4,
  shape = 18
  # https://ggplot2.tidyverse.org/articles/ggplot2-specs.html#sec:shape-spec
  )
```



Point Estimates and Confidence Intervals

```
rm(list = ls())
source("utility/setup.r")
head(data, 10)
```

```
##     PID site race age education gender married smoker drinker wgt_base month wgt_final tx height com
## 1  1003    1    2  53         1      1       1      0       0    106.9    15     106.8  0    158
## 2  1004    1    2  51         0      1       1      0       0     83.4    NA        NA  1    162
## 3  1008    1    2  63         0      1       0      0       0    108.2    14     110.7  0    172
## 4  1012    1    2  18         0      1       0      0       0    127.5    13     130.1  0    169
## 5  1013    1    2  34         0      1       0      0       1    101.0    NA        NA  1    168
## 6  1015    1    2  42         1      0       1      1       0    123.2    NA        NA  0    176
## 7  1016    1    2  48         0      1       1      0       0    100.0    13      97.4  1    158
## 8  1017    1    2  56         1      0       1      0       0    140.8    13     143.8  0    181
## 9  1018    1    2  58         0      1       0      0       0     83.6    14      79.9  1    158
## 10 1021    1    2  53         0      1       1      0       0    100.3    13     102.7  0    159
##                                    Race Gender     Marital Tobacco Alcohol   Education    Condition
## 1  African American (non-Hispanic black) Female     Married      No      No  > 12 years      Control
## 2  African American (non-Hispanic black) Female     Married      No      No <= 12 years Intervention
## 3  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 4  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 5  African American (non-Hispanic black) Female Not married      No     Yes <= 12 years Intervention
```

```
## 6  African American (non-Hispanic black)   Male     Married     Yes      No  > 12 years      Control
## 7  African American (non-Hispanic black) Female     Married      No      No <= 12 years Intervention
## 8  African American (non-Hispanic black)   Male     Married      No      No  > 12 years      Control
## 9  African American (non-Hispanic black) Female Not married      No      No <= 12 years Intervention
## 10 African American (non-Hispanic black) Female     Married      No      No <= 12 years      Control
```

Point Estimates and Confidence Intervals (Limit to no more than 1 page)

- Obtain point estimates for population parameters of interest.
- Obtain confidence intervals for those parameters, using both parametric (e.g., asymptotic) and nonparametric (e.g., bootstrap) methods. Indicate which interval you used. Interpret your results.

Hypothesis Testing

```
rm(list = ls())
source("utility/setup.r")
head(data, 10)
```

```
##     PID site race age education gender married smoker drinker wgt_base month wgt_final tx height com
## 1  1003    1    2  53         1      1       1      0       0    106.9    15     106.8  0    158
## 2  1004    1    2  51         0      1       1      0       0     83.4    NA        NA  1    162
## 3  1008    1    2  63         0      1       0      0       0    108.2    14     110.7  0    172
## 4  1012    1    2  18         0      1       0      0       0    127.5    13     130.1  0    169
## 5  1013    1    2  34         0      1       0      0       1    101.0    NA        NA  1    168
## 6  1015    1    2  42         1      0       1      1       0    123.2    NA        NA  0    176
## 7  1016    1    2  48         0      1       1      0       0    100.0    13      97.4  1    158
## 8  1017    1    2  56         1      0       1      0       0    140.8    13     143.8  0    181
## 9  1018    1    2  58         0      1       0      0       0     83.6    14      79.9  1    158
## 10 1021    1    2  53         0      1       1      0       0    100.3    13     102.7  0    159
##                                    Race Gender     Marital Tobacco Alcohol   Education    Condition
## 1  African American (non-Hispanic black) Female     Married      No      No  > 12 years      Control
## 2  African American (non-Hispanic black) Female     Married      No      No <= 12 years Intervention
## 3  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 4  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 5  African American (non-Hispanic black) Female Not married      No     Yes <= 12 years Intervention
## 6  African American (non-Hispanic black)   Male     Married     Yes      No  > 12 years      Control
## 7  African American (non-Hispanic black) Female     Married      No      No <= 12 years Intervention
## 8  African American (non-Hispanic black)   Male     Married      No      No  > 12 years      Control
## 9  African American (non-Hispanic black) Female Not married      No      No <= 12 years Intervention
## 10 African American (non-Hispanic black) Female     Married      No      No <= 12 years      Control
```

Hypothesis Testing (Limit to no more than 1 page)

- Translate the scientific questions of your interest into null and alternative hypotheses.
- Choose appropriate approaches for hypothesis testing. Justify your choice of test and discuss potential issues that you might have in making your choice.
- Carry out the test and interpret your results. Check assumptions if applicable.

Linear Model

```
rm(list = ls())
source("utility/setup.r")
head(data, 10)
```

```
##     PID site race age education gender married smoker drinker wgt_base month wgt_final tx height com
## 1  1003    1    2  53         1      1       1      0       0    106.9    15     106.8  0    158
## 2  1004    1    2  51         0      1       1      0       0     83.4    NA        NA  1    162
## 3  1008    1    2  63         0      1       0      0       0    108.2    14     110.7  0    172
## 4  1012    1    2  18         0      1       0      0       0    127.5    13     130.1  0    169
## 5  1013    1    2  34         0      1       0      0       1    101.0    NA        NA  1    168
## 6  1015    1    2  42         1      0       1      1       0    123.2    NA        NA  0    176
## 7  1016    1    2  48         0      1       1      0       0    100.0    13      97.4  1    158
## 8  1017    1    2  56         1      0       1      0       0    140.8    13     143.8  0    181
## 9  1018    1    2  58         0      1       0      0       0     83.6    14      79.9  1    158
## 10 1021    1    2  53         0      1       1      0       0    100.3    13     102.7  0    159
##                                      Race Gender     Marital Tobacco Alcohol   Education    Condition
## 1  African American (non-Hispanic black) Female     Married      No      No  > 12 years      Control
## 2  African American (non-Hispanic black) Female     Married      No      No <= 12 years Intervention
## 3  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 4  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 5  African American (non-Hispanic black) Female Not married      No     Yes <= 12 years Intervention
## 6  African American (non-Hispanic black)   Male     Married     Yes      No  > 12 years      Control
## 7  African American (non-Hispanic black) Female     Married      No      No <= 12 years Intervention
## 8  African American (non-Hispanic black)   Male     Married      No      No  > 12 years      Control
## 9  African American (non-Hispanic black) Female Not married      No      No <= 12 years Intervention
## 10 African American (non-Hispanic black) Female     Married      No      No <= 12 years      Control
```

##Linear Model (Limit to no more than 2 pages, excluding figures and tables)

- Construct linear model(s) to answer ONE OR MORE of the research questions of interest.
- Justify the procedures or reasons for model and variable selection.
- Fit the model and interpret parameter estimates and confidence intervals if applicable.
- Do model diagnosis using graphical and statistical tools and discuss appropriate remedy procedures if needed.

Categorical Data Analysis

```
rm(list = ls())
source("utility/setup.r")
head(data, 10)
```

```
##     PID site race age education gender married smoker drinker wgt_base month wgt_final tx height com
## 1  1003    1    2  53         1      1       1      0       0    106.9    15     106.8  0    158
## 2  1004    1    2  51         0      1       1      0       0     83.4    NA        NA  1    162
## 3  1008    1    2  63         0      1       0      0       0    108.2    14     110.7  0    172
## 4  1012    1    2  18         0      1       0      0       0    127.5    13     130.1  0    169
```

```
## 5  1013    1    2  34         0      1       0      0       1    101.0    NA        NA  1    168
## 6  1015    1    2  42         1      0       1      1       0    123.2    NA        NA  0    176
## 7  1016    1    2  48         0      1       1      0       0    100.0    13      97.4  1    158
## 8  1017    1    2  56         1      0       1      0       0    140.8    13     143.8  0    181
## 9  1018    1    2  58         0      1       0      0       0     83.6    14      79.9  1    158
## 10 1021    1    2  53         0      1       1      0       0    100.3    13     102.7  0    159
##                                      Race Gender     Marital Tobacco Alcohol   Education    Condition
## 1  African American (non-Hispanic black) Female     Married      No      No  > 12 years      Control
## 2  African American (non-Hispanic black) Female     Married      No      No <= 12 years Intervention
## 3  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 4  African American (non-Hispanic black) Female Not married      No      No <= 12 years      Control
## 5  African American (non-Hispanic black) Female Not married      No     Yes <= 12 years Intervention
## 6  African American (non-Hispanic black)   Male     Married     Yes      No  > 12 years      Control
## 7  African American (non-Hispanic black) Female     Married      No      No <= 12 years Intervention
## 8  African American (non-Hispanic black)   Male     Married      No      No  > 12 years      Control
## 9  African American (non-Hispanic black) Female Not married      No      No <= 12 years Intervention
## 10 African American (non-Hispanic black) Female     Married      No      No <= 12 years      Control
```

Categorical Data Analysis (Limit to no more than 2 pages, excluding figures and tables)

• Perform simple association test and interpret the results. • Construct a simple generalized linear regression model

Future Study Planning

Conclusion and Discussion