Exercise 4

RK

2025-09-29

This is a report of the NHANES dataset for Exercise 4 in Week 4's assessment.

About NHANES

The National Health and Nutrition Examination Survey is a national survey that measures the health and nutrition of adults and children in the United States. This report will focus on key aspects of a cleaned version of the NHANES dataset that has been used in previous week's assessments.

Sample size by wave and gender

This table displays the sample size in 2015, 2016, and 2017 in raw counts, and by the number of males in each wave. In all three waves, males made up just under half of the sample size.

```
# sample size by wave and gender

sample_size_wave <- nhanes_data %>%
  group_by(wave_id) %>%
  summarise(
    sample_size = n(),
    males_n = sum(gender == "Male"),
    males_pct = percent(males_n / sample_size)) %>%
  ungroup() %>%
  select(Wave = wave_id, 'Sample size' = sample_size, 'Males (n)' = males_n, 'Males (%)' = males_pct, )

# create table using kable function
sample_size_wave %>%
  kable(
    booktabs = TRUE,
    caption = "NHANES sample size and male composition by wave") %>%
  kable_styling(latex_options = c("hold_position", "striped", "threeparttable", full_width = FALSE))
```

Table 1: NHANES sample size and male composition by wave

Wave	Sample size	Males (n)	Males (%)
2015	10175	5003	49%
2016	9971	4892	49%
2017	9254	4557	49%

Sample size by ethnicity

This table displays the counts and percentages of each ethnicity in 2015, 2016, and 2017. In all three waves, the greatest proportion of participants were white.

```
# sample size by ethnicity
ethnicity_waves <-nhanes_data %>%
  count(wave_id, ethnicity_2, name = "n") %>%
  group_by(wave_id) %>%
  mutate(pct = n / sum(n),
         cell = sprintf("%d (%.1f%%)", n, 100 * pct)) %>%
  select(wave_id, ethnicity_2, cell) %>%
  pivot_wider(names_from = ethnicity_2, values_from = cell) %>%
  arrange(wave_id) %>%
  rename(Wave = wave_id)

#create table
ethnicity_waves %>%
  kable(caption = "Ethnicity by Wave: n (percent)",
        booktabs = TRUE) %>%
  kable_styling(latex_options = c("hold_position", "striped")) %>%
  column_spec(1, bold = TRUE)
```

Table 2: Ethnicity by Wave: n (percent)

Wave	Asian	Black	Mexican	Other	Other Hispanic	White
2015	1074 (10.6%)	2267 (22.3%)	1730 (17.0%)	470 (4.6%)	960 (9.4%)	3674 (36.1%)
2016	1042 (10.5%)	2129 (21.4%)	1921 (19.3%)	505 (5.1%)	1308 (13.1%)	3066 (30.7%)
2017	1168 (12.6%)	2115 (22.9%)	1367 (14.8%)	634 (6.9%)	820 (8.9%)	3150 (34.0%)

Summary statistics by wave

This table highlights some key summary statistics of the average systolic blood pressure per wave. The third wave in 2017 saw the highest numbers for all reported statistics.

```
# calculate average spb

sbp_cols <- paste0("systolic_bp_", 1:4)

nhanes_data <- nhanes_data %>%
  mutate(
    avg_sbp = {
      m <- rowMeans(across(all_of(sbp_cols)), na.rm = TRUE)
    }
  )

# show summary statistics

sbp_summary <- nhanes_data %>%
  group_by(wave_id) %>%
```

```r
  summarise(
    Min    = min(avg_sbp, na.rm = TRUE),
    Q2     = median(avg_sbp, na.rm = TRUE),
    Mean   = mean(avg_sbp, na.rm = TRUE),
    SD     = sd(avg_sbp, na.rm = TRUE),
    Max    = max(avg_sbp, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(across(c(Min, Q2, Mean, SD, Max), ~ round(.x, 1))) %>%
  arrange(wave_id) %>%
  rename(Wave = wave_id, Median = Q2)

sbp_summary %>%
  kable(booktabs = TRUE, caption = "Average systolic blood pressure by wave (mmHg)") %>%
  kable_styling(latex_options = c("hold_position", "striped")) %>%
  column_spec(1, bold = TRUE)
```

Table 3: Average systolic blood pressure by wave (mmHg)

Wave	Min	Median	Mean	SD	Max
2015	64.7	115.3	118.3	18.1	228.7
2016	74.0	117.3	120.4	18.4	231.3
2017	72.7	118.0	121.7	20.3	234.0