STA540 case1

Feilin Feng

```r
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4      v readr     2.1.5
v forcats   1.0.1      v stringr   1.5.2
v ggplot2   4.0.0      v tibble    3.3.0
v lubridate 1.9.4      v tidyr     1.3.1
v purrr     1.1.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
library(lubridate)
library(knitr)
library(kableExtra)
```

```
Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

    group_rows
```

```r
library(emmeans)
```

```
Welcome to emmeans.
Caution: You lose important information if you filter this package's results.
See '? untidy'
```

```
library(gt)
```

```
Warning: package 'gt' was built under R version 4.4.3
```

```
df_raw = read.csv("/Users/personalarea/Desktop/STA540/Case 1/CTN_FINAL.csv")
```

Data Preprocess

```
df = df_raw %>%
  filter(SITE != "Yahoo", WAVE != 3, PO_FLAG == "Include")

nrow(df)
```

```
[1] 254
```

Reproduce Table 1

- Q3_1: How old are you?

- Q4_1: What sex were you assigned at birth, on your original birth certificate? (1=Male; 2=Female)

- Q5_1: Are you Hispanic and/or Latino? (1=Yes; 2=No)

- Q5_3: Do you self-identify as... (25= American Indian or Alaska Native; 26= Asian; 24= Black or African American; 27= Native Hawaiian or Pacific Islander; 23= White; 28= Other, please specify) Q5_3_28_TEXT

- Q6_2: Have you ever taken PrEP? (Also known as Pre-Exposure Prophylactis)

- Q11_2: How many male sexual partners have you found in 90 days?

- Q11_3: How often do you use condoms? (1= Never; 2= Sometimes; 3 = About half the time; 4 = Most of the time; 5 = Always)

- Q11_4: Have you had condomless receptive anal sex in the past 90 days? (1=Yes;2=No)

- Q11_5: Have you ever been tested for HIV in your lifetime?

- Q11_7: If you have not been tested for HIV, which one of the following would you say is the MAIN reason why you have not been tested? (1= It's unlikely you've been exposed to HIV; 2 = You are afraid to find out if you are HIV positive; 3 = You didn't want to think about HIV or about being HIV positive; 4 = You were worried your name would be reported to the government if you tested positive; 5 = You don't like needles; 6 = You don't trust the results to be confidential; 7 = You didn't know where to get tested; 8= Some other reason. Please specify.)

```
df %>%
  count(Q6_2) %>%
  mutate(
    percent = n / sum(n) * 100
  )
```

```
  Q6_2   n    percent
1    1  22   8.661417
2    3 232  91.338583
```

```
df %>%
  count(Q5_1) %>%
  mutate(
    percent = n / sum(n) * 100
  )
```

```
  Q5_1   n  percent
1    1  66 25.98425
2    2 188 74.01575
```

```
df %>%
  count(Q5_3) %>%
  mutate(
    percent = n / sum(n) * 100
  )
```

```
    Q5_3   n     percent
1           4   1.5748031
2     23  28  11.0236220
3     24 196  77.1653543
4  24,23   5   1.9685039
5  24,28   1   0.3937008
```

```
6            25   1  0.3937008
7         25,24   2  0.7874016
8      25,24,23   1  0.3937008
9  25,24,27,23   1  0.3937008
10        26,24   1  0.3937008
11           28  14  5.5118110
```

```r
quantile(df$Q11_2, probs = c(0.25, 0.5, 0.75), na.rm = TRUE)
```

```
25% 50% 75%
  3   4   6
```

```r
quantile(df$Q3_1, probs = c(0.25, 0.5, 0.75), na.rm = TRUE)
```

```
25% 50% 75%
 21  25  27
```

```r
df %>%
  count(Q11_3) %>%
  mutate(
    percent = n / sum(n) * 100
  )
```

```
  Q11_3   n   percent
1     1  36 14.173228
2     2 108 42.519685
3     3  37 14.566929
4     4  68 26.771654
5     5   5  1.968504
```

```r
df %>%
  count(Q11_4) %>%
  mutate(
    percent = n / sum(n) * 100
  )
```

```
  Q11_4   n  percent
1     1 210 82.67717
2     2  44 17.32283
```

```r
df %>%
  count(Q11_5) %>%
  mutate(
    percent = n / sum(n) * 100
  )
```

```
  Q11_5   n  percent
1     1 191 75.19685
2     2  63 24.80315
```

```r
df %>%
  count(Q11_7) %>%
  mutate(
    percent = n / sum(n) * 100
  )
```

```
  Q11_7   n   percent
1     1   8  3.149606
2     2  26 10.236220
3     3   8  3.149606
4     4   3  1.181102
5     5   5  1.968504
6     6   3  1.181102
7     7   7  2.755906
8     8   4  1.574803
9  9999 190 74.803150
```

```r
# help function for quantile calculation
fmt_median_iqr = function(x) {
  qs = quantile(x, probs = c(0.25, 0.5, 0.75), na.rm = TRUE, type = 7)
  sprintf("%d (%d-%d)", as.integer(round(qs[2])), as.integer(round(qs[1])), as.integer(round
}

# ensures denominator excludes missing by default for categorical vars
count_pct = function(dat, var) {
  v = dat[[var]]
  tab = table(v, useNA = "no")
  denom = sum(tab)
  tibble(
    level = names(tab),
    n = as.integer(tab),
```

```r
    value = sprintf("%d (%.1f)", n, round(100 * n / denom, 1))
  )
}

# age
age_row = tibble(
  Characteristic = "Age in years, median (IQR)",
  Value = fmt_median_iqr(df$Q3_1)
)

# ethnicity
eth_tab = df %>%
  mutate(eth = case_when(
    Q5_1 == 1 ~ "Hispanic/Latinx",
    Q5_1 == 2 ~ "Not Hispanic/Latinx",
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(eth)) %>%
  count(eth) %>%
  mutate(denom = sum(n),
         Value = sprintf("%d (%.0f)", n, round(100 * n / denom, 0))) %>%  # paper shows integ
  select(eth, Value)

eth_rows = tibble(
  Characteristic = c("Ethnicity, n (%)", paste0("  ", eth_tab$eth)),
  Value = eth_tab$Value[eth_tab$eth == "Hispanic/Latinx"]
)


# race
race_map = c(
  '25' = "American Indian or Alaskan Native",
  '24' = "Black or African American",
  '23' = "White",
  '28' = "Other",
  '26' = "Asian",
  '27' = "Native Hawaiian or Pacific Islander"
)

race_tab = df %>%
  mutate(Q5_3_str = trimws(as.character(Q5_3))) %>%
  mutate(race = case_when(
```

```r
    grepl(",", Q5_3_str) ~ "Multiracial",
    Q5_3_str == "" | is.na(Q5_3_str) ~ NA_character_,
    Q5_3_str %in% names(race_map) ~ unname(race_map[Q5_3_str]),
    TRUE ~ "Other"
  )) %>%
  filter(!is.na(race)) %>%
  count(race) %>%
  mutate(denom = sum(n),
         Value = sprintf("%d (%.1f)", n, round(100 * n / denom, 1))) %>%

  arrange(match(race, c("American Indian or Alaskan Native",
                        "Black or African American",
                        "White",
                        "Other",
                        "Multiracial")))

race_rows = tibble(
  Characteristic = c("Race, n (%)", paste0("  ", race_tab$race)),
  Value = c("", race_tab$Value)
)


# history of PrEP uptake
prep_tab = df %>%
  mutate(prep_hist = case_when(
    Q6_2 == 3 ~ "Never taken PrEP",
    Q6_2 == 1 ~ "In the past 6 months",
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(prep_hist)) %>%
  count(prep_hist) %>%
  mutate(denom = sum(n),
         Value = sprintf("%d (%.1f)", n, round(100 * n / denom, 1))) %>%
  arrange(match(prep_hist, c("Never taken PrEP", "In the past 6 months")))

prep_rows = tibble(
  Characteristic = c("History of PrEP uptake, n (%)", paste0("  ", prep_tab$prep_hist)),
  Value = c("", prep_tab$Value)
)

# number of male sex partners in past 90 days
partners_row = tibble(
```

```r
  Characteristic = "Number of male sex partners in the past 90 days, median (IQR)",
  Value = fmt_median_iqr(df$Q11_2)
)

# condom use
condom_map = c(
  '1' = "Never",
  '2' = "Sometimes",
  '3' = "About half the time",
  '4' = "Most of the time",
  '5' = "Always"
)

condom_tab = df %>%
  mutate(condom = case_when(
    Q11_3 %in% as.integer(names(condom_map)) ~ unname(condom_map[as.character(Q11_3)]),
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(condom)) %>%
  count(condom) %>%
  mutate(denom = sum(n),
         Value = sprintf("%d (%.1f)", n, round(100 * n / denom, 1))) %>%
  arrange(match(condom, condom_map))

condom_rows = tibble(
  Characteristic = c("Condom use, n (%)", paste0("  ", condom_tab$condom)),
  Value = c("", condom_tab$Value)
)

# condomless receptive anal sex in past 90 days
cras_tab = df %>%
  mutate(cras = case_when(
    Q11_4 == 1 ~ "Yes",
    Q11_4 == 2 ~ "No",
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(cras)) %>%
  count(cras) %>%
  mutate(denom = sum(n),
         Value = sprintf("%d (%.1f)", n, round(100 * n / denom, 1))) %>%
  arrange(match(cras, c("Yes", "No")))
```

```r
cras_row = tibble(
  Characteristic = "Condomless receptive anal sex in the past 90 days, n (%)",
  Value = cras_tab$Value[cras_tab$cras == "Yes"]
)

# ever tested for HIV
ever_tab = df %>%
  mutate(ever = case_when(
    Q11_5 == 1 ~ "Yes",
    Q11_5 == 2 ~ "No",
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(ever)) %>%
  count(ever) %>%
  mutate(denom = sum(n),
         Value = sprintf("%d (%.1f)", n, round(100 * n / denom, 1))) %>%
  arrange(match(ever, c("Yes", "No")))

ever_row = tibble(
  Characteristic = "Ever tested for HIV during lifetime, n (%)",
  Value = ever_tab$Value[ever_tab$ever == "Yes"]
)

not_tested_row = tibble(
  Characteristic = "If not tested for HIV, n (%)",
  Value = ever_tab$Value[ever_tab$ever == "No"]
)

# months since last HIV test among those ever tested
df_time = df %>%
  filter(Q11_5 == 1) %>%
  mutate(
    last_hiv_test_date = as.Date(Q11_6_NUMERIC),
    baseline_date = as.Date(DATE_COMPLETED_BASELINE),
    months_diff_float = time_length(interval(last_hiv_test_date, baseline_date), unit = "mont
    months_since_last_hiv_test = round(months_diff_float)
  )

months_row = tibble(
  Characteristic = "Months since last HIV test",
  Value = fmt_median_iqr(df_time$months_since_last_hiv_test)
)
```

```r
# main reason for not getting tested among those not tested
reason_map = c(
  '1' = "Unlikely to be exposed to HIV",
  '2' = "Afraid of testing HIV-positive",
  '3' = "Did not want to think about HIV/HIV-positive",
  '4' = "Worried about names being reported if positive",
  '5' = "Dislike for needles",
  '6' = "Unable to trust that the results will be confidential",
  '7' = "Unaware of where to get tested",
  '8' = "Other reasons"
)

reasons_tab = df %>%
  filter(Q11_5 == 2) %>%
  mutate(reason = case_when(
    Q11_7 %in% 1:8 ~ unname(reason_map[as.character(Q11_7)]),
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(reason)) %>%
  count(reason) %>%
  mutate(denom = sum(n),
         Value = sprintf("%d (%.1f)", n, round(100 * n / denom, 1))) %>%
  arrange(match(reason, reason_map))

reason_rows = tibble(
  Characteristic = c("Main reasons cited by the 63 participants for not getting tested, n (%)
                     paste0("  ", reasons_tab$reason)),
  Value = c("", reasons_tab$Value)
)


## form table 1
table1_out = bind_rows(
  age_row,
  eth_rows,
  race_rows,
  prep_rows,
  partners_row,
  condom_rows,
  cras_row,
  ever_row,
  months_row,
```

```
    not_tested_row,
    reason_rows
)


kable(table1_out)
```

Characteristic	Value
Age in years, median (IQR)	25 (21-27)
Ethnicity, n (%)	66 (26)
Hispanic/Latinx	66 (26)
Not Hispanic/Latinx	66 (26)
Race, n (%)	
American Indian or Alaskan Native	1 (0.4)
Black or African American	196 (78.4)
White	28 (11.2)
Other	14 (5.6)
Multiracial	11 (4.4)
History of PrEP uptake, n (%)	
Never taken PrEP	232 (91.3)
In the past 6 months	22 (8.7)
Number of male sex partners in the past 90 days, median (IQR)	4 (3-6)
Condom use, n (%)	
Never	36 (14.2)
Sometimes	108 (42.5)
About half the time	37 (14.6)
Most of the time	68 (26.8)
Always	5 (2.0)
Condomless receptive anal sex in the past 90 days, n (%)	210 (82.7)
Ever tested for HIV during lifetime, n (%)	191 (75.2)
Months since last HIV test	11 (6-21)
If not tested for HIV, n (%)	63 (24.8)
Main reasons cited by the 63 participants for not getting tested, n (%)	
Unlikely to be exposed to HIV	8 (12.7)
Afraid of testing HIV-positive	26 (41.3)
Did not want to think about HIV/HIV-positive	8 (12.7)
Worried about names being reported if positive	3 (4.8)
Dislike for needles	5 (7.9)
Unable to trust that the results will be confidential	3 (4.8)
Unaware of where to get tested	7 (11.1)
Other reasons	3 (4.8)

```
tab_md1 = knitr::kable(table1_out, format = "pipe")
writeLines(c("### Replicated Table 1", "", tab_md1), "README.md")
```

The primary analysis, including fitting the poisson model, estimating the cell-specific rates, and testing the contrasts

```
df_long = df %>%
  select(WAVE, SITE_TYPE, SITE, ORA_WITHIN60_YESNO) %>%
  mutate(WAVE = case_when(
    WAVE == 1 ~ 1,
    WAVE == 4 ~ 1,
    WAVE == 2 ~ 2
  )) %>%
  mutate(
    WAVE_TIME = case_when(
      WAVE == 1 ~ 70,
      WAVE == 2 ~ 38
    )
  )
```

```
df_primary = df_long %>%
  group_by(WAVE, SITE) %>%
  summarise(
    ORD = sum(ORA_WITHIN60_YESNO == "Yes", na.rm = TRUE),
    WAVE_TIME = first(WAVE_TIME),
    SITE_TYPE = first(SITE_TYPE),
    .groups = "drop"
  ) %>%
  mutate(
    ORD_RATE = ORD / WAVE_TIME
  )

df_primary = df_primary %>%
  mutate(
    WAVE = factor(WAVE),
    SITE_TYPE = factor(SITE_TYPE)
  )
```

```
bing_row = tibble(
  WAVE = factor(2, levels = levels(df_primary$WAVE)),
  SITE_TYPE = factor("Info Sites", levels = levels(df_primary$SITE_TYPE)),
  SITE = "Bing",
  ORD = 0,
  WAVE_TIME = 38,
  ORD_RATE = 0
)
df_primary = bind_rows(df_primary, bing_row)
```

```
tab_cell = knitr::kable(df_primary, format = "pipe", digits = 3)
writeLines(c("### Replicated cell-specific rate", "", tab_cell), "Replicated cell-specific ra
```

```
fit_primary = glm(ORD ~ WAVE * SITE_TYPE + offset(log(WAVE_TIME)), family = poisson(link = "l

emm_rate = emmeans(
  fit_primary,
  ~ SITE_TYPE | WAVE,
  type = "response",
  at = list(WAVE_TIME = 1)
)
as.data.frame(emm_rate)
```

```
WAVE = 1:
 SITE_TYPE         rate         SE  df asymp.LCL asymp.UCL
 Dating Apps   0.128571 0.04285714 Inf 0.0668976         0
 Info Sites    0.242857 0.05890151 Inf 0.1509747         0
 Social Media  0.185714 0.05150788 Inf 0.1078362         0

WAVE = 2:
 SITE_TYPE         rate         SE  df asymp.LCL asymp.UCL
 Dating Apps   3.289474 0.29421947 Inf 2.7605310         4
 Info Sites    0.000000 0.00000023 Inf 0.0000000       Inf
 Social Media  0.342105 0.09488293 Inf 0.1986456         1

Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

```
res_prm = contrast(emm_rate,
                   method = "pairwise",
                   adjust = "BH")
```

```
res_prm = as.data.frame(res_prm)
tab_prm1 = knitr::kable(res_prm, format = "pipe", digits = 3)
writeLines(c("### Replicated Primary Results", "", tab_prm1), "Replicated Primary Results.md"
```

Secondary analysis

- Q12_2: In the PAST 12 MONTHS, how often have you had 5 or more drinks containing alcohol in one day? (1=Daily or almost daily; 4= Weekly; 5= Monthly; 6= Less than monthly; 7= never)
- Q12_3: In the PAST 12 MONTHS, how often have you used any drugs including marijuana, cocaine or crack, heroine, methamphetamine (crystal meth), hallucinogens, ecstasy/MDMA?

Table b: Stage of Health Behavior Change

- Q15_1: Which of these statements is most true for you? (1=I don't see any need to regularly test for HIV; 2=I think I should get tested for HIV regularly, but I am not sure; 3= I'm ready to start getting regularly tested for HIV; 4=I'm trying to get tested regularly for HIV; 5=I've been getting testing for HIV regularly over the past few years.)

```
df = df %>%
  mutate(ORA = case_when(
    ORA_WITHIN60_YESNO == "Yes" ~ "Yes",
    ORA_WITHIN60_YESNO == "No" ~ "No",
    ORA_WITHIN60_YESNO == "Missing" ~ "No"
  ))
```

```
df %>%
  group_by(ORA) %>%
  count(Q15_1)
```

```
# A tibble: 10 x 3
# Groups:   ORA [2]
   ORA   Q15_1     n
   <chr> <int> <int>
 1 No        1     2
 2 No        2     8
 3 No        3    49
 4 No        4    12
 5 No        5     6
```

```
 6 Yes        1     7
 7 Yes        2    31
 8 Yes        3    87
 9 Yes        4    40
10 Yes        5    12
```

```r
stage_desc = tibble::tibble(
  Q15_1 = factor(
    c("Precontemplation","Contemplation","Determination","Action","Maintenance"),
    levels = c("Precontemplation","Contemplation","Determination","Action","Maintenance")
  ),
  Description = c(
    "I do not see any need to regularly test for HIV",
    "I think I should get tested for HIV regularly, but I am not sure",
    "I am ready to start getting regularly tested for HIV",
    "I am trying to get tested regularly for HIV",
    "I have been getting tested for HIV regularly over the past few years"
  )
)

df_stage = df %>%
  select(ORA, Q15_1) %>%
  filter(!is.na(ORA), !is.na(Q15_1)) %>%
  mutate(
    ORA = factor(ORA),
    Q15_1 = factor(
      Q15_1,
      levels = 1:5,
      labels = c(
        "Precontemplation",
        "Contemplation",
        "Determination",
        "Action",
        "Maintenance"
      )
    )
  )

# Fisher's exact test p-value
tab_Q15 = table(df_stage$Q15_1, df_stage$ORA)
p_val_Q15 = fisher.test(tab_Q15)$p.value
```

```r
# form the table b in Appendix 3
denom = df_stage %>%
  count(ORA, name = "N")

counts = df_stage %>%
  count(Q15_1, ORA, name = "n") %>%
  left_join(denom, by = "ORA") %>%
  mutate(
    percent = 100 * n / N,
    n_out_of_N = paste0(n, "/", N)
  )

tab_b = counts %>%
  select(Q15_1, ORA, n_out_of_N, percent) %>%
  pivot_wider(
    names_from = ORA,
    values_from = c(n_out_of_N, percent),
    names_glue = "{ORA}_{.value}"
  ) %>%
  arrange(Q15_1) %>%
  mutate(
    'P-value' = if_else(row_number() == 1,
                        format.pval(p_val_Q15, digits = 3, eps = 0.001),
                        "")
  ) %>%
  left_join(stage_desc, by = "Q15_1") %>%
  # add description col
  relocate(Description, .after = Q15_1) %>%
  # change name according to the table b in Appendix 3
  rename(
    'Ordered test kit (n=177) n/N' = 'Yes_n_out_of_N',
    'Percent_yes' = 'Yes_percent',
    'Did not order test kit (n=77) n/N' = 'No_n_out_of_N',
    'Percent_no' = 'No_percent'
  )


kbl(tab_b, booktabs = TRUE, caption = "b. Stage of Health Behavior Change") %>%
  kable_styling(latex_options = c("scale_down", "repeat_header", "hold_position"))
```

Table 2: b. Stage of Health Behavior Change

Q15_1	Description	Did not order test kit (n=77) n/N	Ordered test kit (n=177) n/N	Percent_no	Percent_yes	P-value
Precontemplation	I do not see any need to regularly test for HIV	2/77	7/177	2.597403	3.954802	0.251
Contemplation	I think I should get tested for HIV regularly, but I am not sure	8/77	31/177	10.389610	17.514124	
Determination	I am ready to start getting regularly tested for HIV	49/77	87/177	63.636364	49.152542	
Action	I am trying to get tested regularly for HIV	12/77	40/177	15.584416	22.598870	
Maintenance	I have been getting tested for HIV regularly over the past few years	6/77	12/177	7.792208	6.779661	

```
gt(tab_b) |>
  gtsave("table_b.png")


readr::write_csv(tab_b, "tab_b.csv")
```

Table c: Attitudes toward human immunodeficiency virus (HIV) testing

- Q15_3: Getting tested for HIV helps people feel better. (1=Agree; 2=Disagree)
- Q15_4: Getting tested for HIV helps people from getting HIV. (1=Agree; 2=Disagree)
- Q15_5: People in my life would leave if I had HIV. (1=Agree; 2=Disagree)
- Q15_6: People who tested positive for HIV should hide it from others. (1=Agree; 2=Disagree)
- Q15_7: I would rather not know if I have HIV. (1=Agree; 2=Disagree)

```
att_long = df %>%
  select(ORA, Q15_3, Q15_4, Q15_5, Q15_6, Q15_7) %>%
  pivot_longer(
    cols = starts_with("Q15_"),
    names_to = "Item",
    values_to = "Resp"
  ) %>%
  filter(!is.na(Resp)) %>%
  mutate(
    Resp = factor(Resp, levels = c(1, 2), labels = c("Agree", "Disagree")),
    Item = factor(
      Item,
      levels = c("Q15_3","Q15_4","Q15_5","Q15_6","Q15_7"),
      labels = c(
        "Getting tested for HIV helps people feel better",
        "Getting tested for HIV helps people from getting HIV",
        "People in my life would leave if I had HIV",
        "People who tested positive for HIV should hide it from others",
        "I would rather not know if I have HIV"
      )
    ),
```

```r
    ORA = factor(ORA)
  )

# calculate p-value
pvals_tabc = att_long %>%
  group_by(Item) %>%
  summarise(
    p_value = fisher.test(table(Resp, ORA))$p.value,
    .groups = "drop"
  ) %>%
  mutate(p_value_fmt = format.pval(p_value, digits = 3, eps = 0.001))

denom = att_long %>%
  count(Item, ORA, name = "N")

counts = att_long %>%
  count(Item, ORA, Resp, name = "n") %>%
  left_join(denom, by = c("Item","ORA")) %>%
  mutate(
    nN = paste0(n, "/", N),
    percent = round(100 * n / N, 1)
  )

# form table c in Appendix 3
tab_c = counts %>%
  select(Item, Resp, ORA, nN, percent) %>%
  pivot_wider(
    names_from = ORA,
    values_from = c(nN, percent),
    names_glue = "{ORA}_{.value}"
  ) %>%
  arrange(Item, Resp) %>%
  left_join(pvals_tabc %>% select(Item, p_value_fmt), by = "Item") %>%
  group_by(Item) %>%
  mutate('P-value' = if_else(row_number() == 1, p_value_fmt, "")) %>%
  ungroup() %>%
  select(-p_value_fmt) %>%
  rename(
    'Ordered test kit (n=177) n/N' = 'Yes_nN',
    'Percent_yes' = 'Yes_percent',
    'Did not order test kit (n=77) n/N' = 'No_nN',
    'Percent_no' = 'No_percent'
```

```
  )
```

```r
kbl(tab_c, booktabs = TRUE, caption = "c. Attitudes toward human immunodeficiency virus (HIV)
  kable_styling(latex_options = c("scale_down", "repeat_header", "hold_position"))
```

Table 3: c. Attitudes toward human immunodeficiency virus (HIV) testing

Item	Resp	Did not order test kit (n=77) n/N	Ordered test kit (n=177) n/N	Percent_no	Percent_yes	P-value
Getting tested for HIV helps people feel better	Agree	70/77	169/177	90.9	95.5	0.1599
Getting tested for HIV helps people feel better	Disagree	7/77	8/177	9.1	4.5	
Getting tested for HIV helps people from getting HIV	Agree	62/76	148/177	81.6	83.6	0.7167
Getting tested for HIV helps people from getting HIV	Disagree	14/76	29/177	18.4	16.4	
People in my life would leave if I had HIV	Agree	37/77	59/175	48.1	33.7	0.0352
People in my life would leave if I had HIV	Disagree	40/77	116/175	51.9	66.3	
People who tested positive for HIV should hide it from others	Agree	9/77	18/177	11.7	10.2	0.8250
People who tested positive for HIV should hide it from others	Disagree	68/77	159/177	88.3	89.8	
I would rather not know if I have HIV	Agree	15/77	27/177	19.5	15.3	0.4628
I would rather not know if I have HIV	Disagree	62/77	150/177	80.5	84.7	

```r
gt(tab_c) |>
  gtsave("table_c.png")
```

```r
readr::write_csv(tab_c, "tab_c.csv")
```

Table d: Attitudes toward human immunodeficiency virus (HIV) treatment (continuous scale from 1 [strongly disagree] to 7 [strongly agree])

- Q94_1 I am less threatened by the idea of being HIV positive than I used to be.
- Q94_5 I am less worried about HIV infection than I used to be.
- Q94_6 I think HIV/AIDS is less of a problem than it used to be.
- Q94_7 I think HIV/AIDS is a less serious threat than it used to be because of new HIV/AIDS treatments.
- Q94_8 I am much less concerned about becoming HIV positive myself because of new HIV/AIDS treatments.
- Q94_9 I think that condom use during sex is less necessary now that new HIV/AIDS treatments are available.
- Q94_10 I think that someone who is HIV positive now needs to care less about condom us.
- Q94_11 I think that the need for condom use is less than it used to be, because you can always start new treatments.
- Q94_12 I think that someone who is HIV positive and uses new HIV/AIDS treatments can be cured.
- Q94_13 I think that new HIV/AIDS treatments can eradicate the virus from your body.

```r
q94_vars = c("Q94_1","Q94_5","Q94_6","Q94_7","Q94_8","Q94_9","Q94_10","Q94_11","Q94_12","Q94

q94_labels = c(
  Q94_1  = "I am less threatened by the idea of being HIV positive than I used to be",
  Q94_5  = "I am less worried about HIV infection than I used to be",
  Q94_6  = "I think HIV/AIDS is less of a problem than it used to be",
  Q94_7  = "I think HIV/AIDS is a less serious threat than it used to be because of new HIV/
  Q94_8  = "I am much less concerned about becoming HIV positive myself because of new HIV/A
  Q94_9  = "I think that condom use during sex is less necessary now that new HIV/AIDS treat
  Q94_10 = "I think that someone who is HIV positive now needs to care less about condom use
  Q94_11 = "I think that the need for condom use is less than it used to be, because you can
  Q94_12 = "I think that someone who is HIV positive and uses new HIV/AIDS treatments can be
  Q94_13 = "I think that new HIV/AIDS treatments can eradicate the virus from your body"
)


df_tabd = df %>%
  select(ORA, all_of(q94_vars)) %>%
  pivot_longer(
    cols = all_of(q94_vars),
    names_to = "Item",
    values_to = "Score"
  ) %>%
  mutate(
    ORA = factor(ORA),
    Item = factor(Item, levels = q94_vars),
    Statements = unname(q94_labels[as.character(Item)]),
    Score = as.numeric(Score)
  )

# calculate mean and sd
sumstats_d = df_tabd %>%
  group_by(Item, Statements, ORA) %>%
  summarise(
    mean = mean(Score, na.rm = TRUE),
    sd   = sd(Score, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(mean_sd = paste0(round(mean, 1), " (", round(sd, 1), ")")) %>%
  select(Item, Statements, ORA, mean_sd) %>%
  pivot_wider(
    names_from = ORA,
```

```
    values_from = mean_sd,
    names_glue = "{ORA}_Mean(SD)"
  )

# Wilcoxon rank test
pvals_d = df_tabd %>%
  group_by(Item, Statements) %>%
  summarise(
    p_value = suppressWarnings(wilcox.test(Score ~ ORA)$p.value),
    .groups = "drop"
  ) %>%
  mutate('P-value (Wilcoxon rank test)' = format.pval(p_value, digits = 3, eps = 0.001)) %>%
  select(Item, 'P-value (Wilcoxon rank test)')

# form table d in Appendix 3
tab_d = sumstats_d %>%
  left_join(pvals_d, by = "Item") %>%
  arrange(Item) %>%
  select(
    Statements,
    ends_with("Mean(SD)"),
    'P-value (Wilcoxon rank test)'
  ) %>%
  rename('Did not order test kit (n=77) Mean (SD)' = 'No_Mean(SD)',
         'Ordered test kit (n=177) Mean (SD)' = 'Yes_Mean(SD)')

kbl(tab_d, booktabs = TRUE, caption = "d. Attitudes toward human immunodeficiency virus (HIV)
  kable_styling(latex_options = c("scale_down", "repeat_header", "hold_position"))
```

Table 4: d. Attitudes toward human immunodeficiency virus (HIV) treatment (continuous scale from 1 [strongly disagree] to 7 [strongly agree])

Statements	Did not order test kit (n=77) Mean (SD)	Ordered test kit (n=177) Mean (SD)	P-value (Wilcoxon rank test)
I am less threatened by the idea of being HIV positive than I used to be	3.8 (2.2)	4.1 (2)	0.4214
I am less worried about HIV infection than I used to be	3.5 (1.8)	3.7 (2)	0.3995
I think HIV/AIDS is less of a problem than it used to be	3.6 (2.1)	3.9 (2.1)	0.4128
I think HIV/AIDS is a less serious threat than it used to be because of new HIV/AIDS treatments	3.8 (2.1)	4.2 (2)	0.2247
I am much less concerned about becoming HIV positive myself because of new HIV/AIDS treatments	3 (1.9)	3.4 (2)	0.2738
I think that condom use during sex is less necessary now that new HIV/AIDS treatments are available	2.2 (1.5)	2.4 (1.8)	0.9707
I think that someone who is HIV positive now needs to care less about condom use	2.5 (2.1)	1.9 (1.7)	0.0638
I think that the need for condom use is less than it used to be, because you can always start new treatments	2.5 (2)	2.5 (1.9)	0.7672
I think that someone who is HIV positive and uses new HIV/AIDS treatments can be cured	3.6 (2.2)	3.2 (1.9)	0.1989
I think that new HIV/AIDS treatments can eradicate the virus from your body	3.8 (2.1)	3.1 (1.9)	0.0290

```
gt(tab_d) |>
  gtsave("table_d.png")

readr::write_csv(tab_d, "tab_d.csv")
```

e. Human immunodeficiency virus (HIV)-related stigma among study participants

- Q14_2: I feel afraid of people living with HIV/AIDS. (1=Strongly agree; 2=Agree; 3=Somewhat agree; 4=Neither agree nor disagree; 5=Somewhat disagree; 6=Disagree; 7=Strongly disagree)
- Q14_3: I could not be friends with someone who has HIV/AIDS.
- Q14_4: People who get HIV/AIDS through sex or drug use got what they deserve.
- Q14_5: I feel anger toward people with HIV/AIDS.

```r
stig_long = df %>%
  select(ORA, Q14_2, Q14_3, Q14_4, Q14_5) %>%
  pivot_longer(
    cols = starts_with("Q14_"),
    names_to = "Item",
    values_to = "Resp_num"
  ) %>%
  mutate(
    ORA = factor(ORA),
    Item = factor(Item, levels = c("Q14_2","Q14_3","Q14_4","Q14_5")),
    Statement = recode(
      Item,
      Q14_2 = "I feel afraid of people living with HIV/AIDS",
      Q14_3 = "I could not be friends with someone who has HIV/AIDS",
      Q14_4 = "People who get HIV/AIDS through sex or drug use got what they deserve",
      Q14_5 = "I feel anger toward people with HIV/AIDS"
    ),
    Resp = factor(
      Resp_num,
      levels = 1:7,
      labels = c("Strongly agree","Agree","Somewhat agree",
                 "Neither agree nor disagree","Somewhat disagree",
                 "Disagree","Strongly disagree")
    )
  )


# calculate p-value
pvals_e = stig_long %>%
  filter(!is.na(Resp_num), !is.na(ORA)) %>%
  group_by(Statement) %>%
  summarise(
    p_value = suppressWarnings(wilcox.test(Resp_num ~ ORA)$p.value),
```

```r
    .groups = "drop"
  ) %>%
  mutate(p_fmt = format.pval(p_value, digits = 3, eps = 0.001))


denom_e = stig_long %>%
  filter(!is.na(Resp), !is.na(ORA)) %>%
  count(Statement, ORA, name = "N")

counts_e = stig_long %>%
  filter(!is.na(Resp), !is.na(ORA)) %>%
  count(Statement, Resp, ORA, name = "n") %>%
  complete(Statement, Resp, ORA, fill = list(n = 0)) %>%
  left_join(denom_e, by = c("Statement", "ORA")) %>%
  mutate(
    nN = paste0(n, "/", N),
    percent = round(100 * n / N, 1)
  )


# form the table
tab_e = counts_e %>%
  select(Statement, Resp, ORA, nN, percent) %>%
  pivot_wider(
    names_from = ORA,
    values_from = c(nN, percent),
    names_glue = "{ORA}_{.value}"
  ) %>%
  arrange(Statement, Resp) %>%
  left_join(pvals_e %>% select(Statement, p_fmt), by = "Statement") %>%
  group_by(Statement) %>%
  mutate(P_value = if_else(row_number() == 1, p_fmt, "")) %>%
  ungroup() %>%
  select(-p_fmt) %>%
  rename('Response' = 'Resp',
         'Did not order test kit (n=77) n/N' = 'No_nN',
         'Ordered test kit (n=177) n/N' = 'Yes_nN',
         'Percent_yes' = 'Yes_percent',
         'Percent_no' = 'No_percent')


kbl(tab_e, booktabs = TRUE, caption = "e. Human immunodeficiency virus (HIV)-related stigma a
```

```
kable_styling(latex_options = c("scale_down", "repeat_header", "hold_position"))
```

Table 5: e. Human immunodeficiency virus (HIV)-related stigma among study participants

Statement	Response	Did not order test kit (n=77) n/N	Ordered test kit (n=177) n/N	Percent_no	Percent_yes	P_value
I feel afraid of people living with HIV/AIDS	Strongly agree	7/77	12/177	9.1	6.8	0.6129
I feel afraid of people living with HIV/AIDS	Agree	2/77	9/177	2.6	5.1	
I feel afraid of people living with HIV/AIDS	Somewhat agree	9/77	21/177	11.7	11.9	
I feel afraid of people living with HIV/AIDS	Neither agree nor disagree	9/77	20/177	11.7	11.3	
I feel afraid of people living with HIV/AIDS	Somewhat disagree	6/77	11/177	7.8	6.2	
I feel afraid of people living with HIV/AIDS	Disagree	17/77	32/177	22.1	18.1	
I feel afraid of people living with HIV/AIDS	Strongly disagree	27/77	72/177	35.1	40.7	
I could not be friends with someone who has HIV/AIDS	Strongly agree	4/77	1/177	5.2	0.6	0.0325
I could not be friends with someone who has HIV/AIDS	Agree	0/77	1/177	0.0	0.6	
I could not be friends with someone who has HIV/AIDS	Somewhat agree	1/77	3/177	1.3	1.7	
I could not be friends with someone who has HIV/AIDS	Neither agree nor disagree	6/77	9/177	7.8	5.1	
I could not be friends with someone who has HIV/AIDS	Somewhat disagree	1/77	7/177	1.3	4.0	
I could not be friends with someone who has HIV/AIDS	Disagree	19/77	26/177	24.7	14.7	
I could not be friends with someone who has HIV/AIDS	Strongly disagree	46/77	130/177	59.7	73.4	
People who get HIV/AIDS through sex or drug use got what they deserve	Strongly agree	0/77	2/177	0.0	1.1	0.3319
People who get HIV/AIDS through sex or drug use got what they deserve	Agree	1/77	2/177	1.3	1.1	
People who get HIV/AIDS through sex or drug use got what they deserve	Somewhat agree	2/77	3/177	2.6	1.7	
People who get HIV/AIDS through sex or drug use got what they deserve	Neither agree nor disagree	7/77	8/177	9.1	4.5	
People who get HIV/AIDS through sex or drug use got what they deserve	Somewhat disagree	2/77	6/177	2.6	3.4	
People who get HIV/AIDS through sex or drug use got what they deserve	Disagree	12/77	24/177	15.6	13.6	
People who get HIV/AIDS through sex or drug use got what they deserve	Strongly disagree	53/77	132/177	68.8	74.6	
I feel anger toward people with HIV/AIDS	Strongly agree	0/77	1/177	0.0	0.6	0.2130
I feel anger toward people with HIV/AIDS	Agree	0/77	0/177	0.0	0.0	
I feel anger toward people with HIV/AIDS	Somewhat agree	0/77	0/177	0.0	0.0	
I feel anger toward people with HIV/AIDS	Neither agree nor disagree	5/77	11/177	6.5	6.2	
I feel anger toward people with HIV/AIDS	Somewhat disagree	0/77	3/177	0.0	1.7	
I feel anger toward people with HIV/AIDS	Disagree	16/77	19/177	20.8	10.7	
I feel anger toward people with HIV/AIDS	Strongly disagree	56/77	143/177	72.7	80.8	

```
gt(tab_e) |>
  gtsave("table_e.png")

readr::write_csv(tab_e, "tab_e.csv")
```

f. Medical mistrust

- Q16_1 You'd better be cautious when dealing with healthcare organizations. (28=Strongly agree; 30=Agree; 33=Disagree; 34=Strongly disagree)
- Q16_2 Patients have sometimes been deceived or misled by health care organizations. (1=Strongly agree; 2=Agree; 6=Disagree; 7=Strongly disagree)
- Q16_3 When health care organizations make mistakes they usually cover it up. (1=Strongly agree; 2=Agree; 6=Disagree; 7=Strongly disagree)
- Q16_4 Health care organizations have sometimes done harmful experiments on patients without their knowledge. (1=Strongly agree; 2=Agree; 6=Disagree; 7=Strongly disagree)
- Q16_5 Health care organizations don't always keep your information totally private. (1=Strongly agree; 2=Agree; 6=Disagree; 7=Strongly disagree)
- Q16_6 Sometimes I wonder if health care organizations really know what they are doing. (1=Strongly agree; 2=Agree; 6=Disagree; 7=Strongly disagree)
- Q16_7 Mistakes are common in health care organizations. (1=Strongly agree; 2=Agree; 6=Disagree; 7=Strongly disagree)

```r
mistrust_long = df %>%
  select(ORA, Q16_1:Q16_7) %>%
  pivot_longer(cols = starts_with("Q16_"),
               names_to = "Item", values_to = "Resp_num") %>%
  mutate(
    ORA = factor(ORA),
    Statement = recode(Item,
      Q16_1 = "You'd better be cautious when dealing with healthcare organizations",
      Q16_2 = "Patients have sometimes been deceived or misled by health care organizations"
      Q16_3 = "When health care organizations make mistakes they usually cover it up",
      Q16_4 = "Health care organizations have sometimes done harmful experiments on patients
      Q16_5 = "Health care organizations don't always keep your information totally private"
      Q16_6 = "Sometimes I wonder if health care organizations really know what they are doi
      Q16_7 = "Mistakes are common in health care organizations"
    ),
    Resp_num = as.numeric(Resp_num),
    Resp_num = ifelse(Item == "Q16_1" & Resp_num == 28, 1, Resp_num),
    Resp_num = ifelse(Item == "Q16_1" & Resp_num == 30, 2, Resp_num),
    Resp_num = ifelse(Item == "Q16_1" & Resp_num == 33, 6, Resp_num),
    Resp_num = ifelse(Item == "Q16_1" & Resp_num == 34, 7, Resp_num),
    Resp = factor(
      Resp_num,
      levels = c(1,2,6,7),
      labels = c("Strongly agree","Agree","Disagree","Strongly disagree")
    )
  )


# calculate p-values based on Wilcoxon rank test
pvals_f = mistrust_long %>%
  filter(!is.na(Resp_num), !is.na(ORA)) %>%
  group_by(Statement) %>%
  summarise(
    p_value = wilcox.test(Resp_num ~ ORA)$p.value,
    .groups = "drop"
  ) %>%
  mutate(p_fmt = format.pval(p_value, digits = 3, eps = 0.001))


denom_f = mistrust_long %>%
  filter(!is.na(Resp), !is.na(ORA)) %>%
  count(Statement, ORA, name = "N")
```

```r
counts_f = mistrust_long %>%
  filter(!is.na(Resp), !is.na(ORA)) %>%
  count(Statement, Resp, ORA, name = "n") %>%
  complete(Statement, Resp, ORA, fill = list(n = 0)) %>%
  left_join(denom_f, by = c("Statement","ORA")) %>%
  mutate(
    nN = paste0(n, "/", N),
    percent = round(100 * n / N, 1)
  )

# form the table
tab_f = counts_f %>%
  select(Statement, Resp, ORA, nN, percent) %>%
  pivot_wider(
    names_from = ORA,
    values_from = c(nN, percent),
    names_glue = "{ORA}_{.value}"
  ) %>%
  arrange(Statement, Resp) %>%
  left_join(pvals_f %>% select(Statement, p_fmt), by = "Statement") %>%
  group_by(Statement) %>%
  mutate(`P-value` = if_else(row_number() == 1, p_fmt, "")) %>%
  ungroup() %>%
  select(-p_fmt) %>%
  rename('Response' = 'Resp',
         'Did not order test kit (n=77) n/N' = 'No_nN',
         'Ordered test kit (n=177) n/N' = 'Yes_nN',
         'Percent_yes' = 'Yes_percent',
         'Percent_no' = 'No_percent')


kbl(tab_f, booktabs = TRUE, caption = "f. Medical mistrust") %>%
  kable_styling(latex_options = c("scale_down", "repeat_header", "hold_position"))


gt(tab_f) |>
  gtsave("table_f.png")

readr::write_csv(tab_f, "tab_f.csv")
```

Table 6: f. Medical mistrust

Statement	Response	Did not order test kit (n=77) n/N	Ordered test kit (n=177) n/N	Percent_no	Percent_yes	P-value
Health care organizations don't always keep your information totally private	Strongly agree	11/76	36/175	14.5	20.6	0.371
Health care organizations don't always keep your information totally private	Agree	29/76	64/175	38.2	36.6	
Health care organizations don't always keep your information totally private	Disagree	23/76	47/175	30.3	26.9	
Health care organizations don't always keep your information totally private	Strongly disagree	13/76	28/175	17.1	16.0	
Health care organizations have sometimes done harmful experiments on patients without their knowledge	Strongly agree	14/77	30/176	18.2	17.0	0.413
Health care organizations have sometimes done harmful experiments on patients without their knowledge	Agree	25/77	73/176	32.5	41.5	
Health care organizations have sometimes done harmful experiments on patients without their knowledge	Disagree	26/77	51/176	33.8	29.0	
Health care organizations have sometimes done harmful experiments on patients without their knowledge	Strongly disagree	12/77	22/176	15.6	12.5	
Mistakes are common in health care organizations	Strongly agree	8/75	21/176	10.7	11.9	0.638
Mistakes are common in health care organizations	Agree	36/75	89/176	48.0	50.6	
Mistakes are common in health care organizations	Disagree	23/75	46/176	30.7	26.1	
Mistakes are common in health care organizations	Strongly disagree	8/75	20/176	10.7	11.4	
Patients have sometimes been deceived or misled by health care organizations	Strongly agree	11/77	32/176	14.3	18.2	0.413
Patients have sometimes been deceived or misled by health care organizations	Agree	34/77	83/176	44.2	47.2	
Patients have sometimes been deceived or misled by health care organizations	Disagree	22/77	33/176	28.6	18.8	
Patients have sometimes been deceived or misled by health care organizations	Strongly disagree	10/77	28/176	13.0	15.9	
Sometimes I wonder if health care organizations really know what they are doing	Strongly agree	9/77	22/176	11.7	12.5	0.965
Sometimes I wonder if health care organizations really know what they are doing	Agree	28/77	66/176	36.4	37.5	
Sometimes I wonder if health care organizations really know what they are doing	Disagree	30/77	60/176	39.0	34.1	
Sometimes I wonder if health care organizations really know what they are doing	Strongly disagree	10/77	28/176	13.0	15.9	
When health care organizations make mistakes they usually cover it up	Strongly agree	10/76	29/174	13.2	16.7	0.222
When health care organizations make mistakes they usually cover it up	Agree	29/76	83/174	38.2	47.7	
When health care organizations make mistakes they usually cover it up	Disagree	30/76	36/174	39.5	20.7	
When health care organizations make mistakes they usually cover it up	Strongly disagree	7/76	26/174	9.2	14.9	
You'd better be cautious when dealing with healthcare organizations	Strongly agree	17/76	37/177	22.4	20.9	0.503
You'd better be cautious when dealing with healthcare organizations	Agree	32/76	66/177	42.1	37.3	
You'd better be cautious when dealing with healthcare organizations	Disagree	14/76	42/177	18.4	23.7	
You'd better be cautious when dealing with healthcare organizations	Strongly disagree	13/76	32/177	17.1	18.1	