

# Untitled

Yunting Chiu

9/27/2020

5. Use the lead.csv data set from HW 1. Revisit HW 1 for a description of this data set and study. For this problem you will compare the Wechsler full-scale IQ scores (the variable IQF) between the different lead exposure groups, denoted by the GROUP variable.

```
library(mosaic)
```

```
## 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.
##
## Attaching package: 'mosaic'
##
## The following objects are masked from 'package:dplyr':
##
##   count, do, tally
##
## The following object is masked from 'package:Matrix':
##
##   mean
##
## The following object is masked from 'package:ggplot2':
##
##   stat
##
## The following objects are masked from 'package:stats':
##
##   binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##   quantile, sd, t.test, var
##
## The following objects are masked from 'package:base':
##
##   max, mean, min, prod, range, sample, sum
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
## v tibble  3.0.3      v purrr   0.3.4
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
##
## -- Conflicts ----- tidyverse_conflicts() --
## x mosaic::count()      masks dplyr::count()
## x purrr::cross()       masks mosaic::cross()
```

```

## x mosaic::do()           masks dplyr::do()
## x tidyr::expand()        masks Matrix::expand()
## x dplyr::filter()        masks stats::filter()
## x ggstance::geom_errorbarh() masks ggplot2::geom_errorbarh()
## x dplyr::lag()           masks stats::lag()
## x tidyr::pack()          masks Matrix::pack()
## x mosaic::stat()         masks ggplot2::stat()
## x mosaic::tally()        masks dplyr::tally()
## x tidyr::unpack()        masks Matrix::unpack()

library(readr)
bloodlead <- read_csv(file = "lead.csv")

## Parsed with column specification:
## cols(
##   .default = col_double()
## )

## See spec(...) for full column specifications.

print(bloodlead)

## # A tibble: 124 x 40
##       id area age sex iqv_inf iqv_comp iqv_ar iqv_ds iqv_raw iqp_pc iqp_bd
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 101 3 1101 1 3 4 3 5 15 10 8
## 2 102 3 905 1 7 9 7 6 29 8 7
## 3 103 3 1101 1 4 9 5 3 21 10 7
## 4 104 2 611 1 4 6 6 6 22 5 8
## 5 105 1 1103 1 5 4 8 5 22 5 10
## 6 106 2 606 1 5 12 11 9 37 14 7
## 7 107 3 611 1 7 9 10 7 33 10 8
## 8 108 1 1500 2 3 1 3 6 13 6 2
## 9 109 2 702 2 13 10 14 13 50 8 15
## 10 110 2 703 1 7 9 12 9 37 6 9
## # ... with 114 more rows, and 29 more variables: iqp_oa <dbl>, iqp_cod <dbl>,
## # iqp_raw <dbl>, hh_index <dbl>, iqv <dbl>, iqp <dbl>, iqf <dbl>,
## # iq_type <dbl>, lead_typ <dbl>, ld72 <dbl>, ld73 <dbl>, fst2yrs <dbl>,
## # totyrs <dbl>, pica <dbl>, colic <dbl>, clumsi <dbl>, irrit <dbl>,
## # convul <dbl>, `@2plat_r` <dbl>, `@2plat_l` <dbl>, visrea_r <dbl>,
## # visrea_l <dbl>, audrea_r <dbl>, audrea_l <dbl>, fwt_r <dbl>, fwt_l <dbl>,
## # hyperact <dbl>, MAXFT <dbl>, GROUP <dbl>

bloodlead %>%
  filter(GROUP == 1) -> leadGroup1
bloodlead %>%
  filter(GROUP == 2) -> leadGroup2
nrow(leadGroup1)

## [1] 46

nrow(leadGroup2)

## [1] 78

```

- Compute the mean, standard deviation, standard error, and 95% confidence interval for the population mean IQ score for **each** lead exposure group, separately. Summarize each confidence interval.

- IQF = the Wechsler full-scale IQ score.

```
# mean and standard deviation of group 1
favstats(~iqf, data = leadGroup1)

## min Q1 median Q3 max mean sd n missing
## 46 80 88 93.75 114 88.02174 12.20654 46 0

# mean and standard deviation of group 2
favstats(~iqf, data = leadGroup2)

## min Q1 median Q3 max mean sd n missing
## 50 85 94 101 141 92.88462 15.34451 78 0

# standard error of group 1, 2
library(plotrix)

##
## Attaching package: 'plotrix'

## The following object is masked from 'package:mosaic':
##
## rescale

std.error(leadGroup1$iqf, na.rm = TRUE)

## [1] 1.799756

std.error(leadGroup2$iqf, na.rm = TRUE)

## [1] 1.737424

# 95% confidence interval of group 1
tout1 <- t.test(~iqf, data = leadGroup1, conf.level = 0.95)
attributes(tout1)

## $names
## [1] "statistic" "parameter" "p.value" "conf.int" "estimate"
## [6] "null.value" "stderr" "alternative" "method" "data.name"
##
## $class
## [1] "htest"

# 95% confidence interval of group 2
tout2 <- t.test(~iqf, data = leadGroup2, conf.level = 0.95)
tout2

##
## One Sample t-test
##
## data: iqf
## t = 53.461, df = 77, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 89.42496 96.34427
## sample estimates:
## mean of x
## 92.88462

attributes(tout2)
```

```
## $names
## [1] "statistic" "parameter" "p.value" "conf.int" "estimate"
## [6] "null.value" "stderr" "alternative" "method" "data.name"
##
## $class
## [1] "htest"
```

- b. Researchers were interested in assessing the difference in the mean IQ score between the two exposure group populations. Give the estimate mean difference, the standard error, and the 95% confidence interval for the difference in population mean IQ scores. Summarize the confidence interval.

```
# group1 and group2 two sample t-test
```

```
t.test(iqf~GROUP, data = bloodlead, conf.level = 0.95) # iqf by each Group
```

```
##
## Welch Two Sample t-test
##
## data: iqf by GROUP
## t = -1.9439, df = 111.41, p-value = 0.05442
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -9.81966762 0.09391511
## sample estimates:
## mean in group 1 mean in group 2
## 88.02174 92.88462
```

```
# group1 and group2 two sample t-test
```

```
toutland2_01 <- t.test(iqf~GROUP, mu = 0, data = bloodlead, conf.level = 0.95) # iqf by each Group
toutland2_01
```

```
##
## Welch Two Sample t-test
##
## data: iqf by GROUP
## t = -1.9439, df = 111.41, p-value = 0.05442
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -9.81966762 0.09391511
## sample estimates:
## mean in group 1 mean in group 2
## 88.02174 92.88462
```

```
toutland2_01$stderr # standard error
```

```
## [1] 2.501552
```

- c. Researchers hypothesized that the exposed group (GROUP = 1) would have a lower population mean IQ score than the control group (GROUP = 2). Set up and conduct a statistical hypothesis test to address the research hypothesis. Carefully state the null and alternative hypotheses to be tested. Give the parameter of interest, the estimate of this parameter, the standard error of the estimate, the test statistic, and the p-value. Summarize the results of the test.

```
# group1 and group2 two sample t-test
```

```
toutland2_02 <- t.test(iqf~GROUP, mu = 0, data = bloodlead,
                      conf.level = 0.95, alternative = "greater") # iqf by each Group
toutland2_02
```

```
##
```

```
## Welch Two Sample t-test
##
## data:  iqf by GROUP
## t = -1.9439, df = 111.41, p-value = 0.9728
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -9.012065      Inf
## sample estimates:
## mean in group 1 mean in group 2
##      88.02174      92.88462
toutland2_02$stderr
## [1] 2.501552
```