HW8

Natalie Brewer

2023-11-01

Problem 8A

```
grades <- matrix (c(8, 15, 13,
                  14, 19, 15,
                  15, 4, 7,
                  3, 1, 4), nrow = 4, byrow = TRUE)
colnames(grades) <- c("Pysch", "Bio", "Other")
rownames(grades) <- c("A", "B", "C", "D-F")
grades
```

```
##     Pysch Bio Other
## A       8  15    13
## B      14  19    15
## C      15   4     7
## D-F     3   1     4
```

```
chitest <- chisq.test(grades)
```

```
## Warning in chisq.test(grades): Chi-squared approximation may be incorrect
```

```
chitest
```

```
##
##  Pearson's Chi-squared test
##
## data:  grades
## X-squared = 12.183, df = 6, p-value = 0.058
```

```
chitest$expected
```

```
##         Pysch       Bio     Other
## A   12.203390 11.898305 11.898305
## B   16.271186 15.864407 15.864407
## C    8.813559  8.593220  8.593220
## D-F  2.711864  2.644068  2.644068
```

Problem 8B

```
drugs <- matrix(c( 165, 95,
                   152, 52,
                   85, 52,
                   67, 35,
                   85, 37), nrow=5, byrow=TRUE)

drugs <- cbind(drugs, drugs[,1] - drugs[,2])
rownames(drugs) <- c("Placebo","Chlorpromazine","Dimenhydrinate","Pentobarbital 100","Pe
ntobarbital 150")
colnames(drugs) <- c("Patients", "Nausea", "No Nausea")
drugs
```

```
##                   Patients Nausea No Nausea
## Placebo                165     95        70
## Chlorpromazine         152     52       100
## Dimenhydrinate          85     52        33
## Pentobarbital 100       67     35        32
## Pentobarbital 150       85     37        48
```

```
# Compare drugs to eachother
chitest_B <- chisq.test(drugs[2:5,2:3])
chitest_B
```

```
##
##  Pearson's Chi-squared test
##
## data:  drugs[2:5, 2:3]
## X-squared = 17.603, df = 3, p-value = 0.0005311
```

```
# Compare drugs to placebo

test_chlor <- chisq.test(drugs[1:2, 2:3], correct=FALSE) # Chlorpromazine
test_dimen <- chisq.test(drugs[c(1,3), 2:3], correct=FALSE) # Dimenhydrinate
test_pento100 <- chisq.test(drugs[c(1,4), 2:3], correct=FALSE) # Pentobarbital 100
test_pento150 <- chisq.test(drugs[c(1,5), 2:3], correct=FALSE) # Pentobarbital 150

all_tests <- matrix(c(test_chlor$statistic, test_dimen$statistic, test_pento100$statisti
c, test_pento150$statistic,
                      test_chlor$p.value, test_dimen$p.value, test_pento100$p.value, tes
t_pento150$p.value), nrow=2, byrow=TRUE)
colnames(all_tests) <- c("Chlorpromazine","Dimenhydrinate","Pentobarbital 100","Pentobar
bital 150")
rownames(all_tests) <- c("Statistic", "p value")
all_tests
```

```
##           Chlorpromazine Dimenhydrinate Pentobarbital 100 Pentobarbital 150
## Statistic   1.736844e+01      0.3002378         0.5509215        4.44133623
## p value     3.078970e-05      0.5837334         0.4579414        0.03507878
```

Problem 8D

```r
B = 10000

fun <- function() {
  x1 <- rnorm(1)
  y <- rnorm(2, mean=1)
  p <- prod(y<x1)
}

vec <- replicate(B, fun())
mean(vec)
```

```
## [1] 0.1163
```

Problem 8E

```r
heat_data <- read.table("/Users/nataliebrewer/Desktop/Stat 135/HW8/11.29.txt", sep= " ")
heat_df <- data.frame(heat_data)
heat_df <- heat_df[-1,]
heat_df$V1 <- as.numeric(heat_df$V1)
heat_df$V2 <- as.numeric(heat_df$V2)
head(heat_df)
```

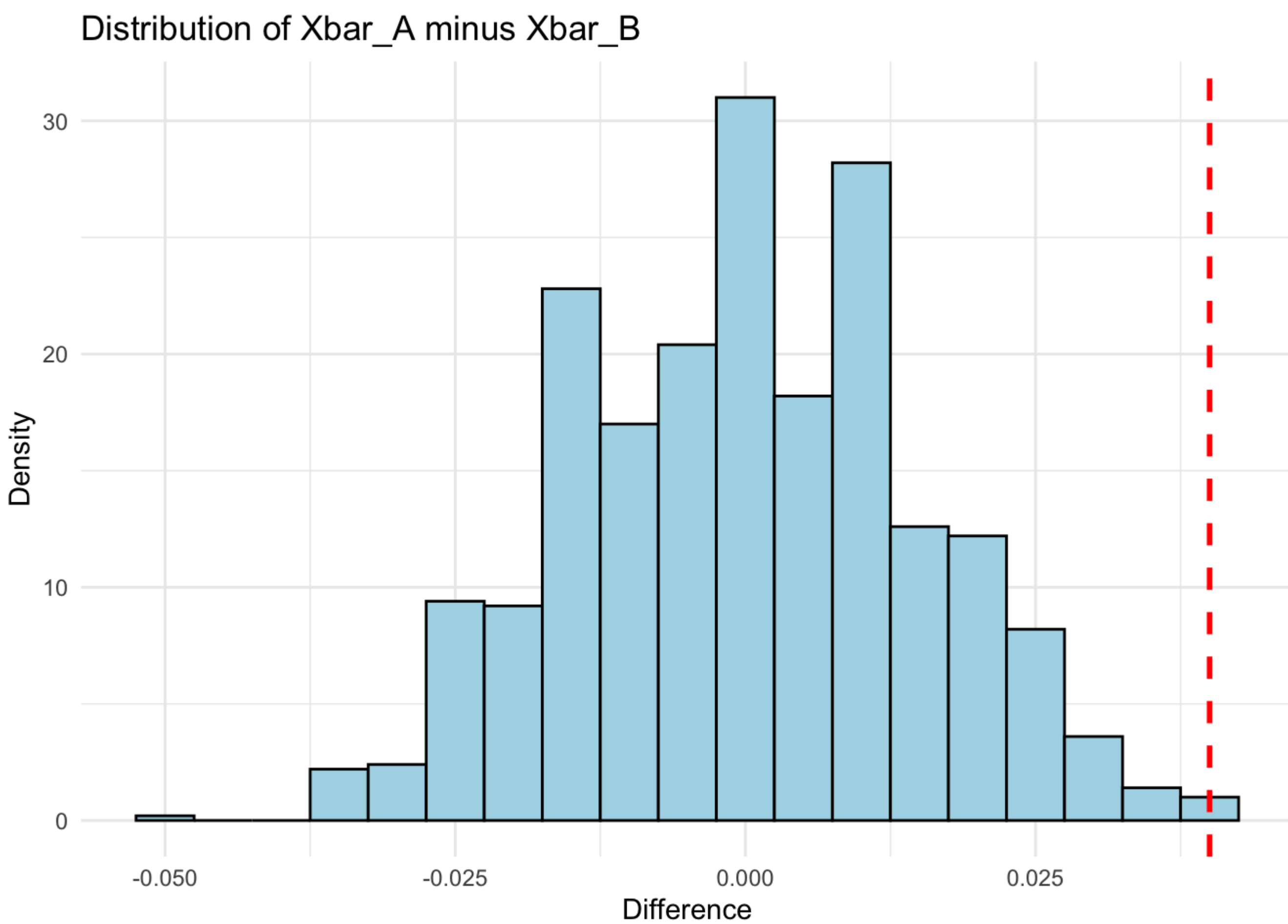
```
##      V1 V2
## 2 79.98  1
## 3 80.04  1
## 4 80.02  1
## 5 80.04  1
## 6 80.03  1
## 7 80.03  1
```

```r
xbar_diff <- 80.02 - 79.98

get_xbar_diff <- function() {
  indices <- sample(1:21, 8)
  xbarA <- mean(heat_df[-indices, 1])
  xbarB <- mean(heat_df[indices, 1])
  diff <- xbarA - xbarB
  return(diff)
}

partitions <-  data.frame(diff=replicate(1000, get_xbar_diff()))

ggplot(partitions, aes(x=diff)) +
  geom_histogram(aes(y=after_stat(density)), fill="lightblue", color="black",binwidth=0.
005) +
  geom_vline(aes(xintercept = xbar_diff), color = "red", linetype = "dashed", linewidth
= 1) +
  labs(title="Distribution of Xbar_A minus Xbar_B",
       x="Difference",
       y="Density") +
  theme_minimal()
```



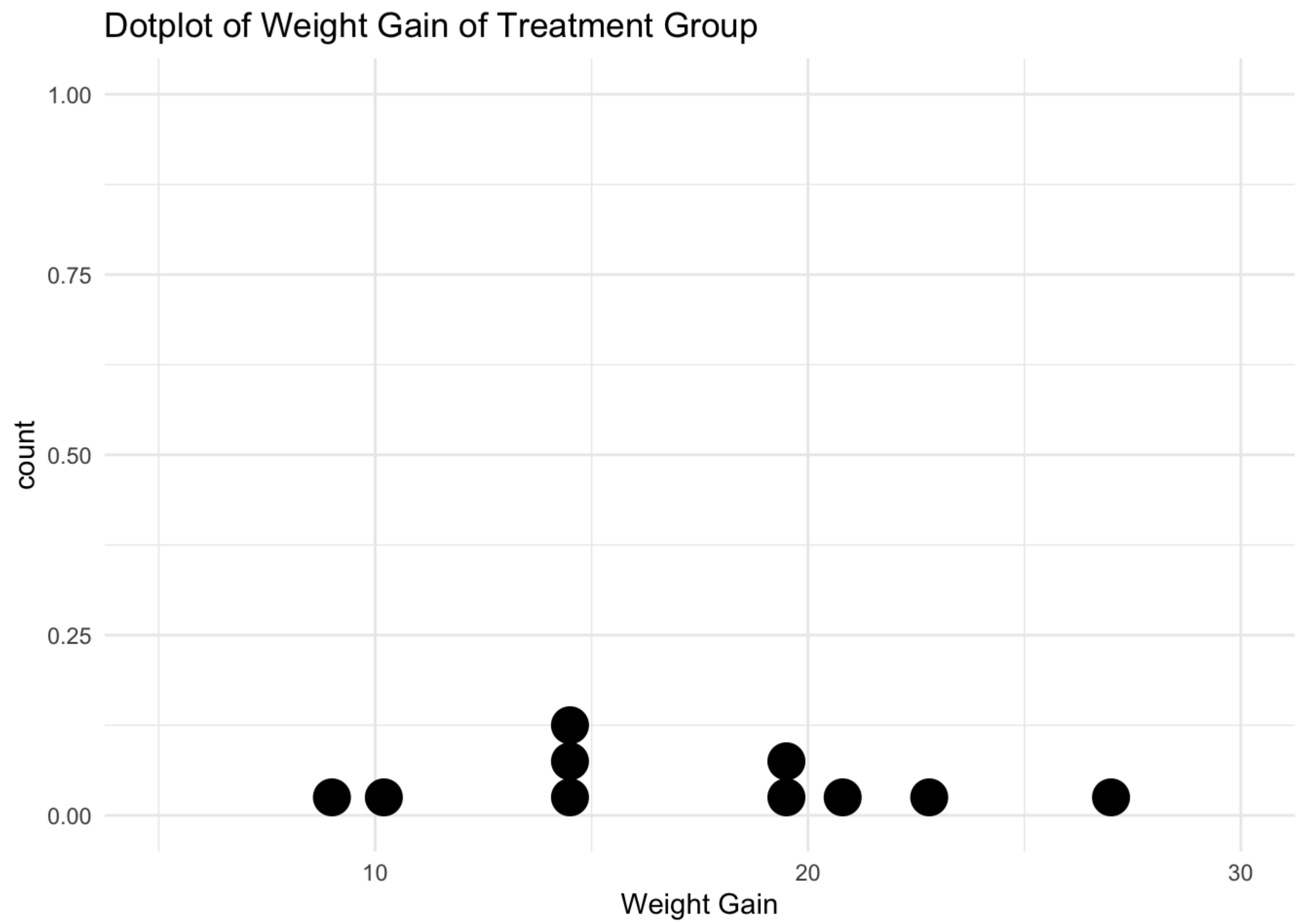
Problem 8F

```
mice_data <- read.table("/Users/nataliebrewer/Desktop/Stat 135/HW8/11.40.txt", sep= " ")
mice_df <- data.frame(mice_data)
mice_df <- mice_df[-1,]
mice_df$V1 <- as.numeric(mice_df$V1)
mice_df$V2 <- as.numeric(mice_df$V2)
mice_df
```

```
##      V1   V2
## 2  22.8 23.5
## 3  10.2 31.0
## 4  20.8 19.5
## 5  27.0 26.2
## 6  19.2 26.5
## 7   9.0 25.2
## 8  14.2 24.5
## 9  19.8 23.8
## 10 14.5 27.8
## 11 14.8 22.0
```

```
ggplot(mice_df) +
  geom_dotplot(aes(x=mice_df$V1))+
  labs(title="Dotplot of Weight Gain of Treatment Group",
       x="Weight Gain") +
  xlim(5, 30) +
  theme_minimal()
```

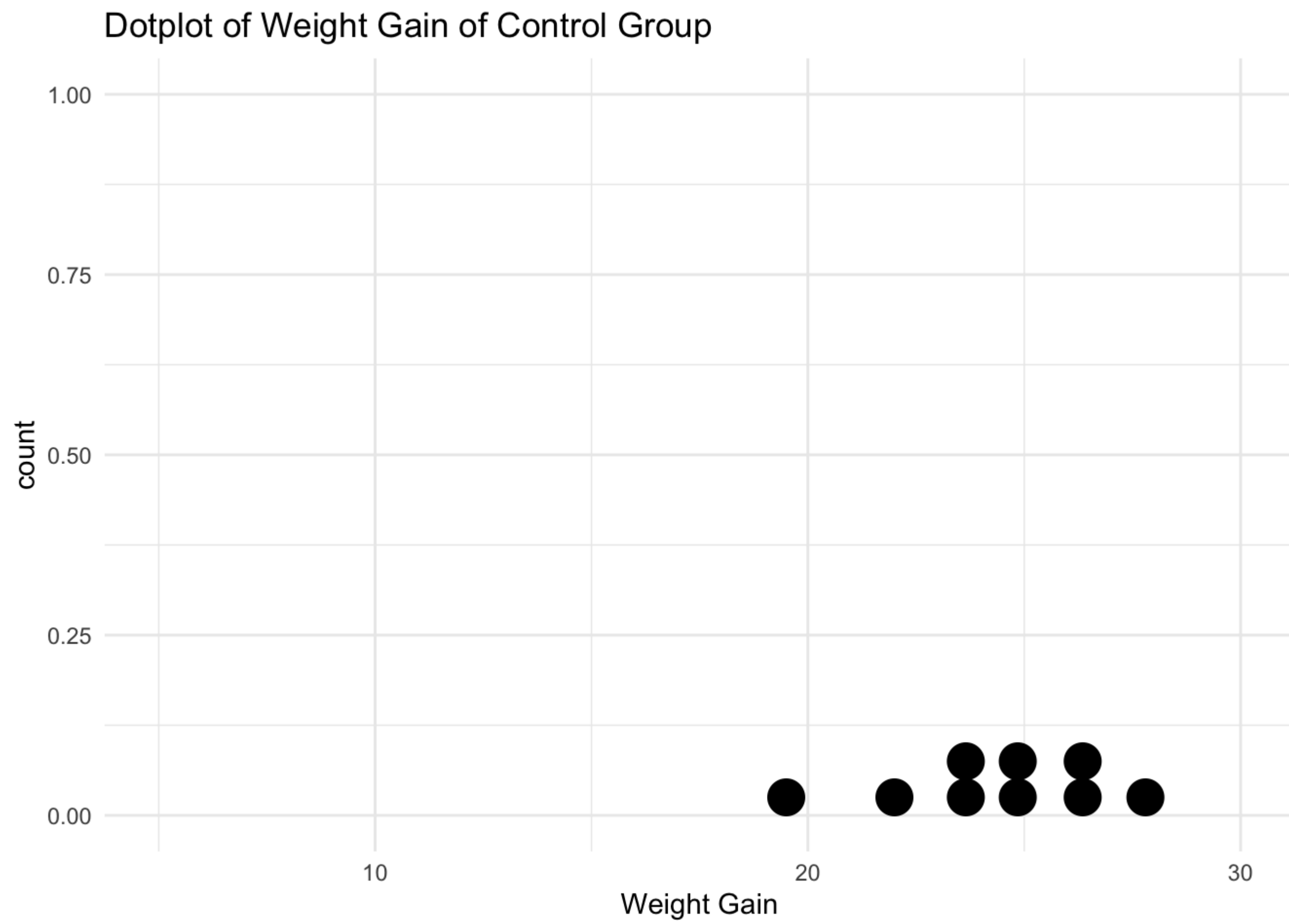
```
## Bin width defaults to 1/30 of the range of the data. Pick better value with
## `binwidth`.
```




```
ggplot(mice_df) +
  geom_dotplot(aes(x=mice_df$V2))+
  labs(title="Dotplot of Weight Gain of Control Group",
       x="Weight Gain") +
  xlim(5, 30) +
  theme_minimal()
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with
## `binwidth`.
```

```
## Warning: Removed 1 rows containing missing values (`stat_bindot()`).
```



Problem 8G

```
# Part b
xbar <- mean(mice_df$V1)
xbar
```

```
## [1] 17.23
```

```
ybar <- mean(mice_df$V2)
ybar
```

```
## [1] 25
```

```
xbar_minus_ybar <- xbar - ybar
xbar_minus_ybar
```

```
## [1] -7.77
```

```
var_x <- var(mice_df$V1)
var_x
```

```
## [1] 32.15567
```

```
var_y <- var(mice_df$V2)
var_y
```

```
## [1] 10.08444
```

```
pooled_sd <- sqrt((9*var_x + 9*var_y)/(18))
sd_diff <- pooled_sd * sqrt(2/10)
sd_diff
```

```
## [1] 2.05524
```

```
CI <- c(xbar_minus_ybar + (qt(.025,18)*sd_diff), xbar_minus_ybar - (qt(.025,18)*sd_dif
f))
CI
```

```
## [1] -12.087899  -3.452101
```

```
# Part c
test_result <- t.test(mice_df$V1, mice_df$V2, var.equal = TRUE)
test_result
```

```
##
##  Two Sample t-test
##
## data:  mice_df$V1 and mice_df$V2
## t = -3.7806, df = 18, p-value = 0.001369
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12.087899  -3.452101
## sample estimates:
## mean of x mean of y
##     17.23     25.00
```

Problem 8H

```
MW_test <- wilcox.test(mice_df$V1, mice_df$V2)
MW_test
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  mice_df$V1 and mice_df$V2
## W = 12, p-value = 0.002879
## alternative hypothesis: true location shift is not equal to 0
```

Problem 8I

```
# Part a
bodytemps <- read.table("/Users/nataliebrewer/Desktop/Stat 135/HW8/bodytemp.txt", sep= "
")
bodytemps_df <- data.frame(bodytemps)
bodytemps_df <-  bodytemps_df[-1,]
bodytemps_df$V1 <- as.numeric(bodytemps_df$V1)
bodytemps_df$V2 <- as.numeric(bodytemps_df$V2)
bodytemps_df$V3 <- as.numeric(bodytemps_df$V3)
colnames(bodytemps_df) <- c("temp","gender","rate")
head(bodytemps_df)
```

```
##   temp gender rate
## 2 96.3      1   70
## 3 96.7      1   71
## 4 96.9      1   74
## 5 97.0      1   80
## 6 97.1      1   73
## 7 97.1      1   75
```

```
xbar_m <- mean(bodytemps_df$temp[bodytemps_df$gender == 1])
xbar_f <- mean(bodytemps_df$temp[bodytemps_df$gender == 2])
m_minus_f <- xbar_m - xbar_f
m_minus_f
```

```
## [1] -0.2892308
```

```
var_m <- var(bodytemps_df$temp[bodytemps_df$gender == 1])
var_m
```

```
## [1] 0.4882596
```

```
var_f <- var(bodytemps_df$temp[bodytemps_df$gender == 2])
var_f
```

```
## [1] 0.552774
```

```
mf_pooled_sd <- sqrt((64*var_m + 64*var_f)/(128))
sd_mf <- mf_pooled_sd * sqrt(2/65)
sd_mf
```

```
## [1] 0.126554
```

```
CI_mf <- c(m_minus_f + (qt(.025,128)*sd_mf), m_minus_f - (qt(.025,128)*sd_mf))
CI_mf
```

```
## [1] -0.53963938 -0.03882216
```

```
# Part b
xbar_m_rate <- mean(bodytemps_df$rate[bodytemps_df$gender == 1])
xbar_f_rate <- mean(bodytemps_df$rate[bodytemps_df$gender == 2])
rate_diff <- xbar_m_rate - xbar_f_rate
rate_diff
```

```
## [1] -0.7846154
```

```
var_m_rate <- var(bodytemps_df$rate[bodytemps_df$gender == 1])
var_m_rate
```

```
## [1] 34.51779
```

```
var_f_rate <- var(bodytemps_df$rate[bodytemps_df$gender == 2])
var_f_rate
```

```
## [1] 65.69471
```

```
sd_diff_rate <- sqrt((var_m_rate/65) + (var_f_rate/65))
sd_diff_rate
```

```
## [1] 1.241665
```

```
degfree <- t.test(bodytemps_df$rate[bodytemps_df$gender == 1], bodytemps_df$rate[bodytem
ps_df$gender == 2])$parameter
degfree
```

```
##       df
## 116.7044
```

```
CI_rate <- c(rate_diff + (qt(.025,degfree)*sd_diff_rate), rate_diff - (qt(.025,degfree)*
sd_diff_rate))
CI_rate
```

```
## [1] -3.243732  1.674501
```

```
# Part c

# t test for temperature
t.test(bodytemps_df$temp[bodytemps_df$gender == 1], bodytemps_df$temp[bodytemps_df$gende
r == 2])
```

```
##
##  Welch Two Sample t-test
##
## data:  bodytemps_df$temp[bodytemps_df$gender == 1] and bodytemps_df$temp[bodytemps_df
$gender == 2]
## t = -2.2854, df = 127.51, p-value = 0.02394
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.53964856 -0.03881298
## sample estimates:
## mean of x mean of y
##  98.10462  98.39385
```

```
# Mann Whitney for temperature
wilcox.test(bodytemps_df$temp[bodytemps_df$gender == 1], bodytemps_df$temp[bodytemps_df
$gender == 2])
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  bodytemps_df$temp[bodytemps_df$gender == 1] and bodytemps_df$temp[bodytemps_df
$gender == 2]
## W = 1637, p-value = 0.02676
## alternative hypothesis: true location shift is not equal to 0
```

```
# t test for rate
t.test(bodytemps_df$rate[bodytemps_df$gender == 1], bodytemps_df$rate[bodytemps_df$gende
r == 2])
```

```
##
##  Welch Two Sample t-test
##
## data:  bodytemps_df$rate[bodytemps_df$gender == 1] and bodytemps_df$rate[bodytemps_df
$gender == 2]
## t = -0.63191, df = 116.7, p-value = 0.5287
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.243732  1.674501
## sample estimates:
## mean of x mean of y
##  73.36923  74.15385
```

```
# Mann Whitney for rate
wilcox.test(bodytemps_df$rate[bodytemps_df$gender == 1], bodytemps_df$rate[bodytemps_df
$gender == 2])
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  bodytemps_df$rate[bodytemps_df$gender == 1] and bodytemps_df$rate[bodytemps_df
$gender == 2]
## W = 1927.5, p-value = 0.3898
## alternative hypothesis: true location shift is not equal to 0
```