

22 September, 2020

```
## 0 healthy, 1 has disease
X <- "HHHDDDDHDDHHHHHHHDDHHHHHHHHHHHHDDHHHHDDDDHHHHDDHHHDH" %>%
  stringr::str_split("") %>% unlist() %>% `==`("D") %>% as.integer()

n <- length(X)
n_0 <- n - sum(X)
n_1 <- sum(X)
```

```

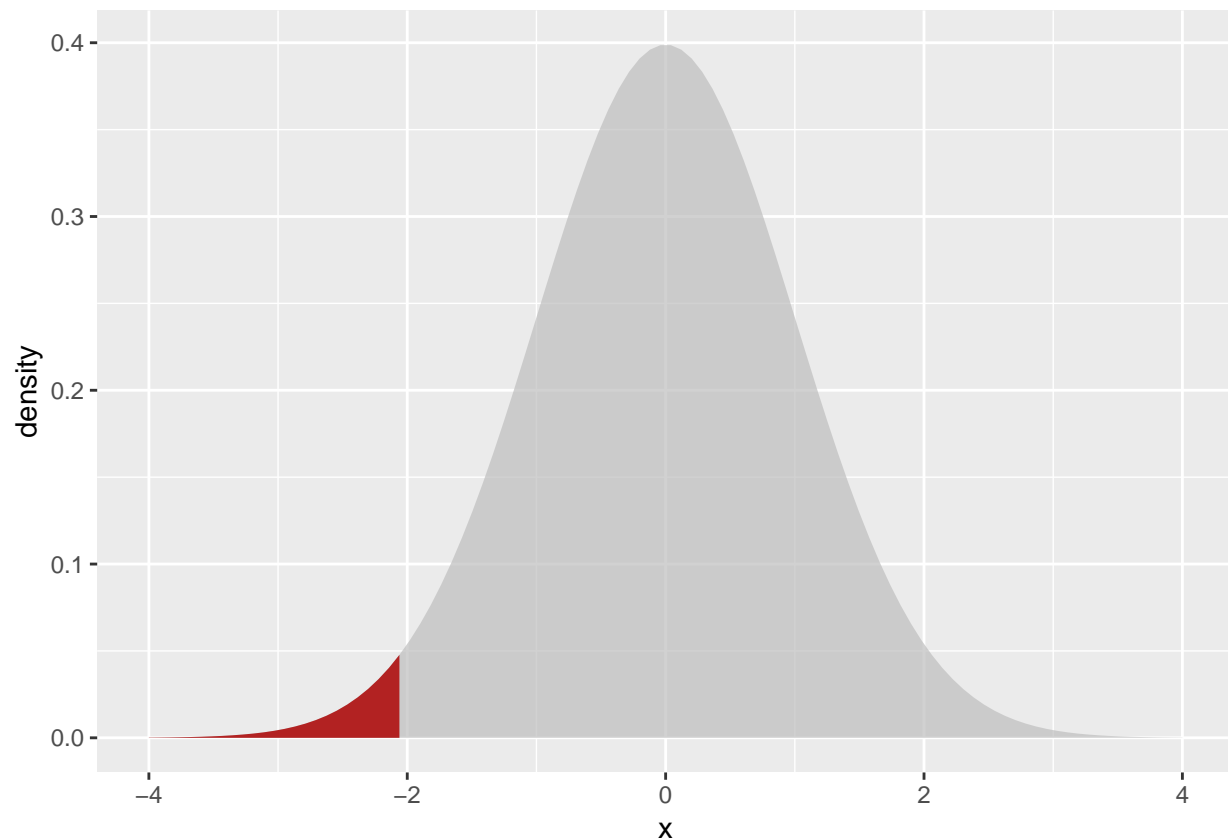
R <- 1 + sum(X[2:n] != X[1:n-1])
R_est <- 1 + 2*n_0*n_1/n
R_var <- (2*n_0*n_1*(2*n_0*n_1 - n))/(n^2*(n - 1))
Z <- (R - R_est)/sqrt(R_var)
p <- pnorm(Z)
cbind(R, R_est, R_var, Z, p)

##      R      R_est      R_var      Z      p
## [1,] 15 20.65957 7.974767 -2.004125 0.02252834

xs <- data.frame(x = seq(-4, 4, length.out = 100)) %>% dplyr::mutate(density = dnorm(x))

ggplot(xs, aes(x, density)) +
  geom_area(fill = "firebrick") +
  gghighlight(x < Z)

```



With  $p = 0.0225$  we reject  $H_0$  at the 5% level. The values are not randomly ordered.

d)