5240 Workshop 06

Digvijay Jondhale 0862899

Q1. What is the population of interest?

Ans : All Grade 8 students who participated in 5km race.

Q2. What is the parameter of interest?

Ans : Proportion of grade 8 students who finished their 5km race under 25 minutes.

Q3. What is the probability of observing 74 or more students finishing the 5km race in under 25 minutes if the normal distribution above is assumed to be correct (this question is asking you to find a -value!).

Ans :
As it is a normal distribution , which means the mean is the central value below which 50% of population would fall. So probability of a student running the race under 25 minutes is 50% or 0.5

- From the given info
  - Ho (null hypothesis) : Only 50% of grade 8 students can run the 5kms race under 25 mins
  - Ha (alternate hypothesis) : More than 50% of grade 8 students can run the 5 kms race under 25 mins.
    - Total possible students : n = 123
    - Favourable students who finished their race before 25 mins : X = 74
    - p = 0.5

```r
n <- 123
p <- 0.5
observed <- 74


mean <- n * p
sd <- sqrt(n * p * (1 - p))


z <- (observed - mean) / sd

p <- pnorm(z, lower.tail = FALSE) # Lower tails is set to false as we want result for right p
p
```

```
[1] 0.01209261
```

- As the p value (0.01) is < 0.05 (significance level), we can say that the result is significant and reject the null hypothesis.
- Hence, the probability of observing 74 or more students finish their 5 kms rae under 25 minutes is 1.2%

Q4. Create a visualization of the -value you calculated in the previous question.

Ans : P value Visualization :

```
library(ggplot2)


x <- seq(mean - 4*sd, mean +  4*sd, length.out = 100)
y <- dnorm(x, mean = mean, sd = sd)


plot01 <- data.frame(x = x, y = y)
ggplot(plot01, aes(x = x, y = y)) +
  geom_line(color = "blue") +
  geom_area(data = subset(plot01, x >= observed), aes(x = x, y = y), fill = "green", alpha =
  geom_vline(xintercept = observed, color = "black", linetype = "dashed") +
  labs(title = "P Value ",
       x = "No. of Students Finishing Race under 25 mins",
       y = "Density") +
  annotate("text", x = observed + 5, y = 0.012, label = paste("P value:", round(p, 3)), colo
  theme_minimal()
```