1.1 The median is (1): 107

1.2 The concentration of DDT, in milligrams per liter, is: (3) a continuous variable

1.3 If the interquartile range is zero, you can conclude that: (3) at least 50% of the observations have the same value

1.4 The species of each insect found in a plot of cropland is: (1) a nominal variable

1.5 Which statement(s) is (are) correct?: (3) One-quarter of the scores are greater than 120.

1.6 Which statement(s) is (are) correct?: (2) The middle 50% of scores are between 70 and 120.

1.7 Identify the correct statement: (3) Both (1) and (2) are correct.

1.8 A measurable characteristic of a population is: (1) a parameter

1.9 (2) All values of the variable have the same value.

1.10 A subset of a population is: (4) a sample

1.11 (2) the distribution is skewed

1.12 The mean number of tickets (rounded to the nearest tenth) is: (2) 2.0

Problem 2

Bad checks - 18, 15, 13, 12, 8, 3, 7, 14, 16, 3

Sample Mean:

Mean = Sum of all observations / Number of observations

Mean = (18 + 15 + 13 + 12 + 8 + 3 + 7 + 14 + 16 + 3) / 10 = 10.9

Sample Median:

Arrange the bad checks in arranging order

Bad checks in ascending order: 3,3,7,8,12,13,14,15,16,18

Median = (12 + 13) / 2 = 12.5

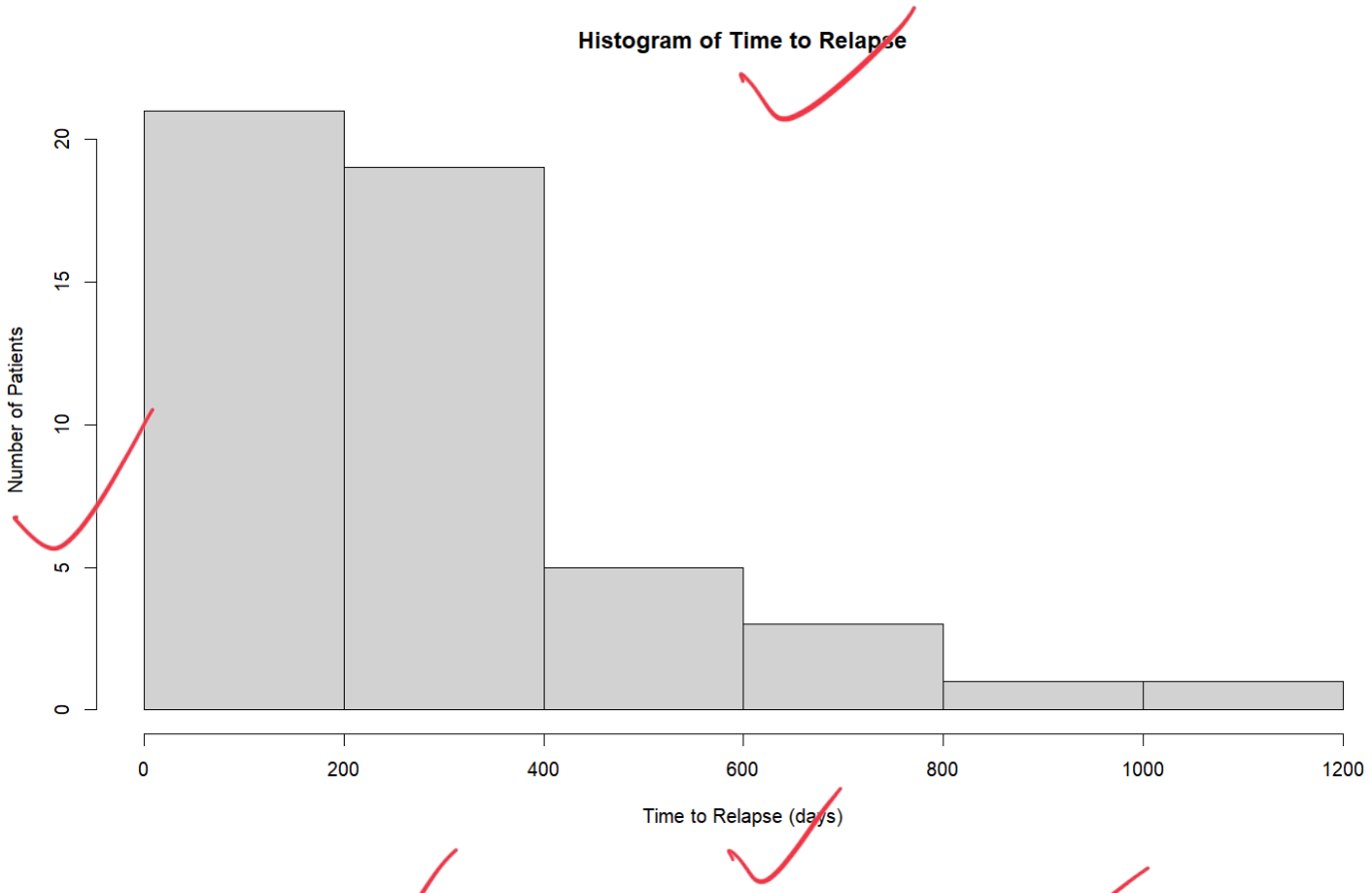
Sample Variance = $\Sigma(x_i - \mu)^2 / (n - 1)$ where μ is mean

Sample Variance = (18 - 10.9) ^ 2 + (12 - 10.9) ^ 2 + ……..+ (3 - 10.9) ^ 2 / (10 - 1) = 28.54444

Sample Standard deviation (s) = sqrt(sample variance) = sqrt(28.54444) = 5.3427

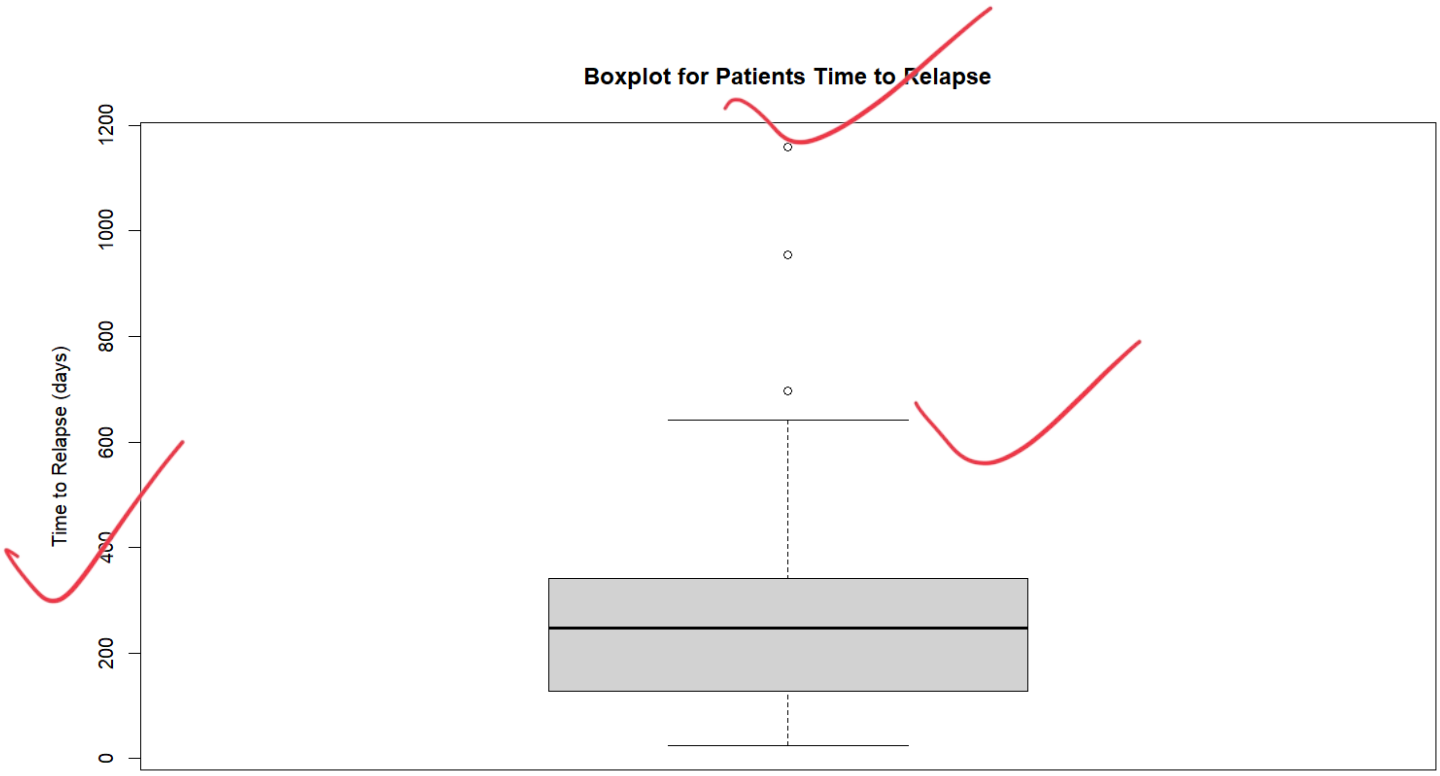
Problem 3)

(1) Histogram using R with function hist with default settings for patients Time to Relapse in days



The distribution is right-skewed, as most relapse times occur at lower values (0 - 400) with tail extending at higher values (1200). Most patients' relapse occurs in the 0 - 200 days range. The data shows a wide range, with relapse times extending from close to 0 days up to 1200 days. Very few relapse in higher values.

(2) Boxplot using R function boxplot for patients time to relapse (days)



From the above boxplot we can say that the distribution is skewed. There appears to be three outliers from the boxplot. The 1st Quartile and 3rd Quartile lie between above 100 and under 400. The median lies between 400 and 200.

(3)
Sample Mean: 287.88
Sample Median: 248
Sample Variance: 53065.33
Sample Standard Deviation: 230.3591
Lower Quartile: 131.75
Upper Quartile: 338.75

(4)
Interquartile Range (IQR): Q3 - Q1
Step: Step = 1.5 * IQR
Lower Inner Face: Q1 - step
Upper Inner Face: Q2 + step
Lower outer fence: Q1 - 2 * IQR
Upper outer fence: Q3 + 2 * IQR

Interquartile Range (IQR): 207
Step: 310.5
Lower Inner Fence: -178.75
Upper Inner Fence: 649.25
Lower Outer Fence: -282.25
Upper Outer Fence: 752.75

```r
# R code

# Problem 1
# 1.1 data: 95 87 96 110 150 104 112 110
# sort the data in ascending order
scores <- c(95, 87, 96, 110, 150, 104, 112, 110)
median_value <- median(scores)
print(median_value)


# problem 2
bad_check <- c(3,3,7,8,12,13,14,15,16,18 )

mean(bad_check)
median(bad_check)
sample_var <- var(bad_check)
sample_var
std_dev <- sqrt(sample_var)
std_dev

# problem 3
anl_data <- read.csv("anl.csv", stringsAsFactors = FALSE)
anl_data
str(anl_data)

# histogram
hist(
  anl_data$days,
  main = "Histogram of Time to Relapse",
  xlab = "Time to Relapse (days)",
  ylab = "Number of Patients",

)


# box plot
boxplot(
  anl_data$days,
  main = "Boxplot for Patients Time to Relapse",
  ylab = "Time to Relapse (days)",
  xlab = ""
)

# Calculate summary statistics
mean_value <- mean(anl_data$days)
median_value <- median(anl_data$days)
variance_value <- var(anl_data$days)
```

```
sd_value <- sd(anl_data$days)
quartiles <- quantile(anl_data$days)

# Print the results
cat("Sample Mean:", mean_value, "\n")
cat("Sample Median:", median_value, "\n")
cat("Sample Variance:", variance_value, "\n")
cat("Sample Standard Deviation:", sd_value, "\n")
cat("Lower Quartile:", quartiles["25%"], "\n")
cat("Upper Quartile:", quartiles["75%"], "\n")


# Calculate boxplot values
IQR_value <- IQR(anl_data$days)
step <- 1.5 * IQR_value
lower_inner <- quartiles["25%"] - step
upper_inner <- quartiles["75%"] + step
lower_outer <- quartiles["25%"] - 2 * IQR_value
upper_outer <- quartiles["75%"] + 2 * IQR_value

# Print the results
cat("Interquartile Range (IQR):", IQR_value, "\n")
cat("Step:", step, "\n")
cat("Lower Inner Fence:", lower_inner, "\n")
cat("Upper Inner Fence:", upper_inner, "\n")
cat("Lower Outer Fence:", lower_outer, "\n")
cat("Upper Outer Fence:", upper_outer, "\n")
```