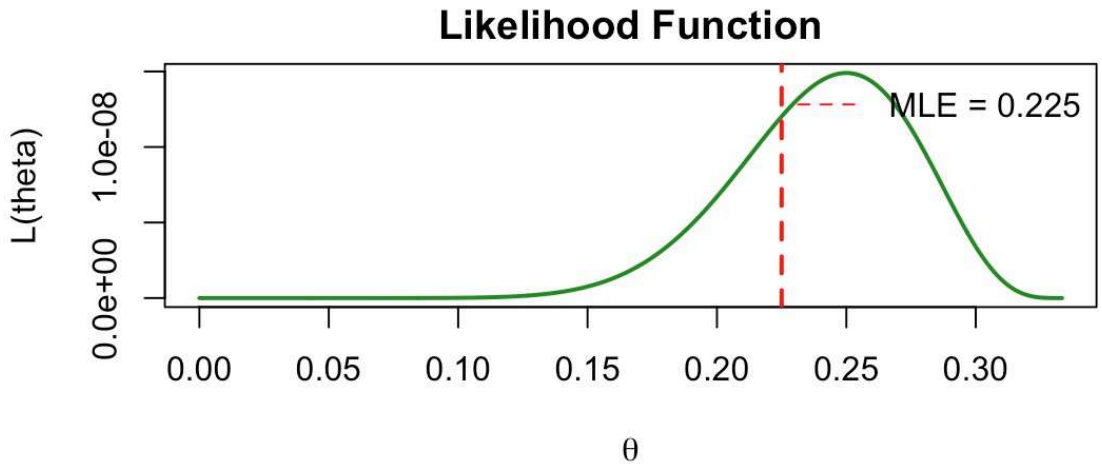
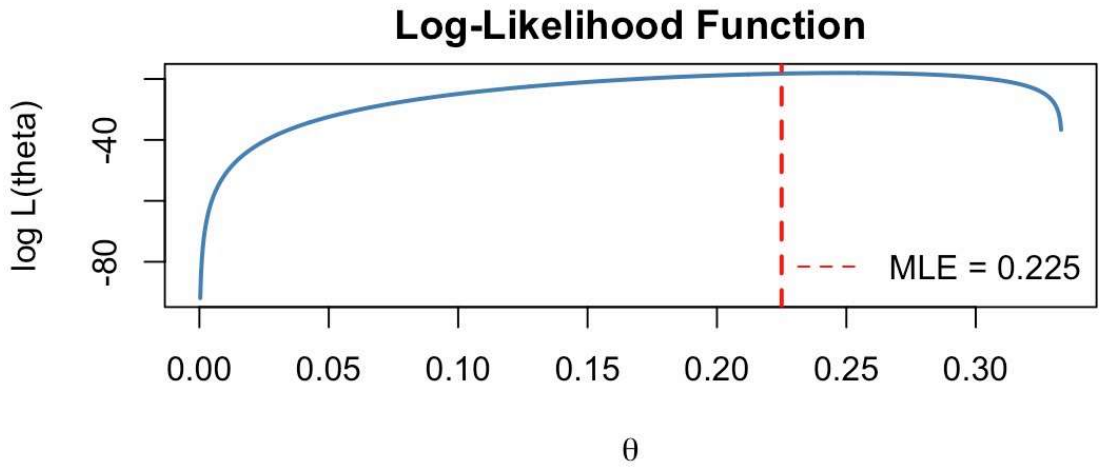
Benjamin Pomianek (100559097), Andrew Ramsahoi (100559088) Leo Dorfman (100558113)

Problem 1:

```
# MLE formula: theta_hat = (2*n1 + n2) / (5*n)
theta.hat <- (2 * n1 + n2) / (5 * n)
cat("MLE of θ =", theta.hat, "\n")  # should print 0.225

# --- Vectorized Likelihood Functions ------------------------------

# P(X=1) = 2θ, P(X=2) = θ, P(X=3) = 1−3θ valid only for θ in [0,1/3]

likelihood <- function(theta) {
  # raw likelihood
  out <- (2*theta)^n1 * (theta)^n2 * (1 - 3*theta)^n3
  # mask invalid θ
  out[ theta < 0 | theta > 1/3 ] <- NA
  out
}

loglikelihood <- function(theta) {
  # raw log-likelihood
  out <- n1*log(2*theta) +
    n2*log(theta) +
    n3*log(1 - 3*theta)
  # mask invalid θ (and avoid log of non-positive)
  out[ theta <= 0 | theta >= 1/3 ] <- NA
  out
}

# --- Plotting -----------------------------------------------------

# You can plot them in one column layout:
par(mfrow = c(2, 1), mar = c(4,4,2,1))

# 1) Likelihood
curve(likelihood, from = 0, to = 1/3, n = 1000,
      xlab = expression(theta), ylab = "L(theta)",
      main = "Likelihood Function", lwd = 2, col = "forestgreen")
abline(v = theta.hat, col = "red", lty = 2, lwd = 2)
legend("topright", legend = paste0("MLE = ", round(theta.hat, 3)),
       lty = 2, col = "red", bty = "n")

# 2) Log-Likelihood
curve(loglikelihood, from = 0, to = 1/3, n = 1000,
      xlab = expression(theta), ylab = "log L(theta)",
      main = "Log-Likelihood Function", lwd = 2, col = "steelblue")
abline(v = theta.hat, col = "red", lty = 2, lwd = 2)
```

```
legend("bottomright", legend = paste0("MLE = ", round(theta.hat, 3)),
       lty = 2, col = "red", bty = "n")
```

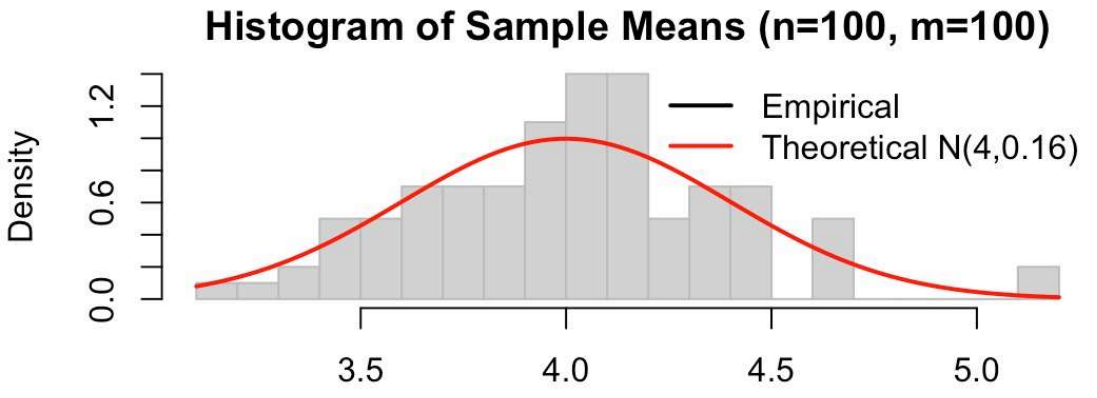


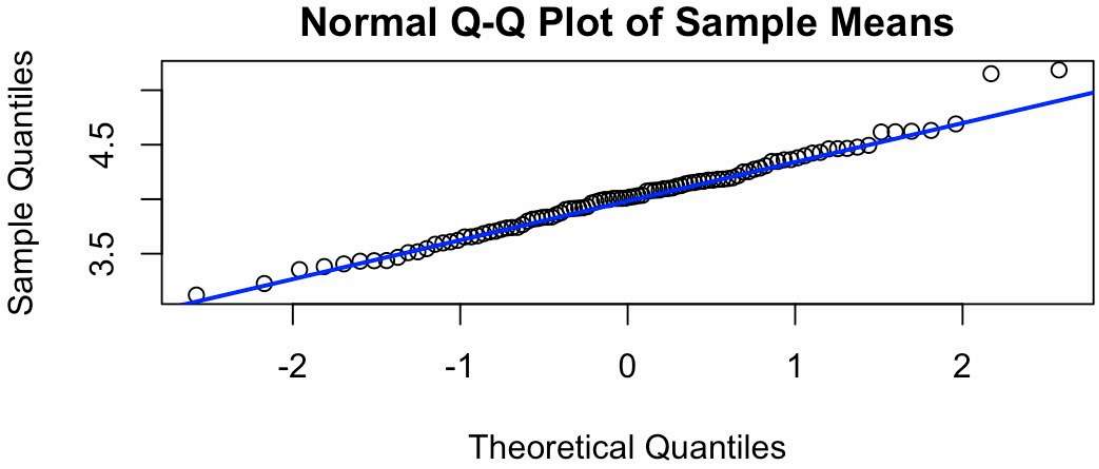


Problem 2;

```
set.seed(123)

m <- 100    # number of samples
n <- 100    # sample size
# simulate m sample-means from Exp(mean=4) => rate = 1/4
sample_means <- replicate(m, mean(rexp(n, rate = 1/4)))

# Theoretical distribution by CLT:
#   mean = 4,  var = Var(X)/n = 16/100 = 0.16,  sd = 0.4
theo_mean <- 4
theo_sd   <- sqrt(16 / n)  # = 0.4

# 1) Histogram + theoretical Normal overlay
hist(sample_means, probability = TRUE, breaks = 15,
     main = "Histogram of Sample Means (n=100, m=100)",
     xlab = "Sample Mean",
     border = "gray")
```

```
curve(dnorm(x, mean = theo_mean, sd = theo_sd),
      col = "red", lwd = 2, add = TRUE)
legend("topright",
       legend = c("Empirical", "Theoretical N(4,0.16)"),
       lwd = 2, col = c("black","red"), bty="n")

# 2) Normal Q–Q plot
qqnorm(sample_means,
       main = "Normal Q-Q Plot of Sample Means")
qqline(sample_means, col = "blue", lwd = 2)

# 3) Goodness-of-fit tests
shapiro_res <- shapiro.test(sample_means)
ks_res      <- ks.test(sample_means,
                       "pnorm",
                       mean = theo_mean,
                       sd   = theo_sd)

cat("Shapiro-Wilk test:\n")
print(shapiro_res)
cat("\nKolmogorov-Smirnov test vs N(4,0.16):\n")
print(ks_res)
```

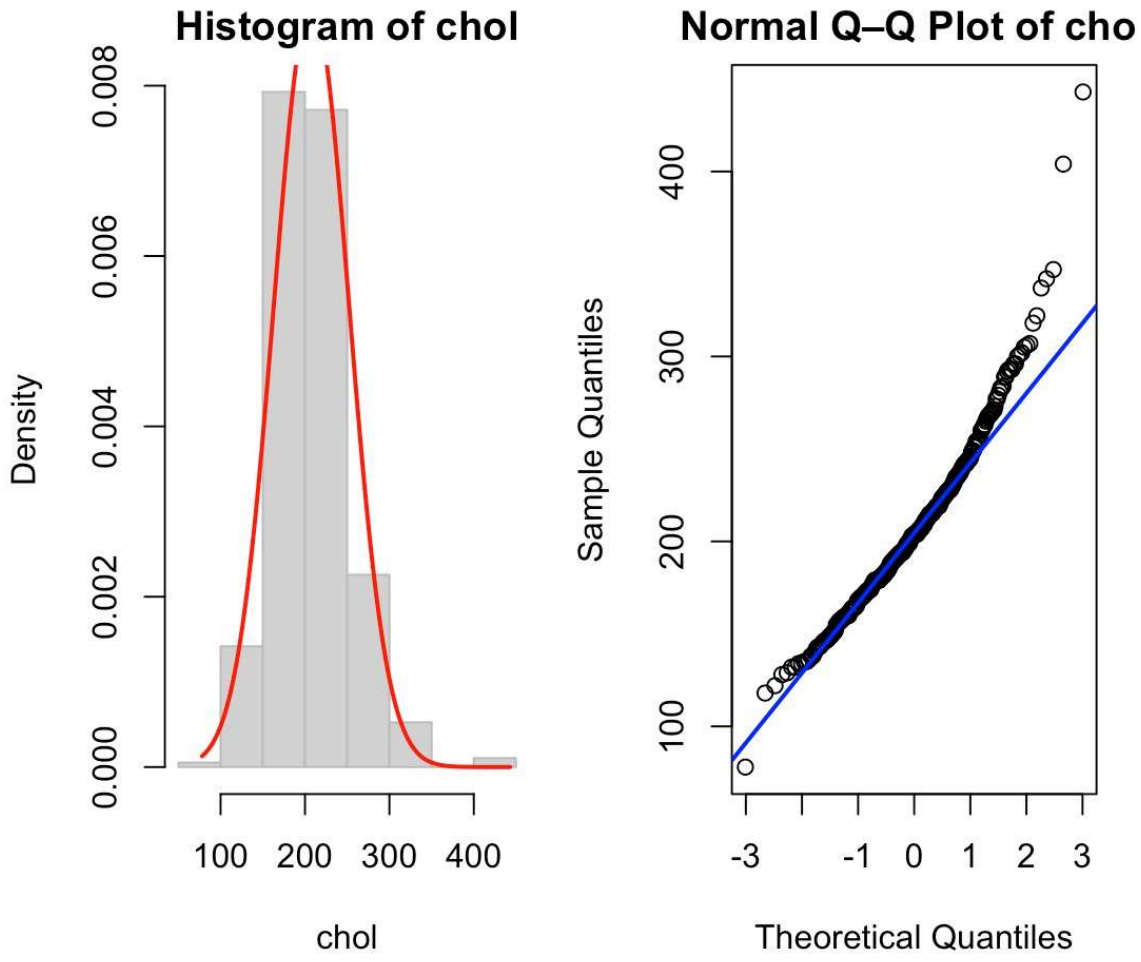




Problem 3:

```
# 1) Extract the cholesterol vector and ensure it's numeric
  chol <- diabetes$chol
  chol <- as.numeric(chol)       # in case it was imported as factor/character
  stopifnot(is.numeric(chol))    # should now be numeric

  # 2) Graphical checks for normality

  # Histogram with overlaid Normal density
  hist(chol,
       probability = TRUE,
       main        = "Histogram of chol",
       xlab        = "chol",
       border      = "gray")
  # add Normal curve
  x.seq <- seq(min(chol), max(chol), length.out = 100)
  lines(x.seq,
        dnorm(x.seq, mean = mean(chol), sd = sd(chol)),
        col = "red", lwd = 2)

  # Normal Q–Q plot
```

```
qqnorm(chol, main = "Normal Q–Q Plot of chol")
qqline(chol, col = "blue", lwd = 2)


# 3) Shapiro–Wilk test for normality
shapiro_res <- shapiro.test(chol)
print(shapiro_res)


# 4) 95% two–sided confidence interval for the mean
ci_res <- t.test(chol, conf.level = 0.95)
print(ci_res)


# 5) One–sided test H0: μ = 210 vs H1: μ > 210 at α = 0.02
mu0   <- 210
alpha <- 0.02
tt_res <- t.test(chol,
                 mu          = mu0,
                 alternative = "greater",
                 conf.level  = 1 - alpha)   # 98% one–sided CI
print(tt_res)


# 6) Conclusion
if (tt_res$p.value < alpha) {
  cat("\nConclusion: p-value =", round(tt_res$p.value,4),
      "<", alpha,
      "⇒ reject H0. Evidence suggests mean(chol) > 210.\n")
} else {
  cat("\nConclusion: p-value =", round(tt_res$p.value,4),
      "≥", alpha,
      "⇒ do not reject H0. No strong evidence mean(chol) exceeds 210.\n")
}
```

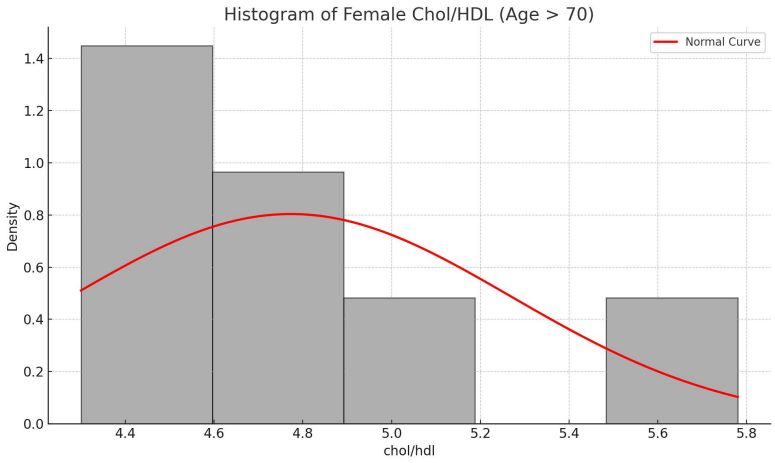


QUESTION 4

```
# 1) Load and prepare data
diabetes <- read.table(".../diabetes.txt", header = TRUE)
diabetes$chol_hdl <- as.numeric(diabetes$chol / diabetes$hdl)

# 2) Filter for males and females over 70
male70 <- subset(diabetes, gender == "male" & age > 70)$chol_hdl
female70 <- subset(diabetes, gender == "female" & age > 70)$chol_hdl

# 3) Normality check (females > 70)
hist(female70, probability = TRUE, main = "Histogram of Female Chol/HDL (Age > 70)", xlab
= "chol/hdl", border = "gray")
x_seq <- seq(min(female70), max(female70), length.out = 100)
lines(x_seq, dnorm(x_seq, mean = mean(female70), sd = sd(female70)), col = "red", lwd =
2)

qqnorm(female70, main = "Q-Q Plot of Female Chol/HDL (Age > 70)")
qqline(female70, col = "blue", lwd = 2)

shapiro.test(female70)
```

# 4) Two-sample t-test for male vs female
t.test(male70, female70, alternative = "two.sided", conf.level = 0.99)

#5 p-value = 0.0069 < α = 0.01 ⇒ reject H 0. There is strong evidence that the mean cholesterol/HDL ratio differs between men and women over 70.



QUESTION 5
# 1) Extract hemoglobin data by gender
female_hba1c <- subset(diabetes, gender == "female")$glyhb
male_hba1c <- subset(diabetes, gender == "male")$glyhb

# 2) Number of females with glyhb > 7.0
x_female <- sum(female_hba1c > 7.0)
n_female <- length(female_hba1c)

# 3) 90% CI for female proportion
prop.test(x = x_female, n = n_female, conf.level = 0.90)

# 4) Compare with males
x_male <- sum(male_hba1c > 7.0)
n_male <- length(male_hba1c)

prop.test(x = c(x_male, x_female), n = c(n_male, n_female), alternative = "two.sided", conf.level = 0.90)

#5) p-value = 0.0838 < α = 0.1 ⇒ reject H_0.

There is moderate evidence that the proportion of individuals with glyhb > 7.0 differs between males and females.