# SGupta_S540_Exam1_v1

## Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see https://quarto.org.

## Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
# ----- 1. Likelihood and Graphical Assessment -----
# Load libraries and data
library(dplyr)
```

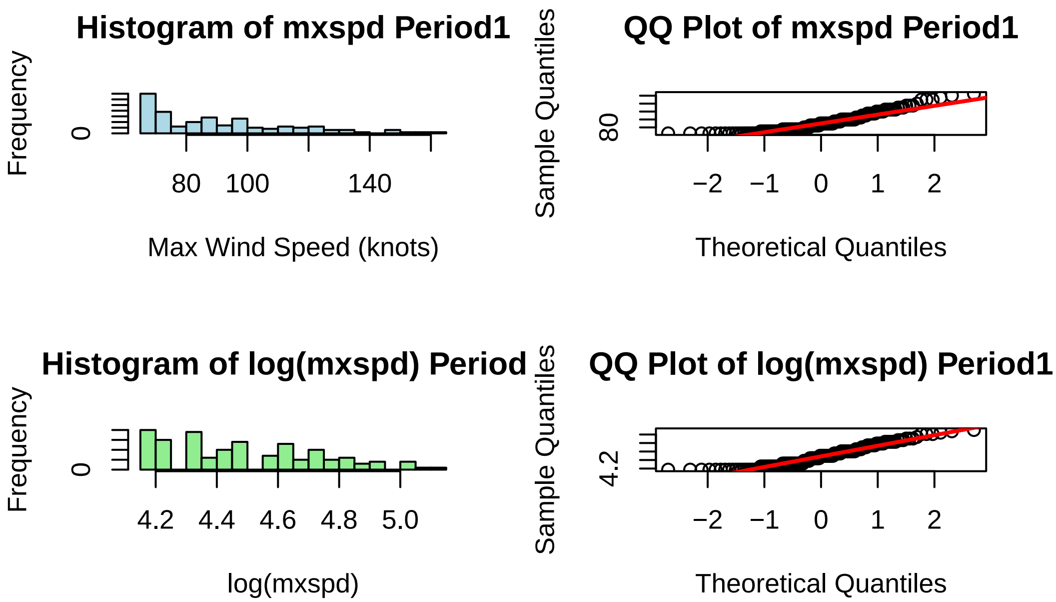
```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

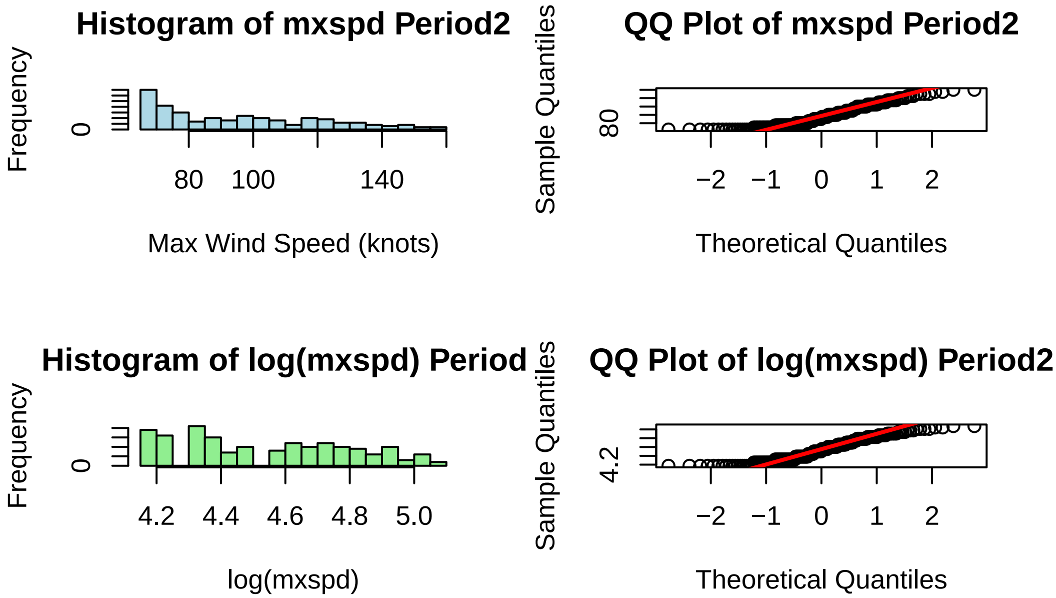
```
data("storms")

# Filter data to hurricanes (wind speed >= 65 knots, non-missing category)
storms <- storms %>% filter(!is.na(category) & wind >= 65)
head(storms)
```

```
# A tibble: 6 x 13
  name      year month   day  hour   lat  long status    category  wind pressure
  <chr>    <dbl> <dbl> <int> <dbl> <dbl> <dbl> <fct>        <dbl> <int>    <int>
1 Blanche   1975     7    27     6  35.9 -70   hurricane        1    65      987
2 Blanche   1975     7    27    12  36.9 -69   hurricane        1    70      984
3 Blanche   1975     7    27    18  37.9 -68   hurricane        1    75      981
4 Blanche   1975     7    28     0  39.3 -67.2 hurricane        1    75      980
5 Blanche   1975     7    28     6  41.2 -66.4 hurricane        1    70      980
6 Caroline  1975     8    30     0  23.3 -94.2 hurricane        1    65      990
# i 2 more variables: tropicalstorm_force_diameter <int>,
#   hurricane_force_diameter <int>
```

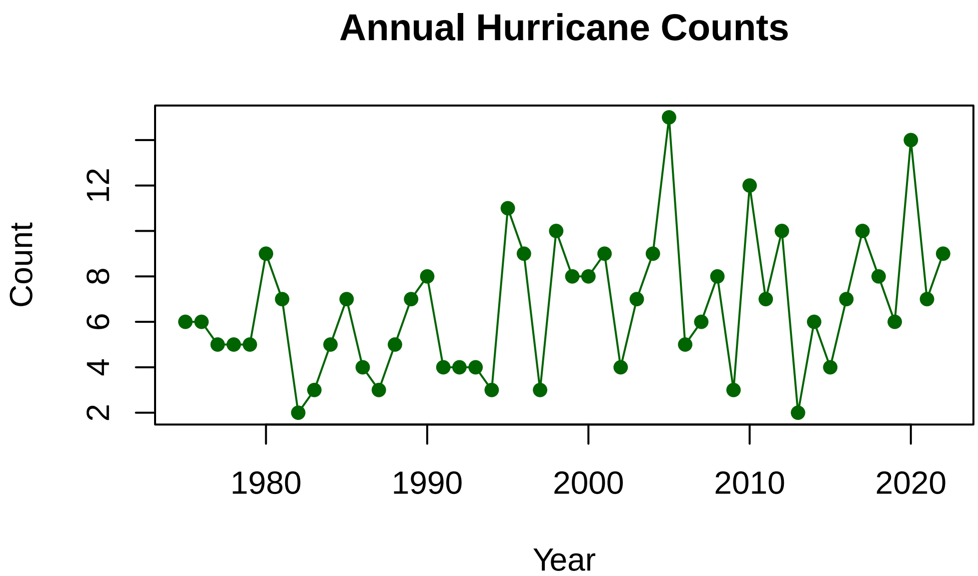
```
# Create unique storm identifier and compute max wind speed and storm year
storms <- storms %>% mutate(storm_id = paste0(name, "_", year))
storm_summary <- storms %>%
  group_by(storm_id) %>%
  summarize(year_storm = min(year), mxspd = max(wind)) %>%
  ungroup()

# Compute annual hurricane counts
year_counts <- storm_summary %>%
  group_by(year_storm) %>%
  summarize(cnt = n()) %>%
  ungroup() %>%
  rename(year = year_storm)

# Split into periods:
#   Period1: 1975-1999, Period2: 2000-2021
storm_summary <- storm_summary %>%
  mutate(period = ifelse(year_storm < 2000, "Period1", "Period2"))
year_counts <- year_counts %>%
  mutate(period = ifelse(year < 2000, "Period1", "Period2"))

# (a) Maximum wind speed:
#    We treat mxspd as continuous. Visualize raw mxspd and its log-transform.
par(mfrow = c(2,2))
for(p in c("Period1", "Period2")) {
  raw_data <- storm_summary %>% filter(period == p) %>% pull(mxspd)
  log_data <- log(raw_data)

  # Plot histogram and QQ plot of raw data
  hist(raw_data, main = paste("Histogram of mxspd", p),
```

```
      xlab = "Max Wind Speed (knots)", col = "lightblue", breaks = 20)
  qqnorm(raw_data, main = paste("QQ Plot of mxspd", p)); qqline(raw_data, col = "red", lwd =

  # Plot histogram and QQ plot of log-transformed data
  hist(log_data, main = paste("Histogram of log(mxspd)", p),
      xlab = "log(mxspd)", col = "lightgreen", breaks = 20)
  qqnorm(log_data, main = paste("QQ Plot of log(mxspd)", p)); qqline(log_data, col = "red",
}
```






```
par(mfrow = c(1,1))

# (b) Hurricane counts:
#    We assume Y ~ Poisson( ) for annual counts.
plot(year_counts$year, year_counts$cnt, type="o",
     main = "Annual Hurricane Counts",
     xlab = "Year", ylab = "Count", col = "darkgreen", pch = 16)
```




You can add options to executable code like this

```
# --------------------------------------------------------------
# 2. Conjugate Uninformative Prior Distributions
# --------------------------------------------------------------
#
# For Maximum Wind Speed (log-transformed):
#
#   Likelihood: log(mxspd_i) ~ N( ,  ²)
#
#   Priors (vague, conjugate):
#        |  ² ~ N( ,  ²/ )   where   = mean(all_log_mxspd),   = 0.001
#       ²     ~ Inverse-Gamma( ,  )   with   = 0.001,   = 0.001
#
# For Hurricane Counts:
#
#   Likelihood: Y ~ Poisson( )
#
#   Prior:
#       ~ Gamma(a , b ) with a  = 0.01 and b  = 0.01

# Create log-transformed mxspd variable for analysis
storm_summary <- storm_summary %>% mutate(log_mxspd = log(mxspd))
```

```r
all_log_mxspd <- storm_summary$log_mxspd
mu0 <- mean(all_log_mxspd)
tau0 <- 0.001;
alpha0 <- 0.001;
beta0 <- 0.001
a0_pois <- 0.01;
b0_pois <- 0.01
```

You can add options to executable code like this

```r
# ------------------------------------------------------------
# 3. Posterior Summaries and Testing for Difference Between Periods
# ------------------------------------------------------------
#
# For log(mxspd):
#
#   Posterior (conjugate for Normal likelihood):
#       =   + n
#       = (   + n·x̄)/( +n)
#       =   + n/2
#       =   + 0.5*(n-1)*s² + ( ·n*(x̄- )²)/(2*( +n))
#
# We compute posterior samples for Period1 and Period2 and then derive the difference.
#
# For Hurricane Counts:
#
#   Posterior:
#      | Y ~ Gamma(a  + ΣY, b  + n)
#
# We similarly derive posterior samples for each period and compute the difference.

# Function to compute posterior parameters for Normal model (log(mxspd))
compute_posterior <- function(data, mu0, tau0, alpha0, beta0) {
  n <- length(data)
  sample_mean <- mean(data)
  sample_var <- var(data)
  tau_n <- tau0 + n
  mu_n <- (tau0 * mu0 + n * sample_mean) / tau_n
  alpha_n <- alpha0 + n/2
  beta_n <- beta0 + 0.5*(n - 1)*sample_var + (tau0 * n * (sample_mean - mu0)^2) / (2 * tau_n
  list(mu_n = mu_n, tau_n = tau_n, alpha_n = alpha_n, beta_n = beta_n, n = n)
}
```

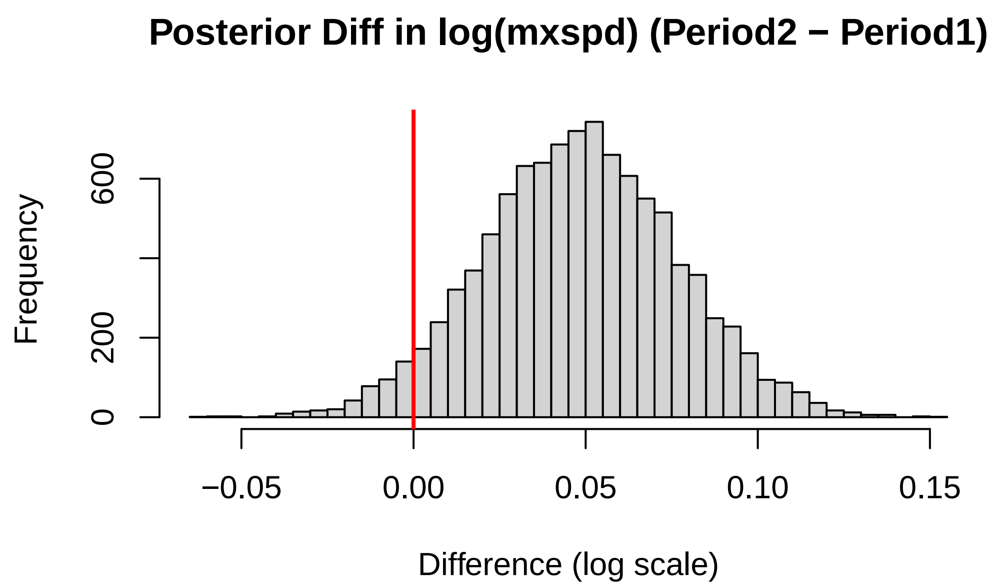
```r
# Compute posterior parameters for each period for log(mxspd)
posterior_p1 <- compute_posterior(
  storm_summary %>% filter(period == "Period1") %>% pull(log_mxspd),
  mu0, tau0, alpha0, beta0)
posterior_p2 <- compute_posterior(
  storm_summary %>% filter(period == "Period2") %>% pull(log_mxspd),
  mu0, tau0, alpha0, beta0)

# Draw posterior samples (Gibbs-style sampling)
set.seed(123)
n_samples <- 10000
samples_p1 <- numeric(n_samples)
samples_p2 <- numeric(n_samples)
for (i in 1:n_samples) {
  sigma2_p1 <- 1 / rgamma(1, shape = posterior_p1$alpha_n, rate = posterior_p1$beta_n)
  samples_p1[i] <- rnorm(1, mean = posterior_p1$mu_n, sd = sqrt(sigma2_p1 / posterior_p1$tau_

  sigma2_p2 <- 1 / rgamma(1, shape = posterior_p2$alpha_n, rate = posterior_p2$beta_n)
  samples_p2[i] <- rnorm(1, mean = posterior_p2$mu_n, sd = sqrt(sigma2_p2 / posterior_p2$tau_
}

# Posterior difference for log(mxspd)
diff_samples <- samples_p2 - samples_p1
point_estimate_diff <- mean(diff_samples)
std_error_diff <- sd(diff_samples)
cat("Posterior difference in mean log(mxspd) (Period2 - Period1):", point_estimate_diff, "\n"
```

```
Posterior difference in mean log(mxspd) (Period2 - Period1): 0.04846583
```

```r
cat("Standard error:", std_error_diff, "\n")
```

```
Standard error: 0.02797023
```

```r
# Layman's interpretation (example): "On average, storms in Period2 have a higher log wind sp

# Plot histogram of the posterior difference (log(mxspd))
hist(diff_samples, breaks = 50, main = "Posterior Diff in log(mxspd) (Period2 - Period1)",
     xlab = "Difference (log scale)", col = "lightgray")
abline(v = 0, col = "red", lwd = 2)
```



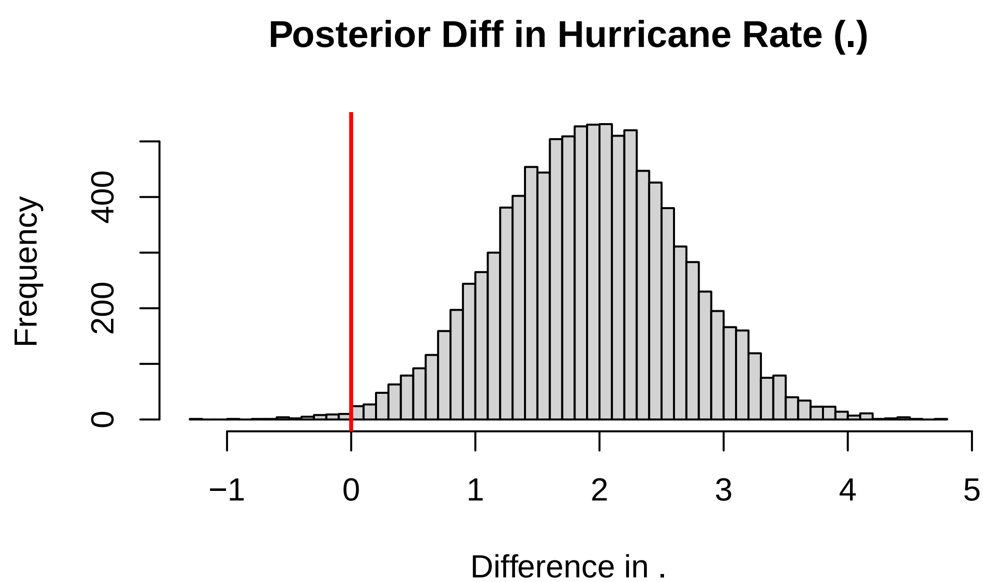

```r
# For hurricane counts:
# Function to compute posterior parameters for Poisson likelihood
compute_poisson_posterior <- function(counts, a0, b0) {
  a_post <- a0 + sum(counts)
  b_post <- b0 + length(counts)
  list(a_post = a_post, b_post = b_post)
}

posterior_cnt_p1 <- compute_poisson_posterior(
  year_counts %>% filter(period == "Period1") %>% pull(cnt), a0_pois, b0_pois)
posterior_cnt_p2 <- compute_poisson_posterior(
  year_counts %>% filter(period == "Period2") %>% pull(cnt), a0_pois, b0_pois)

# Draw posterior samples for
samples_lambda_p1 <- rgamma(n_samples, shape = posterior_cnt_p1$a_post, rate = posterior_cnt_
samples_lambda_p2 <- rgamma(n_samples, shape = posterior_cnt_p2$a_post, rate = posterior_cnt_

# Posterior difference for hurricane counts
diff_lambda <- samples_lambda_p2 - samples_lambda_p1
point_estimate_lambda <- mean(diff_lambda)
std_error_lambda <- sd(diff_lambda)
cat("Posterior difference in   (Period2 - Period1):", point_estimate_lambda, "\n")
```

```
Posterior difference in  (Period2 - Period1): 1.920952
```

```
cat("Standard error:", std_error_lambda, "\n")
```

```
Standard error: 0.7451057
```

```
# Plot histogram of the posterior difference in
hist(diff_lambda, breaks = 50, main = "Posterior Diff in Hurricane Rate ( )",
     xlab = "Difference in  ", col = "lightgray")
abline(v = 0, col = "red", lwd = 2)
```




You can add options to executable code like this

```
# -------------------------------------------------------------
# 4. Sensitivity Analysis: Check if results are sensitive to priors
# -------------------------------------------------------------
# For log(mxspd), use alternative hyperparameters:
#    _alt = 0.01,  _alt = 0.01,  _alt = 0.01
tau0_alt <- 0.01; alpha0_alt <- 0.01; beta0_alt <- 0.01

posterior_p1_alt <- compute_posterior(
  storm_summary %>% filter(period == "Period1") %>% pull(log_mxspd),
```

```r
  mu0, tau0_alt, alpha0_alt, beta0_alt)
posterior_p2_alt <- compute_posterior(
  storm_summary %>% filter(period == "Period2") %>% pull(log_mxspd),
  mu0, tau0_alt, alpha0_alt, beta0_alt)

samples_p1_alt <- numeric(n_samples)
samples_p2_alt <- numeric(n_samples)
for (i in 1:n_samples) {
  sigma2_p1_alt <- 1 / rgamma(1, shape = posterior_p1_alt$alpha_n, rate = posterior_p1_alt$be
  samples_p1_alt[i] <- rnorm(1, mean = posterior_p1_alt$mu_n, sd = sqrt(sigma2_p1_alt / poste

  sigma2_p2_alt <- 1 / rgamma(1, shape = posterior_p2_alt$alpha_n, rate = posterior_p2_alt$be
  samples_p2_alt[i] <- rnorm(1, mean = posterior_p2_alt$mu_n, sd = sqrt(sigma2_p2_alt / poste
}
diff_samples_alt <- samples_p2_alt - samples_p1_alt
cat("Sensitivity Analysis (log(mxspd)) - Difference (Alt Prior):", mean(diff_samples_alt), "\
```

```
Sensitivity Analysis (log(mxspd)) - Difference (Alt Prior): 0.04877928
```

```r
cat("Sensitivity Analysis (log(mxspd)) - Std Error:", sd(diff_samples_alt), "\n")
```

```
Sensitivity Analysis (log(mxspd)) - Std Error: 0.02852238
```

```r
# For hurricane counts, use alternative prior hyperparameters:
#   a0_pois_alt = 0.1, b0_pois_alt = 0.1
a0_pois_alt <- 0.1; b0_pois_alt <- 0.1

posterior_cnt_p1_alt <- compute_poisson_posterior(
  year_counts %>% filter(period == "Period1") %>% pull(cnt), a0_pois_alt, b0_pois_alt)
posterior_cnt_p2_alt <- compute_poisson_posterior(
  year_counts %>% filter(period == "Period2") %>% pull(cnt), a0_pois_alt, b0_pois_alt)

samples_lambda_p1_alt <- rgamma(n_samples, shape = posterior_cnt_p1_alt$a_post, rate = poster
samples_lambda_p2_alt <- rgamma(n_samples, shape = posterior_cnt_p2_alt$a_post, rate = poster
diff_lambda_alt <- samples_lambda_p2_alt - samples_lambda_p1_alt
cat("Sensitivity Analysis (Hurricane Counts) - Difference (Alt Prior):", mean(diff_lambda_alt
```

```
Sensitivity Analysis (Hurricane Counts) - Difference (Alt Prior): 1.928127
```

```r
cat("Sensitivity Analysis (Hurricane Counts) - Std Error:", sd(diff_lambda_alt), "\n")
```

```
Sensitivity Analysis (Hurricane Counts) - Std Error: 0.7451008
```

You can add options to executable code like this

# ——————————————————————

# 5. Final Technician Summary Statement and Report

# ——————————————————————