

# R Programming Practice Questions

## Conditional Probability, Data Visualization, and Statistical Analysis

**Instructions:** Complete all questions by writing R code. Show your work and include comments explaining your approach.

---

### Part 1: Conditional Probability with Data Frames and Tables

#### Question 1: Survey Data Analysis

Create a data frame with the following survey data about 50 students:

**Tasks:** a) Create a data frame with columns: StudentID, Gender, Subject, GradeLevel

- Include 25 males and 25 females
  - Subjects: Math, Science, Arts (distribute as you think appropriate)
  - Grade levels: HighSchool, College
- b) Convert it to a contingency table showing Gender vs Subject using `table()` c) Use `prop.table()` to calculate conditional probability  $P(\text{Math} \mid \text{Male})$  d) Calculate the conditional probability  $P(\text{Female} \mid \text{Science})$  e) Find  $P(\text{Arts} \mid \text{GradeLevel} = \text{College})$

#### Question 2: Medical Testing Scenario

Create a data frame representing medical test results for 100 patients:

**Tasks:** a) Create a data frame with columns: PatientID, Disease (Present/Absent), TestResult (Positive/Negative)

- Include 20 patients with disease present, 80 without disease
  - Among those with disease: 18 test positive, 2 test negative
  - Among those without disease: 8 test positive, 72 test negative
- b) Create a contingency table for Disease vs TestResult c) Use `prop.table()` to calculate the sensitivity  $P(\text{Positive Test} \mid \text{Disease Present})$  d) Calculate the specificity  $P(\text{Negative Test} \mid \text{Disease Absent})$  e) Find the positive predictive value  $P(\text{Disease Present} \mid \text{Positive Test})$

#### Question 3: Customer Purchase Behavior

Create a data frame with purchase data for 60 customers:

**Tasks:** a) Create a data frame with columns: CustomerID, AgeGroup (Young/Middle/Senior), PurchaseCategory (Electronics/Clothing/Books)

- Include 20 customers in each age group
  - Distribute purchases across categories as you think realistic b) Make a contingency table for AgeGroup vs PurchaseCategory c) Calculate  $P(\text{Electronics} \mid \text{Young})$  using `prop.table()` d) Find  $P(\text{Young} \mid \text{Electronics})$  e) Compare purchase patterns across age groups using proportions
- 

## Part 2: Data Visualization

### Question 4: Sales Performance Analysis

Create a dataset with sales data for 3 sales representatives over 4 quarters:

**Tasks:** a) Create a data frame with columns: Rep (Alice, Bob, Carol), Quarter (Q1, Q2, Q3, Q4), Sales

- Manually enter realistic sales figures for each rep per quarter
- Make Alice the top performer, Bob moderate, Carol improving over time b) Make a bar plot showing total sales by representative using `barplot()` c) Create a bar plot showing quarterly sales for each rep d) Add appropriate titles, labels, and colors to your plots e) Make a horizontal bar plot of average quarterly sales by rep

### Question 5: Student Grade Distribution

Create grade data for 30 students in 3 subjects:

**Tasks:** a) Create a data frame with columns: StudentID, Subject (Math/Science/English), Grade

- Manually enter grades (60-100) for each student in each subject
- Make Math grades slightly lower on average, Science moderate, English higher b) Make histograms for each subject's grade distribution using `hist()` c) Create box plots comparing grade distributions across subjects using `boxplot()` d) Make a bar plot showing average grades by subject e) Add proper titles, labels, and colors to all visualizations

### Question 6: Temperature Data Analysis

Create temperature data for 4 seasons with 10 days each:

**Tasks:** a) Create a data frame with columns: Day, Season (Spring/Summer/Fall/Winter), Temperature

- Enter realistic temperatures: Spring (60-75°F), Summer (75-90°F), Fall (50-70°F), Winter (30-50°F) b) Create histograms for temperature in each season c) Make box plots showing temperature distribution by season d) Create a bar plot showing average temperature by season e) Use different colors for each season and add proper labels
-

## Part 3: Integration Challenges

### Question 7: Store Sales Analysis

Create sales data for a small store with 3 product categories over 6 months:

**Tasks:** a) Create a data frame with columns: Month, Product (Electronics/Clothing/Books), Sales, Customers

- Enter data for 6 months, 3 products each month (18 rows total) b) Calculate conditional probabilities:  $P(\text{High Sales} \mid \text{Electronics})$ ,  $P(\text{Electronics} \mid \text{Month} = \text{January})$  c) Create bar plots comparing sales by product and by month d) Make box plots showing sales distribution by product category e) Write a brief interpretation of your findings

### Question 8: Survey Data Integration

Combine conditional probability and visualization for a customer satisfaction survey:

**Tasks:** a) Create a data frame with 40 respondents: Age (Young/Old), Product (A/B), Satisfaction (High/Low)

- Distribute data to show some interesting patterns b) Calculate  $P(\text{High Satisfaction} \mid \text{Product A})$ ,  $P(\text{Product A} \mid \text{Young})$  c) Create contingency tables and use `prop.table()` for analysis d) Make bar plots and box plots to visualize relationships e) Summarize your findings about customer satisfaction patterns

---

## Part 4: Final Challenge

### Question 9: Complete Data Analysis Project

Create a comprehensive dataset and analysis:

**Tasks:** a) Create a data frame with student performance data (30 students, 2 subjects, with gender and study method) b) Perform conditional probability analysis using `table()` and `prop.table()` c) Create multiple types of visualizations (bar plots, histograms, box plots) d) Calculate and interpret at least 3 different conditional probabilities e) Write a summary of your findings with supporting visualizations

#### Deliverables:

- Complete R script with all code and comments
  - Brief written interpretation of results
  - All plots should have proper titles, labels, and colors
-

## Submission Guidelines

1. **Code Quality:** Include comments explaining your logic
2. **Manual Data Entry:** Create all data manually using `c()` and `data.frame()`
3. **Visualization:** Ensure all plots have proper titles, labels, and legends
4. **Analysis:** Provide brief interpretations of your conditional probability findings
5. **Format:** Submit as R script files (.R) with clear section headers

## Key Functions to Use

- `data.frame()` - Create data frames
- `c()` - Create vectors
- `table()` - Create contingency tables
- `prop.table()` - Calculate proportions
- `barplot()` - Create bar plots
- `hist()` - Create histograms
- `boxplot()` - Create box plots

## Grading Criteria

- **Correctness (40%):** Code runs without errors and produces expected results
- **Completeness (25%):** All parts of each question attempted
- **Code Style (15%):** Clean, well-commented, readable code
- **Visualization Quality (20%):** Professional-looking plots with proper formatting

---

**Due Date:** [Insert your due date] **Total Points:** 100 points