Statistical Model Statements with Solutions

Statistical Model Statements

1. Suppose I have data on the heights (in inches) of individuals from 100 different families. I'm interested in modeling the relationship between parental heights (mother and father) on the height (in inches) of their offspring. Write out a statistical model to explore the relationship between offspring height as the dependent variable and mother and father heights as the independent variables.

2. Suppose I have data on the performance ratings (on a scale from 0 to 100) of employees in a company, including their years of experience, education level, and number of training hours completed. Write out a statistical model to explore the relationship between employee performance ratings as the dependent variable and years of experience, education level, and training hours as the independent variables.

3. Suppose I have information on the house prices (in dollars) in a neighborhood, along with their size (in square feet), number of bedrooms, and distance from the city center (in miles). Write out a statistical model to explore the relationship between house prices as the dependent variable and house size, number of bedrooms, and distance from the city center as the independent variables.

4. Suppose I have data on the number of defects found in manufactured products, along with the temperature (in Celsius) during production and the humidity level (in percentage). Write out a statistical model to explore the relationship between defect counts as the dependent variable and temperature and humidity as the independent variables.

5. Suppose I have information on the monthly electricity consumption (in kilowatt-hours) of households, along with their income level (in dollars), number of occupants, and average temperature (in Fahrenheit). Write out a statistical model to explore the relationship between electricity consumption as the dependent variable and income level, number of occupants, and average temperature as the independent variables.

6. Suppose I have data on the number of accidents at different intersections, along with road width (in feet), speed limit (in miles per hour), and presence of traffic signals (binary: 0 for absent, 1 for present). Write out

a statistical model to explore the relationship between accident counts as the dependent variable and road width, speed limit, and presence of traffic signals as the independent variables.

7. Suppose I have information on the lifespan (in years) of a certain species of plant, along with soil pH, sunlight exposure (in hours per day), and precipitation levels (in inches). Write out a statistical model to explore the relationship between plant lifespan as the dependent variable and soil pH, sunlight exposure, and precipitation levels as the independent variables.

8. Suppose I have data on the monthly sales revenue (in dollars) of a retail store, along with advertising expenditure (in dollars), promotions (binary: 0 for no promotions, 1 for promotions), and store location (categorical: urban, suburban, rural). Write out a statistical model to explore the relationship between sales revenue as the dependent variable and advertising expenditure, promotions, and store location as the independent variables.

9. Suppose I have information on the performance scores (on a scale from 0 to 100) of athletes, including their age (in years), diet composition (categorical: balanced, high-protein, high-carb), hours of training (in hours per week), and sleep duration (in hours per night). Write out a statistical model to explore the relationship between athletic performance scores as the dependent variable and age, diet composition, hours of training, and sleep duration as the independent variables.

10. Suppose I have data on the scores of students in a standardized test (ranging from 0 to 100), along with their study hours (in hours per week), socioeconomic status (categorical: low, medium, high), and access to tutoring (binary: 0 for no tutoring, 1 for tutoring). Write out a statistical model to explore the relationship between test scores as the dependent variable and study hours, socioeconomic status, and access to tutoring as the independent variables.

11. Suppose I have information on the monthly sales volume (in units) of a product, along with price changes (in dollars), competitor actions (binary: 0 for no action, 1 for action), and seasonality (categorical: winter, spring, summer, fall). Write out a statistical model to explore the relationship between sales volume as the dependent variable and price changes, competitor actions, and seasonality as the independent variables.

12. Suppose I have data on the number of website visitors, along with marketing expenditure (in dollars), website design changes (binary: 0 for no changes, 1 for changes), and social media activity (measured by engagement rate). Write out a statistical model to explore the relationship between website traffic as the dependent variable and marketing expenditure, website design changes, and social media activity as the independent variables.

13. Suppose I have information on the course performance (on a scale from 0 to 100) of students, including their attendance (in percentage), participation in class discussions (in number of times participated), completion of assignments (binary: 0 for incomplete, 1 for complete), and midterm scores (on a scale from 0 to 100). Write out a statistical model to explore the relationship between course performance as the dependent variable and attendance, participation, assignment completion, and midterm scores as the independent variables.

14. Suppose I have data on the duration of customer service calls (in minutes), along with customer satisfaction ratings (on a scale from 0 to 100), wait times (in minutes), and issue resolution time (in minutes). Write out a statistical model to explore the relationship between call duration as the dependent variable and customer satisfaction ratings, wait times, and issue resolution time as the independent variables.

15. Suppose I have information on the number of patients treated in a hospital, along with staffing levels (number of staff members), bed capacity (number of beds), and availability of medical equipment (binary: 0 for insufficient, 1 for sufficient). Write out a statistical model to explore the relationship between patient volume as the dependent variable and staffing levels, bed capacity, and availability of medical equipment as the independent variables.

16. Suppose I have data on the number of software bugs reported, along with developer experience (in years), coding standards adherence (binary: 0 for non-adherence, 1 for adherence), and project complexity (measured by lines of code). Write out a statistical model to explore the relationship between bug counts as the dependent variable and developer experience, coding standards adherence, and project complexity as the independent variables.

17. Suppose I have information on the performance of stocks in a portfolio (measured by returns), including industry exposure (categorical: technology, healthcare, finance, etc.), company size (categorical: small, medium, large), and financial metrics (e.g., P/E ratio, EPS). Write out a statistical model to explore the relationship between portfolio performance as the dependent variable and industry exposure, company size, and financial metrics as the independent variables.

18. Suppose I have data on the frequency of medical appointments (number of appointments per month), along with patient demographics (age, gender), appointment scheduling method (categorical: online, phone, in-person), and travel distance (in miles). Write out a statistical model to explore the relationship between appointment frequency as the dependent variable and patient demographics, scheduling method, and travel distance as the independent variables.

19. Suppose I have information on the number of votes received by political candidates, along with campaign spending (in dollars), media coverage (measured by airtime), and voter demographics (age, gender, ethnicity). Write out a statistical model to explore the relationship between vote counts as the dependent variable and campaign spending, media coverage, and voter demographics as the independent variables.

20. Suppose I have data on the number of software licenses sold, along with software features (number of features), customer support quality (measured by response time), and pricing strategy (categorical: low, medium, high). Write out a statistical model to explore the relationship between license sales as the dependent variable and software features, customer support quality, and pricing strategy as the independent variables.

Solutions

1. $$\text{offspring\_height}_i \sim \text{Normal}(\mu_i, \sigma^2),$$
$$\mu_i = \beta_0 + \beta_1 \text{mother\_height}_i + \beta_2 \text{father\_height}_i.$$

2. $$\text{employee\_performance}_i \sim \text{Normal}(\mu_i, \sigma^2),$$
$$\mu_i = \beta_0 + \beta_1 \text{years\_of\_experience}_i + \beta_2 \text{education\_level}_i + \beta_3 \text{training\_hours}_i.$$

3. $$\text{house\_price}_i \sim \text{Normal}(\mu_i, \sigma^2),$$
$$\mu_i = \beta_0 + \beta_1 \text{house\_size}_i + \beta_2 \text{number\_of\_bedrooms}_i + \beta_3 \text{distance\_from\_city}_i.$$

4. $$\text{defect\_count}_i \sim \text{Poisson}(\lambda_i),$$
$$\log(\lambda_i) = \beta_0 + \beta_1 \text{temperature}_i + \beta_2 \text{humidity}_i.$$

5. $$\text{accident\_count}_i \sim \text{Poisson}(\lambda_i),$$
$$\log(\lambda_i) = \beta_0 + \beta_1 \text{road\_width}_i + \beta_2 \text{speed\_limit}_i + \beta_3 \text{traffic\_signals}_i.$$

6. $$\text{athletic\_performance}_i \sim \text{Normal}(\mu_i, \sigma^2),$$
$$\mu_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{diet\_composition}_i + \beta_3 \text{training\_hours}_i + \beta_4 \text{sleep\_duration}_i.$$

7. $$\text{test\_scores}_i \sim \text{Normal}(\mu_i, \sigma^2),$$
$$\mu_i = \beta_0 + \beta_1 \text{study\_hours}_i + \beta_2 \text{socieconomic\_status}_i + \beta_3 \text{tutoring\_access}_i.$$

8.
$$\text{website\_traffic}_i \sim \text{Poisson}(\lambda_i),$$
$$\log(\lambda_i) = \beta_0 + \beta_1 \text{marketing\_expenditure}_i + \beta_2 \text{website\_design\_changes}_i + \beta_3 \text{social\_media\_activity}_i.$$

9.
$$\text{course\_performance}_i \sim \text{Normal}(\mu_i, \sigma^2),$$
$$\mu_i = \beta_0 + \beta_1 \text{attendance}_i + \beta_2 \text{participation}_i + \beta_3 \text{assignment\_completion}_i + \beta_4 \text{midterm\_scores}_i.$$

10.
$$\text{patient\_volume}_i \sim \text{Poisson}(\lambda_i),$$
$$\log(\lambda_i) = \beta_0 + \beta_1 \text{staffing\_levels}_i + \beta_2 \text{bed\_capacity}_i + \beta_3 \text{medical\_equipment\_availability}_i.$$

11.
$$\text{bug\_count}_i \sim \text{Poisson}(\lambda_i),$$
$$\log(\lambda_i) = \beta_0 + \beta_1 \text{developer\_experience}_i + \beta_2 \text{coding\_standards\_adherence}_i + \beta_3 \text{project\_complexity}_i.$$

12.
$$\text{portfolio\_performance}_i \sim \text{Normal}(\mu_i, \sigma^2),$$
$$\mu_i = \beta_0 + \beta_1 \text{industry\_exposure}_i + \beta_2 \text{company\_size}_i + \beta_3 \text{financial\_metrics}_i.$$

13.
$$\text{appointment\_frequency}_i \sim \text{Poisson}(\lambda_i),$$
$$\log(\lambda_i) = \beta_0 + \beta_1 \text{patient\_demographics}_i + \beta_2 \text{appointment\_scheduling\_method}_i + \beta_3 \text{travel\_distance}_i.$$

14.
$$\text{vote\_count}_i \sim \text{Poisson}(\lambda_i),$$
$$\log(\lambda_i) = \beta_0 + \beta_1 \text{campaign\_spending}_i + \beta_2 \text{media\_coverage}_i + \beta_3 \text{voter\_demographics}_i.$$