Day 2 Problem Set Solutions

Methods Camp | University of Texas at Austin

Casey Breen

Day 2 Problem Set

This document contains solutions to all exercises from the Day 2 problem set. To reinforce your understanding, please try your best to complete these exercises independently before referring to the solutions.

Exercise 1: Random Variables & Expectation

A survey asks respondents how many times they attended religious services last month. Suppose the distribution is:

Times attended (x)	0	1	2	4
$P(X = x)$	0.4	0.3	0.2	0.1

1.1 Is X discrete or continuous? Why?

1.2 Compute $E[X]$ by hand.

1.3 In your own words, what does $E[X]$ tell us?

1.4 If attendance is recoded as $Y = 2X + 1$, what is $E[Y]$? (Use the linearity property.)

Solutions

1.1 X is **discrete**: it only takes on a countable set of values (0, 1, 2, 4), not any value in a continuous interval.

1.2 $E[X] = 0(0.4) + 1(0.3) + 2(0.2) + 4(0.1) = 0 + 0.3 + 0.4 + 0.4 = 1.1$

1.3 $E[X] = 1.1$ is the long-run average across many respondents — if we surveyed thousands of people and averaged their attendance, we'd land close to 1.1.

1.4 Using linearity, $E[aX+b] = aE[X]+b$: $\quad E[Y] = E[2X+1] = 2E[X]+1 = 2(1.1)+1 = 3.2$

Exercise 2: Distributions

2.1 For each scenario below, name the most appropriate distribution (Bernoulli, Binomial, Poisson, or Normal) and briefly justify your choice:

a. Whether a single respondent owns a home (yes/no)

b. The number of home-owners among 200 randomly sampled respondents

c. The number of hate-crime incidents reported in a city per month

d. Adult height in a large population

2.2 For scenario (b), if $p = 0.65$, what is $E[X]$?

Solutions

2.1 (a) Bernoulli — a single yes/no trial for one respondent. (b) Binomial — the count of "successes" (home-owners) across $n = 200$ independent Bernoulli trials, each with the same probability p. (c) Poisson — a count of relatively rare events occurring over a fixed interval (a month), without a natural fixed number of "trials." (d) Normal — a continuous measurement that tends to cluster symmetrically around an average, the classic bell-curve shape.

2.2 Using $E[X] = np$ for a Binomial: $E[X] = 200 \times 0.65 = 130$

Exercise 3: Regression

Suppose a researcher wants to examine occupational prestige score (Y) as a linear function of their parents' years of schooling (X) across all U.S. families.

3.1 Write the regression equation using index notation. What does the subscript i represent?

3.2 Interpret the slope coefficient.

3.3 Identify the estimand, a plausible estimator, and an example of an estimate.

3.4 Name one plausible factor that could be included in the error term, ε_i.

Solutions

3.1 $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. The subscript i identifies a particular adult child or family.

3.2 β_1 represents the expected difference in the child's occupational prestige score associated with one additional year of parental schooling.

3.3 **Estimand:** the population slope β_1; **Estimator:** ordinary least squares (OLS); **Estimate:** the value obtained from the sample, such as $\hat{\beta}_1 = 1.5$.

3.4 Parental income, neighborhood conditions, or the local labor market.