
8.6.0 End of Chapter Problems

Problem 1

Let X be the weight of a randomly chosen individual from a population of adult men. In order to estimate the mean and variance of X, we observe a random sample X_1,X_2,$\cdots$,X_{10}. Thus, the X_i's are i.i.d. and have the same distribution as X. We obtain the following values (in pounds):

$$165.5,\ 175.4,\ 144.1,\ 178.5,\ 168.0,\ 157.9,\ 170.1,\ 202.5,\ 145.5,\ 135.7$$

Find the values of the sample mean, the sample variance, and the sample standard deviation for the observed sample.

Problem 2

Let X_1, X_2, X_3, $\ldots$, X_n be a random sample with unknown mean $EX_i = \mu$, and unknown variance $\mathrm{Var}(X_i) = \sigma^2$. Suppose that we would like to estimate $\theta = \mu^2$. We define the estimator $\hat{\Theta}$ as

$$\hat{\Theta} = \left(\overline{X}\right)^2 = \left[\frac{1}{n}\sum_{k=1}^{n} X_k\right]^2$$

to estimate θ. Is $\hat{\Theta}$ an unbiased estimator of θ? Why?

Problem 3

Let X_1, X_2, X_3, $\ldots$, X_n be a random sample from the following distribution

$$f_X(x) = \begin{cases} \theta\left(x - \frac{1}{2}\right) + 1 & \text{for } 0 \leq x \leq 1 \\ \\ 0 & \text{otherwise} \end{cases}$$

where $\theta \in [-2, 2]$ is an unknown parameter. We define the estimator $\hat{\Theta}_n$ as

$$\hat{\Theta}_n = 12\overline{X} - 6$$

to estimate θ.

a. Is $\hat{\Theta}_n$ an unbiased estimator of θ?
b. Is $\hat{\Theta}_n$ a consistent estimator of θ?
c. Find the mean squared error (MSE) of $\hat{\Theta}_n$.

Problem 4

Let $X_1, \ldots, X_4$ be a random sample from a $Geometric(p)$ distribution. Suppose we observed (x_1, x_2, x_3, x_4) = $(2, 3, 3, 5)$. Find the likelihood function using $P_{X_i}(x_i; p) = p(1-p)^{x_i - 1}$ as the PMF.

Problem 5

Let $X_1, \ldots, X_4$ be a random sample from an $Exponential(\theta)$ distribution. Suppose we observed (x_1, x_2, x_3, x_4) = $(2.35, 1.55, 3.25, 2.65)$. Find the likelihood function using

$$f_{X_i}(x_i; \theta) = \theta e^{-\theta x_i}, \quad \text{for } x_i \geq 0$$

as the PDF.

Problem 6

Often when working with maximum likelihood functions, out of ease we maximize the log-likelihood rather than the likelihood to find the maximum likelihood estimator. Why is maximizing $L(\mathbf{x}; \theta)$ as a function of θ equivalent to maximizing log $L(\mathbf{x}; \theta)$?

Problem 7

Let X be one observation from a $N(0, \sigma^2)$ distribution.

a. Find an unbiased estimator of σ^2.
b. Find the log likelihood, log$(L(x; \sigma^2))$, using

$$f_X(x;\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$

as the PDF.

c. Find the Maximum Likelihood Estimate (MLE) for the standard deviation σ, $\hat{\sigma}_{ML}$.

Problem 8

Let $X_1, \ldots, X_n$ be a random sample from a $Poisson(\lambda)$ distribution.

a. Find the likelihood equation, $L(x_1, \ldots, x_n; \lambda)$, using

$$P_{X_i}(x_1, \ldots, x_n; \lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

as the PMF.

b. Find the log likelihood function and use that to obtain the MLE for λ, $\hat{\lambda}_{ML}$.

Problem 9

In this problem, we would like to find the CDFs of the order statistics. Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with CDF $F_X(x)$ and PDF $f_X(x)$. Define $X_{(1)}, \ldots, X_{(n)}$ as the order statistics and show that

$$F_{X_{(i)}}(x) = \sum_{k=i}^{n} \binom{n}{k} \big[F_X(x)\big]^k \big[1 - F_X(x)\big]^{n-k}.$$

Hint: Fix $x \in \mathbb{R}$. Let Y be a random variable that counts the number of X_j's $\leq x$. Define $\{X_j \leq x\}$ as a "success" and $\{X_j > x\}$ as a "failure," and show that $Y \sim Binomial(n, p = F_X(x))$.

Problem 10

In this problem, we would like to find the PDFs of order statistics. Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with CDF $F_X(x)$ and PDF $f_X(x)$. Define $X_{(1)}, \ldots, X_{(n)}$ as the order statistics. Our goal here is to show that

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} f_X(x) \big[F_X(x)\big]^{i-1} \big[1 - F_X(x)\big]^{n-i}.$$

One way to do this is to differentiate the CDF (found in Problem 9). However, here, we would like to derive the PDF directly. Let $f_{X_{(i)}}(x)$ be the PDF of $X_{(i)}$. By definition of the PDF, for small δ, we can write

$$f_{X_{(i)}}(x)\delta \approx P(x \leq X_{(i)} \leq x+\delta)\delta.$$

Note that the event $\{x \leq X_{(i)} \leq x+\delta\}$ occurs if $i-1$ of the X_j's are less than x, one of them is in $[x, x+\delta]$, and $n-i$ of them are larger than $x+\delta$. Using this, find $f_{X_{(i)}}(x)$.

Hint: Remember the multinomial distribution. More specifically, suppose that an experiment has 3 possible outcomes, so the sample space is given by

$$S = \{s_1, s_2, s_3\}.$$

Also, suppose that $P(s_i) = p_i$ for $i = 1, 2, 3$. Then for $n = n_1 + n_2 + n_3$ independent trials of this experiment, the probability that each s_i appears n_i times is given by

$$\binom{n}{n_1, n_2, n_3} p_1^{n_1} p_2^{n_2} p_3^{n_3} = \frac{n!}{n_1! n_2! n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}.$$

Problem 11

A random sample X_1, X_2, X_3, ..., X_{100} is given from a distribution with known variance $\mathrm{Var}(X_i) = 81$. For the observed sample, the sample mean is $\overline{X} = 50.1$. Find an approximate 95% confidence interval for $\theta = EX_i$.

Problem 12

To estimate the portion of voters who plan to vote for Candidate A in an election, a random sample of size n from the voters is chosen. The sampling is done with replacement. Let θ be the portion of voters who plan to vote for Candidate A among all voters.

a. How large does n need to be so that we can obtain a 90% confidence interval with 3% margin of error?

b. How large does n need to be so that we can obtain a 99% confidence interval with 3% margin of error?

Problem 13

Let X_1, X_2, X_3, $\ldots$, X_{100} be a random sample from a distribution with unknown variance $\mathrm{Var}(X_i) = \sigma^2 < \infty$. For the observed sample, the sample mean is $\overline{X} = 110.5$, and the sample variance is $S^2 = 45.6$. Find a 95% confidence interval for $\theta = EX_i$.

Problem 14

A random sample X_1, X_2, X_3, $\ldots$, X_{36} is given from a normal distribution with unknown mean $\mu = EX_i$ and unknown variance $\mathrm{Var}(X_i) = \sigma^2$. For the observed sample, the sample mean is $\overline{X} = 35.8$, and the sample variance is $S^2 = 12.5$.

a. Find and compare 90%, 95%, and 99% confidence interval for μ.
b. Find and compare 90%, 95%, and 99% confidence interval for σ^2.

Problem 15

Let X_1, X_2, X_3, X_4, X_5 be a random sample from a $N(\mu, 1)$ distribution, where μ is unknown. Suppose that we have observed the following values

$$5.45, \quad 4.23, \quad 7.22, \quad 6.94, \quad 5.98$$

We would like to decide between

H_0: $\mu = \mu_0 = 5$,

H_1: $\mu \neq 5$.

a. Define a test statistic to test the hypotheses and draw a conclusion assuming $\alpha = 0.05$.
b. Find a 95% confidence interval around $\overline{X}$. Is μ_0 included in the interval? How does the exclusion of μ_0 in the interval relate to the hypotheses we are testing?

Problem 16

Let $X_1, \ldots, X_9$ be a random sample from a $N(\mu, 1)$ distribution, where μ is unknown. Suppose that we have observed the following values

$$16.34, \quad 18.57, \quad 18.22, \quad 16.94, \quad 15.98, \quad 15.23, \quad 17.22, \quad 16.54, \quad 17.54$$

We would like to decide between

H_0: $\mu = \mu_0 = 16$,

H_1: $\mu \neq 16$.

a. Find a 90% confidence interval around $\overline{X}$. Is μ_0 included in the interval? How does this relate to our hypothesis test?
b. Define a test statistic to test the hypotheses and draw a conclusion assuming $\alpha = 0.1$.

Problem 17

Let X_1, X_2 ,..., X_{150} be a random sample from an unknown distribution. After observing this sample, the sample mean and the sample variance are calculated to be

$$\overline{X} = 52.28, \quad S^2 = 30.9$$

Design a level 0.05 test to choose between

H_0: $\mu = 50$,

H_1: $\mu > 50$.

Do you accept or reject H_0?

Problem 18

Let X_1, X_2, X_3, X_4, X_5 be a random sample from a $N(\mu, \sigma^2)$ distribution, where μ and σ are both unknown. Suppose that we have observed the following values

$$27.72, \quad 22.24, \quad 32.86, \quad 19.66, \quad 35.34$$

We would like to decide between

H_0: $\mu \geq 30$,

H_1: $\mu < 30$.

Assuming $\alpha = 0.05$, what do you conclude?

Problem 19

Let X_1, X_2 ,..., X_{121} be a random sample from an unknown distribution. After observing this sample, the sample mean and the sample variance are calculated to be

$$\overline{X} = 29.25, \quad S^2 = 20.7$$

Design a test to decide between

H_0: $\mu = 30$,

H_1: $\mu < 30$,

and calculate the P-value for the observed data.

Problem 20

Suppose we would like to test the hypothesis that at least 10% of students suffer from allergies. We collect a random sample of 225 students and 21 of them suffer from allergies.

a. State the null and alternative hypotheses.
b. Obtain a test statistic and a P-value.
c. State the conclusion at the $\alpha = 0.05$ level.

Problem 21

Consider the following observed values of (x_i, y_i):

$$(-5, -2), \quad (-3, 1), \quad (0, 4), \quad (2, 6), \quad (1, 3).$$

a. Find the estimated regression line

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

based on the observed data.

b. For each x_i, compute the fitted value of y_i using

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

c. Compute the residuals, $e_i = y_i - \hat{y}_i$.

d. Calculate R-squared.

Problem 22

Consider the following observed values of (x_i, y_i):

$$(1, 3), \quad (3, 7).$$

a. Find the estimated regression line

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

based on the observed data.

b. For each x_i, compute the fitted value of y_i using

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

c. Compute the residuals, $e_i = y_i - \hat{y}_i$.

d. Calculate R-squared.

e. Explain the above results. In particular, can you conclude that the obtained regression line is a good model here?

Problem 23
Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where ϵ_i's are independent $N(0, \sigma^2)$ random variables. Therefore, Y_i is a normal random variable with mean $\beta_0 + \beta_1 x_i$ and variance σ^2. Moreover, Y_i's are independent. As usual, we have the observed data pairs (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) from which we would like to estimate β_0 and β_1. In this chapter, we found the following estimators

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}},$$
$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{x}.$$

where

$$s_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2,$$
$$s_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y}).$$

a. Show that $\hat{\beta}_1$ is a normal random variable.
b. Show that $\hat{\beta}_1$ is an unbiased estimator of β_1, i.e.,

$$E[\hat{\beta}_1] = \beta_1.$$

c. Show that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{s_{xx}}.$$

Problem 24
Again consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where ϵ_i's are independent $N(0, \sigma^2)$ random variables, and

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}},$$
$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{x}.$$

a. Show that $\hat{\beta}_0$ is a normal random variable.
b. Show that $\hat{\beta}_0$ is an unbiased estimator of β_0, i.e.,

$$E[\hat{\beta}_0] = \beta_0.$$

c. For any $i = 1, 2, 3, \ldots, n$, show that

$$\text{Cov}(\hat{\beta}_1, Y_i) = \frac{x_i - \overline{x}}{s_{xx}} \sigma^2.$$

d. Show that

$$\text{Cov}(\hat{\beta}_1, \overline{Y}) = 0.$$

e. Show that

$$\text{Var}(\hat{\beta}_0) = \frac{\sum_{i=1}^{n} x_i^2}{n s_{xx}} \sigma^2.$$
