## Exercises

**9.1.** Estimate the unknown parameter $\theta$ from a sample

$$3, 3, 3, 3, 3, 7, 7, 7$$

drawn from a discrete distribution with the probability mass function

$$\begin{cases} P(3) &= \theta \\ P(7) &= 1-\theta \end{cases}.$$

Compute two estimators of $\theta$:

(a) the method of moments estimator;

(b) the maximum likelihood estimator.

Also,

(c) Estimate the standard error of each estimator of $\theta$.

**9.2.** The number of times a computer code is executed until it runs without errors has a Geometric distribution with unknown parameter $p$. For 5 independent computer projects, a student records the following numbers of runs:

$$3 \quad 7 \quad 5 \quad 3 \quad 2$$

Estimate $p$

(a) by the method of moments;

(b) by the method of maximum likelihood.

**9.3.** Use method of moments and method of maximum likelihood to estimate

(a) parameters $a$ and $b$ if a sample from Uniform$(a, b)$ distribution is observed;

(b) parameter $\lambda$ if a sample from Exponential$(\lambda)$ distribution is observed;

(c) parameter $\mu$ if a sample from Normal$(\mu, \sigma)$ distribution is observed, and we already know $\sigma$;

(d) parameter $\sigma$ if a sample from Normal$(\mu, \sigma)$ distribution is observed, and we already know $\mu$;

(e) parameters $\mu$ and $\sigma$ if a sample from Normal$(\mu, \sigma)$ distribution is observed, and both $\mu$ and $\sigma$ are unknown.

**9.4.** A sample of 3 observations ($X_1 = 0.4,\ X_2 = 0.7,\ X_3 = 0.9$) is collected from a continuous distribution with density

$$f(x) = \begin{cases} \theta x^{\theta-1} & \text{if } \ 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Estimate $\theta$ by *your favorite method.*

**9.5.** A sample $(X_1, ..., X_{10})$ is drawn from a distribution with a probability density function

$$\frac{1}{2}\left(\frac{1}{\theta}e^{-x/\theta} + \frac{1}{10}e^{-x/10}\right), \quad 0 < x < \infty$$

The sum of all 10 observations equals 150.

(a) Estimate $\theta$ by the method of moments.

(b) Estimate the standard error of your estimator in (a).

**9.6.** Verify columns 3-5 in Table 9.1 on p. 273. Section 9.4.7 will help you with the last row of the table.

**9.7.** In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation $\sigma = 9.2$.

(a) Construct a 90% confidence interval for the expectation of the number of concurrent users.

(b) At the 1% significance level, do these data provide significant evidence that the mean number of concurrent users is greater than 35?

**9.8.** Installation of a certain hardware takes random time with a standard deviation of 5 minutes.

(a) A computer technician installs this hardware on 64 different computers, with the average installation time of 42 minutes. Compute a 95% confidence interval for the population mean installation time.

(b) Suppose that the population mean installation time is 40 minutes. A technician installs the hardware on your PC. What is the probability that the installation time will be within the interval computed in (a)?

**9.9.** Salaries of entry-level computer engineers have Normal distribution with unknown mean and variance. Three randomly selected computer engineers have salaries (in \$ 1000s):

$$30, 50, 70$$

(a) Construct a 90% confidence interval for the average salary of an entry-level computer engineer.

(b) Does this sample provide a significant evidence, at a 10% level of significance, that the average salary of all entry-level computer engineers is different from \$80,000? Explain.

(c) Looking at this sample, one may think that the starting salaries have a great deal of variability. Construct a 90% confidence interval for the standard deviation of entry-level salaries.

**9.10.** We have to accept or reject a large shipment of items. For quality control purposes, we collect a sample of 200 items and find 24 defective items in it.

(a) Construct a 96% confidence interval for the proportion of defective items in the whole shipment.

(b) The manufacturer claims that at most one in 10 items in the shipment is defective. At the 4% level of significance, do we have sufficient evidence to disprove this claim? Do we have it at the 15% level?

**9.11.** Refer to Exercise 9.10. Having looked at the collected sample, we consider an alternative supplier. A sample of 150 items produced by the new supplier contains 13 defective items.

Is there significant evidence that the quality of items produced by the new supplier is higher than the quality of items in Exercise 9.10? What is the P-value?

**9.12.** An electronic parts factory produces resistors. Statistical analysis of the output suggests that resistances follow an approximately Normal distribution with a standard deviation of 0.2 ohms. A sample of 52 resistors has the average resistance of 0.62 ohms.

(a) Based on these data, construct a 95% confidence interval for the population mean resistance.

(b) If the actual population mean resistance is exactly 0.6 ohms, what is the probability that an average of 52 resistances is 0.62 ohms or higher?

**9.13.** Compute a P-value for the right-tail test in Example 9.25 on p. 272 and state your conclusion about a significant increase in the number of concurrent users.

**9.14.** Is there significant difference in speed between the two servers in Example 9.21 on p. 263?

(a) Use the confidence interval in Example 9.21 to conduct a two-sided test at the 5% level of significance.

(b) Compute a P-value of the two-sided test in (a).

(c) Is server A really faster? How strong is the evidence? Formulate the suitable hypothesis and alternative and compute the corresponding P-value.

State your conclusions in (a), (b), and (c).

**9.15.** According to Example 9.17 on p. 257, there is no significant difference, at the 5% level, between towns A and B in their support for the candidate. However, the level $\alpha = 0.05$ was chosen rather arbitrarily, and the candidate still does not know if he can trust the results when planning his campaign. Can we compare the two towns at *all* reasonable levels of significance? Compute the P-value of this test and state conclusions.

**9.16.** A sample of 250 items from lot A contains 10 defective items, and a sample of 300 items from lot B is found to contain 18 defective items.

(a) Construct a 98% confidence interval for the difference of proportions of defective items.

(b) At a significance level $\alpha = 0.02$, is there a significant difference between the quality of the two lots?

**9.17.** A news agency publishes results of a recent poll. It reports that candidate A leads candidate B by 10% because 45% of the poll participants supported Ms. A whereas only 35% supported Mr. B. What margin of error should be reported for each of the listed estimates, 10%, 35%, and 45%? Notice that 900 people participated in the poll, and the reported margins of error typically correspond to 95% confidence intervals.

**9.18.** Consider the data about the number of blocked intrusions in Exercise 8.1, p. 233.

(a) Construct a 95% confidence interval for the difference between the average number of intrusion attempts per day before and after the change of firewall settings (assume equal variances).

(b) Can we claim a significant reduction in the rate of intrusion attempts? The number of intrusion attempts each day has approximately Normal distribution. Compute P-values and state your conclusions under the assumption of equal variances and without it. Does this assumption make a difference?

**9.19.** Consider two populations ($X$'s and $Y$'s) with different means but the same variance. Two independent samples, sizes $n$ and $m$, are collected. Show that the pooled variance estimator

$$s_p^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2}{n+m-2}$$

estimates their common variance unbiasedly.

**9.20.** A manager questions the assumptions of Exercise 9.8. Her pilot sample of 40 installation times has a sample standard deviation of $s = 6.2$ min, and she says that it is significantly different from the assumed value of $\sigma = 5$ min. Do you agree with the manager? Conduct the suitable test of a standard deviation.

**9.21.** If $[a, b]$ is a $(1-\alpha)100\%$ confidence interval for the population variance (with $a \geq 0$), prove that $[\sqrt{a}, \sqrt{b}]$ is a $(1-\alpha)100\%$ confidence interval for the population standard deviation.

**9.22.** Recall Example 9.21 on p. 263, in which samples of 30 and 20 observations produced standard deviations of 0.6 min and 1.2 min, respectively. In this Example, we assumed unequal variances and used the suitable method only because the reported sample standard deviations *seemed* too different.

(a) Argue for or against the chosen method by testing equality of the population variances.

(b) Also, construct a 95% confidence interval for the ratio of the two population variances.

**9.23.** Anthony says to Eric that he is a stronger student because his average grade for the first six quizzes is higher. However, Eric replies that he is more stable because the variance of his grades is lower. The actual scores of the two friends (presumably, independent and normally distributed) are in the table.

| | Quiz 1 | Quiz 2 | Quiz 3 | Quiz 4 | Quiz 5 | Quiz 6 |
|---|---|---|---|---|---|---|
| Anthony | 85 | 92 | 97 | 65 | 75 | 96 |
| Eric | 81 | 79 | 76 | 84 | 83 | 77 |

(a) Is there significant evidence to support Anthony's claim? State $H_0$ and $H_A$. Test equality of variances and choose a suitable two-sample t-test. Then conduct the test and state conclusions.

(b) Is there significant evidence to support Eric's claim? State $H_0$ and $H_A$ and conduct the test.

For each test, use the 5% level of significance.

**9.24.** Recall Exercise 9.23. Results essentially show that a sample of six quizzes was too small for Anthony to claim that he is a stronger student. We realize that each student has his own population of grades, with his own mean $\mu_i$ and variance $\sigma_i^2$. The observed quiz grades are sampled from this population, and they are different due to all the uncertainty and random factors involved when taking a quiz. Let us estimate the population parameters with some confidence.

(a) Construct a 90% confidence interval for the population mean score for each student.

(b) Construct a 90% confidence interval for the difference of population means. If you have not completed Exercise 9.23(a), start by testing equality of variances and choosing the appropriate method.

(c) Construct a 90% confidence interval for the population variance of scores for each student.

(d) Construct a 90% confidence interval for the ratio of population variances.