I3–TD3
(Confidence Interval Based on a Single Sample)

1. Consider the probability statement

$$P\left(-2.81 < Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 2.75\right) = k$$

where $\bar{X}$ is the mean of a random sample of size n from $N(\mu, \sigma^2)$ distribution with known σ^2.

(a) Find k.

(b) Use this statement to find a confidence interval for μ.

(c) What is the confidence level of this confidence interval?

(d) Find a symmetric confidence interval for μ.

2. Let $X_1, X_2, \dots, X_n$ be a random sample from an exponential distribution $\text{Exp}(\lambda)$.

(a) Show that the mgf of $X \sim \text{Exp}(\lambda)$ is $M_X(t) = (1-\lambda t)^{-1}, \qquad t < \frac{1}{\lambda}$.

(b) Let the r.v. $Y = \frac{2}{\lambda}\sum_{i=1}^{n} X_i$. Find the mgf of Y and deduce that $Y \sim \chi^2(2n)$.

(c) Derive a $100(1-\alpha)\%$ CI for λ.

(d) A random sample of 15 heat pumps of a certain type yielded the following observations on lifetime (in years):

2.0	1.3	6.0	1.9	5.1	0.4	1.0	5.3
15.7	0.7	4.8	0.9	12.2	5.3	0.6	

Assume that the lifetime distribution is exponential. What is a 95% CI for the true average and the standard deviation of the lifetime distribution?

3. Let $X_1, ..., X_n$ be a random sample from an $N(\mu, \sigma^2)$, where the value of σ^2 is unknown.

(a) Construct a $(1-\alpha)100\%$ confidence interval for μ when the value of σ^2 is known.

(b) Construct a $(1-\alpha)100\%$ confidence interval for μ when the value of σ^2 is unknown.

4. Consider the next 1000 95% CIs for μ that a statistical consultant will obtain for various clients. Suppose the data sets on which the intervals are based are selected independently of one another. How many of these 1000 intervals do you expect to capture the corresponding value of μ? What is the probability that between 940 and 960 of these intervals contain the corresponding value of μ? [Hint: Let $Y =$ the number among the 1000 intervals that contain μ. What kind of random variable is Y?]

5. A random sample of size 50 from a particular brand of 16-ounce tea packets produced a mean weight of 15.65 ounces. Assume that the weights of these brands of tea packets are normally distributed with standard deviation of 0.59 ounce. Find a 95% confidence interval for the true mean μ.

6. Use the t-table to determine the values of $t_{\alpha/2}$ that would be used in the construction of a confidence interval for a population mean in each of the following cases:

(a) $\alpha = 0.99, n = 20$

(b) $\alpha = 0.95, n = 18$

(c) $\alpha = 0.90, n = 25$

7. In a large university, the following are the ages of 20 randomly chosen employees:

24	31	28	43	28	56	48	39	52	32
38	49	51	49	62	33	41	58	63	56

Assuming that the data come from a normal population, construct a 95% confidence interval for the population mean μ of the ages of the employees of this university. Interpret your answer.

8. A random sample from a normal population yields the following 25 values:

90	87	121	96	106	107	89	107	83	92
117	93	98	120	97	109	78	87	99	79
104	85	91	107	89					

(a) Calculate an unbiased estimate $\hat{\mu}$ of the population mean.

(b) Give approximate 99% confidence interval for the population mean.

9. The following data represent the rates (micrometers per hour) at which a razor cut made in the skin of anesthetized newts is closed by new cells.

28	20	21	39	32	23	18	31	14	23
18	22	28	24	33	12	23	21	25	25

(a) Can we say that the data are approximately normally distributed?

(b) Find a 95% confidence interval for population mean rate μ for the new cells to close a razor cut made in the skin of anesthetized newts.

(c) Find a 99% confidence interval for μ.

(d) Is the 95% CI wider or narrower than the 99% CI? Briefly explain why.

10. Many mutual funds use an investment approach involving owning stocks whose price/earnings multiples (P/Es) are less than the P/E of the S&P 500. The following data give P/Es of 49 companies a randomly selected mutual fund owns in a particular year.

6.8	5.6	8.5	8.5	8.4	7.5	9.3	9.4	7.8	7.1
9.9	9.6	9.0	9.4	13.7	16.6	9.1	10.1	10.6	11.1
8.9	11.7	12.8	11.5	12.0	10.6	11.1	6.4	12.3	12.3
11.4	9.9	14.3	11.5	11.8	13.3	12.8	13.7	13.9	12.9
14.2	14.0	15.5	16.9	18.0	17.9	21.8	18.4	34.3	

Find a 98% confidence interval for the mean P/E multiples. Interpret the result and state any assumptions you have made.

11. Let $X_1, ..., X_n$ be a random sample from an $N(\mu, \sigma^2)$, where the value of σ^2 is unknown.

(a) Construct a $(1-\alpha)100\%$ confidence interval for σ^2, choosing an appropriate pivot. Interpret its meaning.

(b) Suppose a random sample from a normal distribution gives the following summary statistics: $n = 21, \bar{X} = 44.3$, and $s = 3.96$. Using part (a), find a 90% confidence interval for σ^2. Interpret its meaning.

12. A random sample of size 20 is drawn from a population having a normal distribution. The sample mean and the sample standard deviation from the data are given, respectively, as $\bar{X} = -2.2$ and $s = 1.42$. Construct a 90% confidence interval for the population variance σ^2 and interpret.

13. A random sample from a normal population yields the following 25 values:

90	87	121	96	106	107	89	107	83	92
117	93	98	120	97	109	78	87	99	79
104	85	91	107	89					

(a) Calculate an unbiased estimate of the population variance.

(b) Give approximate 99% confidence interval for the population variance.

(c) Interpret your results and state any assumptions you made in order to solve the problem.

14. A survey indicates that it is important to pay attention to truth in political advertising. Based on a survey of 1200 people, 35% indicated that they found political advertisements to be untrue; 60% say that they will not vote for candidates whose advertisements are judged to be untrue; and of this latter group, only 15% ever complained to the media or to the candidate about their dissatisfaction.

(a) Find a 95% confidence interval for the percentage of people who find political advertising to be untrue.

(b) Find a 95% confidence interval for the percentage of voters who will not vote for candidates whose advertisements are considered to be untrue.

(c) Find a 95% confidence interval for the percentage of those who avoid voting for candidates whose advertisements are considered untrue and who have complained to the media or to the candidate about the falsehood in commercials.

(d) For each case above, interpret the results and state any assumptions you have made.

15. In a random sample of 50 college seniors, 18 indicated that they were planning to pursue a graduate degree. Find a 98% confidence interval for the true proportion of all college seniors planning to pursue a graduate degree, and interpret the result, and state any assumptions you have made.

16. In a random sample of 500 items from a large lot of manufactured items, there were 40 defectives.

(a) Find a 90% confidence interval for the true proportion of defectives in the lot.

(b) Is the assumption of normal approximation valid?

(c) Suppose we suspect that another lot has the same proportion of defectives as in the first lot. What should be the sample size if we want to estimate the true proportion within 0.01 with 90% confidence?

17. Let $\bar{X}$ be the mean of a random sample from the exponential distribution, $Exp(\theta), \theta > 0$.

(a) Show that $\bar{X}$ is an unbiased point estimator of θ.

(b) Using the mgf technique determine the distribution of $\bar{X}$.

(c) Use (b) to show that $Y = 2n\bar{X}/\theta$ has a χ^2 distribution with $2n$ degrees of freedom.

(d) Based on Part (c) , find a 95% confidence interval for θ if $n = 10$. Hint: Find c and d such that $P(c < \frac{2n\bar{X}}{\theta} < d) = 0.95$ and solve the inequalities for θ.

18. Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution on the interval $[0, \theta]$, so that

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

Then if $Y = max(X_i)$, it can be shown that the rv $U = Y/\theta$ has density function

$$f_U(u) = \begin{cases} nu^{n-1} & 0 \le u \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a. Use $f_U(u)$ to verify that

$$P\left((\alpha/2)^{1/n} < \frac{Y}{\theta} < (1-\alpha/2)^{1/n}\right) = 1 - \alpha$$

and use this to derive a $100(1-\alpha)\%$ CI for θ.

b. Verify that $P\left(\alpha^{1/n} \le Y/\theta \le 1\right) = 1 - \alpha$, and derive a $100(1-\alpha)\%$ CI for θ based on this probability statement.

c. Which of the two intervals derived previously is shorter? If my waiting time for a morning bus is uniformly distributed and observed waiting times are $x_1 = 4.2, x_2 = 3.5, x_3 = 1.7, x_4 = 1.2$ and $x_5 = 2.4$, derive a 95% CI for θ by using the shorter of the two intervals.

19. Let $X_1, X_2, \ldots, X_n$ be a random sample from a population X with pdf

$$f(x; \theta) = \begin{cases} \dfrac{2x}{\theta^2} & \text{if } 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a) Determine an estimator T_n for θ, using the moment method. Is T_n efficient.

(b) Find an unbiased estimator $\widehat{\theta}_n$ for θ, from the MLE for θ. Which estimator between T_n and $\widehat{\theta}_n$ is better?

(c) Find a 95% CI for θ when $\max(x_1, ..., x_{20}) = 5$.

20. Let $X_1, X_2, \ldots, X_n$ be a random sample from a population X with pdf

$$f(x;\theta) = \begin{cases} \dfrac{x_0^{\frac{1}{\theta}}}{\theta x^{1+\frac{1}{\theta}}} & \text{if } x > x_0 \\ 0 & \text{if } x \le x_0. \end{cases}$$

where $\theta > 0$ is an unknown parameter and $x_0 > 0$.

(a) Find the MLE $\hat{\theta}_n$ for θ. Is $\hat{\theta}_n$ efficient?

(b) Find a 95% CI for θ when $\prod_{i=1}^{14} x_i = 2565^{14}$ and $x_0 = 1900$.

21. Let $X_1, X_2, \ldots, X_n$ be a random sample from a population X with pdf

$$f(x;\theta) = \begin{cases} \dfrac{1}{2\theta\sqrt{x}} e^{-\frac{\sqrt{x}}{\theta}} & \text{if } x > 0 \\ 0 & \text{otherwise .} \end{cases}$$

where $\theta > 0$ is an unknown paramete.

(a) Find the MLE $\hat{\theta}_n$ for θ. Is $\hat{\theta}_n$ efficient?

(b) Find a 90% CI for θ when $\sum_{i=1}^{20} \sqrt{x_i} = 47.4$.

22. A random sample of size $n_1 = 25$, taken from a normal population with a standard deviation $\sigma_1 = 5$, has a mean $\bar{x}_1 = 80$. A second random sample of size $n_2 = 36$, taken from a different normal population with a standard deviation $\sigma_2 = 3$, has a mean $\bar{x}_2 = 75$. Find a 94% confidence interval for $\mu_1 - \mu_2$.

23. Two kinds of thread are being compared for strength. Fifty pieces of each type of thread are tested under similar conditions. Brand A has an average tensile strength of 78.3 kilograms with a standard deviation of 5.6 kilograms, while brand B has an average tensile strength of 87.2 kilograms with a standard deviation of 6.3 kilograms. Construct a 95% confidence interval for the difference of the population means.

24. A study was conducted to determine if a certain treatment has any effect on the amount of metal removed in a pickling operation. A random sample of 100 pieces was immersed in a bath for 24 hours without the treatment, yielding an average of 12.2 millimeters of metal removed and a sample standard deviation of 1.1 millimeters. A second sample of 200 pieces was exposed to the treatment, followed by the 24-hour immersion in the bath, resulting in an average removal of 9.1 millimeters of metal with a sample standard deviation of 0.9 millimeter. Compute a 98% confidence interval estimate for the difference between the population means. Does the treatment appear to reduce the mean amount of metal removed?

25. Two catalysts in a batch chemical process, are being compared for their effect on the output of the process reaction. A sample of 12 batches was prepared using catalyst 1, and a sample of 10 batches was prepared using catalyst 2. The 12 batches for which catalyst 1 was used in the reaction gave an average yield of 85 with a sample standard

deviation of 4, and the 10 batches for which catalyst 2 was used gave an average yield of 81 and a sample standard deviation of 5. Find a 90% confidence interval for the difference between the population means, assuming that the populations are approximately normally distributed with equal variances.

26. Students may choose between a 3-semester-hour physics course without labs and a 4-semester-hour course with labs. The final written examination is the same for each section. If 12 students in the section with labs made an average grade of 84 with a standard deviation of 4, and 18 students in the section without labs made an average grade of 77 with a standard deviation of 6, find a 99% confidence interval for the difference between the average grades for the two courses. Assume the populations to be approximately normally distributed with equal variances.

27. A study was conducted by the Department of Zoology at the Virginia Tech to estimate the difference in the amounts of the chemical orthophosphorus measured at two different stations on the James River. Orthophosphorus was measured in milligrams per liter. Fifteen samples were collected from station 1, and 12 samples were obtained from station 2. The 15 samples from station 1 had an average orthophosphorus content of 3.84 milligrams per liter and a standard deviation of 3.07 milligrams per liter, while the 12 samples from station 2 had an average content of 1.49 milligrams per liter and a standard deviation of 0.80 milligram per liter. Find a 95% confidence interval for the difference in the true average orthophosphorus contents at these two stations, assuming that the observations came from normal populations.

28. An experiment reported in Popular Science compared fuel economies for two types of similarly equipped diesel mini-trucks. Let us suppose that 12 Volkswagen and 10 Toyota trucks were tested in 90- kilometer-per-hour steady-paced trials. If the 12 Volkswagen trucks averaged 16 kilometers per liter with a standard deviation of 1.0 kilometer per liter and the 10 Toyota trucks averaged 11 kilometers per liter with a standard deviation of 0.8 kilometer per liter, construct a 90% confidence interval for the difference between the average kilometers per liter for these two mini-trucks. Assume that the distances per liter for the truck models are approximately normally distributed.

29. The local branch of the Internal Revenue Service spent an average of 21 minutes helping each of 10 people prepare their tax returns. The standard deviation was 5.6 minutes. A volunteer tax preparer spent an average of 27 minutes helping 14 people prepare their taxes. The standard deviation was 4.3 minutes. Find the 98% confidence interval for the two means.

30. A random sample of enrollments from medical schools that specialize in research and from those that are noted for primary care is listed. Find the 90% confidence interval for the difference in the means.

Research				Primary care			
474	577	605	663	783	605	427	728
783	467	670	414	546	474	371	107
813	443	565	696	442	587	293	277
692	694	277	419	662	555	527	320
884							

31. A certain geneticist is interested in the proportion of males and females in the population who have a minor blood disorder. In a random sample of 1000 males, 250 are found to be afflicted, whereas 275 of 1000 females tested appear to have the disorder. Compute a 95% confidence interval for the difference between the proportions of males and females who have the blood disorder.

32. Ten engineering schools in Cambodia were surveyed. The sample contained 250 electrical engineers, 80 being women; 175 chemical engineers, 40 being women. Compute a 90% confidence interval for the difference between the proportions of women in these two fields of engineering. Is there a significant difference between the two proportions?

33. A clinical trial was conducted to determine if a certain type of inoculation has an effect on the incidence of a certain disease. A sample of 1000 rats was kept in a controlled environment for a period of 1 year, and 500 of the rats were given the inoculation. In the group not inoculated, there were 120 incidences of the disease, while 98 of the rats in the inoculated group contracted it. If p_1 is the probability of incidence of the disease in uninoculated rats and p_2 the probability of incidence in inoculated rats, compute a 90% confidence interval for $p_1 - p_2$.

34. A survey of 1000 students found that 274 chose professional baseball team A as their favorite team. In a similar survey involving 760 students, 240 of them chose team A as their favorite. Compute a 95% confidence interval for the difference between the proportions of students favoring team A in the two surveys. Is there a significant difference?