IE 33000: Probability and Statistics in Engineering II (Fall 2022)
School of Industrial Engineering, Purdue University

Homework 1

Instruction: There are 9 problems in total.
1) Problems 1 to 7 – 10 points
2) Problems 8 to 9 – 15 points

Due September 23, 2022 (11:59 pm)

For all problems, provide both hand-written solutions and R codes.

Problem 1

A manufacturer produces piston rings for an automobile engine. It is known that ring diameter is normally distributed with $\sigma = 0.001$ millimeters. A random sample of 15 rings has a mean diameter of $x = 74.036$ millimeters.

a. Construct a 99% two-sided confidence interval on the mean piston ring diameter.

b. Construct a 99% lower-confidence bound on the mean piston ring diameter. Compare the lower bound of this confidence interval with the one in part (a).

Problem 2

A confidence interval estimate is desired for the gain in a circuit on a semiconductor device. Assume that gain is normally distributed with standard deviation $\sigma = 20$.

a. How large must n be if the length of the 95% CI is to be 40?

b. How large must n be if the length of the 99% CI is to be 40?

Problem 3

In Bioengineering, the adhesion of various biofilms to solid surfaces for possible use in environmental technologies is studied. Adhesion assay is conducted by measuring absorbance. Suppose that for the bacterial strain Acinetobacter, five measurements gave readings of 2.69, 5.76, 2.67, 1.62, and 4.12 dyne-cm^2. Assume that the standard deviation is known to be 0.66 dyne-cm^2.

a. Find a 95% confidence interval for the mean adhesion.

b. If the scientists want the confidence interval to be no wider than 0.55 dyne-cm2, how many observations should they take?

Problem 4

A research engineer for a tire manufacturer is investigating tire life for a new rubber compound and has built 16 tires and tested them to end-of-life in a road test. The sample mean and standard deviation are 60,139.7 and 3645.94 kilometers. Find a 95% confidence interval on mean tire life.

Problem 5

The compressive strength of concrete is being tested by a civil engineer who tests 12 specimens. Assume that compressive strength is normally distributed, and the following data are obtained:

2216	2237	2249	2204
2225	2301	2281	2263
2318	2255	2275	2295

a. Construct a 95% two-sided confidence interval on the mean strength.
b. Construct a 95% lower confidence bound on the mean strength. Compare this bound with the lower bound of the two-sided confidence interval and discuss why they are different.

Problem 6

The percentage of titanium in an alloy used in aerospace castings is measured in 51 randomly selected parts. The sample standard deviation is $s = 0.37$. Construct a 95% two-sided confidence interval for σ.

Problem 7

In a cancer research study, the tumorigenesis of a drug is tested. Rats were randomly selected from litters and given the drug. The times of tumor appearance were recorded as follows:

101, 104, 104, 77, 89, 88, 104, 96, 82, 70, 89, 91, 39, 103, 93, 85, 104, 104, 81, 67, 104, 104, 104, 87, 104, 89, 78, 104, 86, 76, 103, 102, 80, 45, 94, 104, 104, 76, 80, 72, 73

Assuming that the population is normally distributed, calculate a 95% confidence interval on the standard deviation of time until a tumor appearance.

Problem 8

A normal population has known mean $\mu = 50$ and variance $\sigma^2 = 5$. What is the approximate probability that the sample variance is greater than or equal to 7.44? less than or equal to 2.56? For a random sample of size

a. $n = 16$ b. $n = 30$ c. $n = 71$

d. Compare your answers to parts (a)–(c) for the approximate probability that the sample variance is greater than or equal to 7.44. Explain why this tail probability is increasing or decreasing with increased sample size.

e. Compare your answers to parts (a)–(c) for the approximate probability that the sample variance is less than or equal to 2.56. Explain why this tail probability is increasing or decreasing with increased sample size.

Problem 9

During the 1999 and 2000 baseball seasons, there was much speculation that the unusually large number of home runs hit was due at least in part to a livelier ball. One way to test the "liveliness" of a baseball is to launch the ball at a vertical surface with a known velocity VL and measure the ratio of the outgoing velocity VO of the ball to VL. The ratio R = VO/VL is called the coefficient of restitution. Following are measurements of the coefficient of restitution for 40 randomly selected baseballs. Assume that the population is normally distributed. The balls were thrown from a pitching machine at an oak surface.

0.6248	0.6237	0.6118	0.6159	0.6298	0.6192
0.6520	0.6368	0.6220	0.6151	0.6121	0.6548
0.6226	0.6280	0.6096	0.6300	0.6107	0.6392
0.6230	0.6131	0.6223	0.6297	0.6435	0.5978
0.6351	0.6275	0.6261	0.6262	0.6262	0.6314
0.6128	0.6403	0.6521	0.6049	0.6170	
0.6134	0.6310	0.6065	0.6214	0.6141	

a. Find a 99% CI on the mean coefficient of restitution.

b. Find a 99% prediction interval on the coefficient of restitution for the next baseball that will be tested.

c. Find an interval that will contain 99% of the values of the coefficient of restitution with 95% confidence.
d. Explain the difference in the three intervals computed in parts (a), (b), and (c).

Problem 1

A manufacturer produces piston rings for an automobile engine. It is known that ring diameter is normally distributed with σ = 0.001 millimeters. A random sample of 15 rings has a mean diameter of x = 74.036 millimeters.

a. Construct a 99% two-sided confidence interval on the mean piston ring diameter.

b. Construct a 99% lower-confidence bound on the mean piston ring diameter. Compare the lower bound of this confidence interval with the one in part (a).

A) $\bar{X} \pm Z_{(0.005)} \cdot \frac{\sigma}{\sqrt{n}}$

$74.036 \pm 2.58 \cdot \frac{0.001}{\sqrt{15}}$

74.036 ± 0.00067

B) $LC = \bar{x} - Z_{(0.01)} \cdot \frac{\sigma}{\sqrt{n}}$

$= 74.036 - 2.33 \cdot \frac{0.001}{\sqrt{15}}$

$LC = 74.0354$

Interval $= (74.0354, \infty)$

Problem 2

A confidence interval estimate is desired for the gain in a circuit on a semiconductor device. Assume that gain is normally distributed with standard deviation σ = 20.

a. How large must n be if the length of the 95% CI is to be 40?

b. How large must n be if the length of the 99% CI is to be 40?

a) $\alpha = 0.05$ both extremes

$\downarrow \alpha/2$

$Z(0.025) = 1.96$

$CL = \bar{x} \pm Z \cdot \frac{\sigma}{\sqrt{n}}$

$(\cancel{\bar{x}} + Z \cdot \frac{\sigma}{\sqrt{n}}) + (\cancel{\bar{x}} + Z \cdot \frac{\sigma}{\sqrt{n}}) = 40$

$2 \cdot Z \cdot \frac{\sigma}{\sqrt{n}} = 40$

$Z \cdot \frac{\sigma}{\sqrt{n}} = 20$

$1.96 \cdot \frac{20}{\sqrt{n}} = 20$

$n = 3.84$

b) $\alpha = 0.01$

$Z(0.005) = 2.58$

$(\cancel{\bar{x}} + Z \cdot \frac{\sigma}{\sqrt{n}}) + (\cancel{\bar{x}} + Z \cdot \frac{\sigma}{\sqrt{n}}) = 40$

$2.58 \cdot \frac{20}{\sqrt{n}} = 20$

$n = 6.66$

Problem 3

In Bioengineering, the adhesion of various biofilms to solid surfaces for possible use in environmental technologies is studied. Adhesion assay is conducted by measuring absorbance. Suppose that for the bacterial strain Acinetobacter, five measurements gave readings of 2.69, 5.76, 2.67, 1.62, and 4.12 dyne-cm^2. Assume that the standard deviation is known to be 0.66 dyne-cm^2.

a. Find a 95% confidence interval for the mean adhesion.

b. If the scientists want the confidence interval to be no wider than 0.55 dyne-cm2, how many observations should they take?

a) $\bar{x} = \frac{\Sigma x}{n}$

$= \frac{2.69 + 5.76 + 2.67 + 1.62 + 4.12}{5}$

$= 3.37$

$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

$= \sqrt{\frac{(0.46 + 5.7 + 0.49 + 3.07 + 0.56)}{4}}$

$s = 1.60$

$\sigma = 0.66 \quad \alpha = 0.05 \xrightarrow{/2} Z(0.025) = 1.96$

$\bar{x} \pm Z \cdot \frac{\sigma}{\sqrt{n}}$

$3.37 \pm 1.96 \cdot \frac{0.66}{\sqrt{5}}$

$3.37 \pm 0.5785 \rightarrow (2.79, 3.95)$

b) $E = \frac{0.55}{2} \Rightarrow 0.275$

$n = \frac{Z^2 \cdot \sigma^2}{E^2}$

$= \frac{1.96^2 \cdot 0.66^2}{0.275^2}$

$n = 22.13$

Problem 4

A research engineer for a tire manufacturer is investigating tire life for a new rubber compound and has built 16 tires and tested them to end-of-life in a road test. The sample mean and standard deviation are 60,139.7 and 3645.94 kilometers. Find a 95% confidence interval on mean tire life.

$\alpha = 0.05 \rightarrow t(0.05) = 2.13$

$M.E = t \cdot \sqrt{\frac{s^2}{n}}$

$= 2.13 \cdot \sqrt{\frac{3645.94^2}{16}}$

$= 1942.37$

$1 - \alpha = P(\bar{x} - M.E < \mu < \bar{x} + M.E)$

$1 - 0.05 = P(60{,}139.7 - 1942.37 < \mu < 60{,}139.7 + 1942.37)$

$0.95 = P(58{,}197.33 < \mu < 62{,}082.07)$

Problem 5

The compressive strength of concrete is being tested by a civil engineer who tests 12 specimens. Assume that compressive strength is normally distributed, and the following data are obtained:

2216	2237	2249	2204
2225	2301	2281	2263
2318	2255	2275	2295

a. Construct a 95% two-sided confidence interval on the mean strength.
b. Construct a 95% lower confidence bound on the mean strength. Compare this bound with the lower bound of the two-sided confidence interval and discuss why they are different.

a) $\bar{x} = 2259.92$ $\sigma = 35.57$ $\alpha(0.05)$ $n = 12$ $t(0.025) = 2.2$

$\bar{x} \pm t \cdot \frac{\sigma}{\sqrt{n}}$

$2259.92 \pm 2.2 \cdot \frac{35.57}{\sqrt{12}}$

$(2237.33, 2282.51)$

b) $\alpha = 0.05$ $t(0.05) = 1.8$

$\bar{x} - t \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow 2259.17 - 1.8 \cdot \frac{35.57}{\sqrt{12}} \Rightarrow 2241.43$

Problem 6

The percentage of titanium in an alloy used in aerospace castings is measured in 51 randomly selected parts. The sample standard deviation is s = 0.37. Construct a 95% two-sided confidence interval for σ.

$\alpha = 0.05 \rightarrow$ table val 0.025 $n = 51 \Rightarrow$ 50 degrees of freedom

$\sigma = 0.37$

$LC = \sqrt{\dfrac{(51-1)\cdot 0.37^2}{\chi^2_{50} \cdot 0.025}}$

$= 32.36$

$UC = \sqrt{\dfrac{(51-1)\cdot 0.37^2}{\chi^2_{50} \cdot 0.975}}$

$= 71.42$

Problem 7

In a cancer research study, the tumorigenesis of a drug is tested. Rats were randomly selected from litters and given the drug. The times of tumor appearance were recorded as follows:

101, 104, 104, 77, 89, 88, 104, 96, 82, 70, 89, 91, 39, 103, 93, 85, 104, 104, 81, 67, 104, 104, 104, 87, 104, 89, 78, 104, 86, 76, 103, 102, 80, 45, 94, 104, 104, 76, 80, 72, 73

Assuming that the population is normally distributed, calculate a 95% confidence interval on the standard deviation of time until a tumor appearance.

$\bar{x} = \dfrac{\sum x_i}{n}$

$= \dfrac{3640}{41}$

$\bar{x} = 88.78$

$\sigma = \sqrt{\dfrac{\sum (x-\bar{x})^2}{n-1}}$

$= \sqrt{255.78}$

$= 15.99$

$Z(0.10) = 1.645$

$= 88.78 \pm 1.645 \cdot \dfrac{15.99}{\sqrt{41}}$

$= 88.78 \pm 4.11$

$= 88.78 \pm 4.11$

$(84.67, 92.89)$

Problem 8

A normal population has known mean $\mu = 50$ and variance $\sigma^2 = 5$. What is the approximate probability that the sample variance is greater than or equal to 7.44? less than or equal to 2.56? For a random sample of size

a. $n = 16$ b. $n = 30$ c. $n = 71$

d. Compare your answers to parts (a)–(c) for the approximate probability that the sample variance is greater than or equal to 7.44. Explain why this tail probability is increasing or decreasing with increased sample size.

e. Compare your answers to parts (a)–(c) for the approximate probability that the sample variance is less than or equal to 2.56. Explain why this tail probability is increasing or decreasing with increased sample size.

a) $\mu = 50 \quad \sigma^2 = 5$ 15 degrees of freedom

$$P(S^2 \geq 7.44) = P\left(\chi^2 \geq \frac{15 \cdot 7.44}{5}\right)$$
$$= P(\chi^2 \geq 22.32) \Longrightarrow 0.0997$$

random variable χ^2 has chi distribution

$$P(S^2 \leq 2.56) = P\left(\chi^2 \leq \frac{15 \cdot 2.56}{5}\right)$$
$$= P(\chi^2 \leq 7.68) \Longrightarrow 0.064$$

b) $n = 30$

$$P(S^2 \geq 7.44) = P\left(\chi^2 \geq \frac{29 \cdot (7.44)}{5}\right)$$
$$= P(\chi^2 \geq 43.15) \Longrightarrow 0.044$$

$$P(S^2 \leq 2.56) = P\left(\chi^2 \leq \frac{29 \cdot 2.56}{5}\right)$$
$$= P(\chi^2 \leq 14.85) \Longrightarrow 0.014$$

c) $n = 71$

$$P(s^2 \geq 7.44) = P\left(\chi^2 \geq \frac{70 \cdot 7.44}{5}\right)$$

$$= P(\chi^2 \geq 104.16) \implies 0.0051$$

$$P(s^2 \leq 2.56) = P\left(\chi^2 \leq \frac{70 \cdot 2.56}{5}\right)$$

$$= P(\chi^2 \leq 35.84) \implies 0.0009$$

d) The probabilty of the sample variation being ≥ 7.44 decreases with the increased sample size. When sample size increases so does the degrees of freedom. Thus the tail probabilites decrease.

e) The probability of the sample being ≤ 2.56 is smaller as sample size increases. This is due to the same reason as part d. Tail probabilities decrease as degrees of freedom increase.

Problem 9

During the 1999 and 2000 baseball seasons, there was much speculation that the unusually large number of home runs hit was due at least in part to a livelier ball. One way to test the "liveliness" of a baseball is to launch the ball at a vertical surface with a known velocity VL and measure the ratio of the outgoing velocity VO of the ball to VL. The ratio R = VO/VL is called the coefficient of restitution. Following are measurements of the coefficient of restitution for 40 randomly selected baseballs. Assume that the population is normally distributed. The balls were thrown from a pitching machine at an oak surface.

0.6248	0.6237	0.6118	0.6159	0.6298	0.6192
0.6520	0.6368	0.6220	0.6151	0.6121	0.6548
0.6226	0.6280	0.6096	0.6300	0.6107	0.6392
0.6230	0.6131	0.6223	0.6297	0.6435	0.5978
0.6351	0.6275	0.6261	0.6262	0.6262	0.6314
0.6128	0.6403	0.6521	0.6049	0.6170	
0.6134	0.6310	0.6065	0.6214	0.6141	

a. Find a 99% CI on the mean coefficient of restitution.
b. Find a 99% prediction interval on the coefficient of restitution for the next baseball that will be tested.

c. Find an interval that will contain 99% of the values of the coefficient of restitution with 95% confidence.
d. Explain the difference in the three intervals computed in parts (a), (b), and (c).

$$CI = \left(\bar{x} - \frac{t_c \cdot \sigma}{\sqrt{n}}, \bar{x} + \frac{t_c \cdot \sigma}{\sqrt{n}}\right), \quad R = V_0 / V_L$$

a) 99% CI:

$\bar{x} = 0.624 \rightarrow \frac{\Sigma x}{N}$

$\sigma = 0.013$, $\sigma_e = 0.002$

$n = 40$

$t_{crit} = 2.708$

$(0.619, 0.630)$

b) 99% CI on next:

$\bar{x} = 0.624$

$\sigma = 0.013$ $\sigma_e = 0.013$

$n = 40$

$t_{crit} = 2.708$

$(0.589, .660)$

c) 99% of values and 95% confidence:

$\bar{x} = 0.624$

$\sigma = 0.013$ $\sigma_e = 0.002$

$n = 40$ $t_{crit} = 3.797$

$(0.616, 0.632)$

D) The predictions vary due to variability in data.
The prediction interval is widest/largest because it has to incorporate most.

```
#Q1) s = 0.001, n = 15, xbar = 74.036, a = 0.01
##a)
1a1 <- xbar + Z[a/2] * (s / sqrt(n))
1a2 <- xbar - Z[a/2] * (s / sqrt(n))
##b)
1b_LC <- xbar - Z[a] * (x / sqrt(n))

#Q2) s = 20, a = 0.05, CL = 40,
##a)
##CL = Xbar +- Z[] * s / sqrt(N)
2a_N <- (4 *s^2 * Z[a]^2) / (CL)^2
##b)
2b_N <- (4 *s^2 * Z[a/2]^2) / (CL)^2

#Q3) X = c(2.69, 5.76, 2.67, 1.62, 4.12), s = 0.66, n = 5
##a)
3a_xbar <- sum(X) / n
3a_xi <- x - xbar
3a_s <- sqrt(xi^2 / (n-1))
3a_LC <- xbar - Z[a/2] * sqrt(s/sqrt(n))
3a_UC <- xbar + Z[a/2] * sqrt(s/sqrt(n))
##b)
3b_E <- 0.55 / 2
3b_N <- (Z[a/2]^2 * s^2) / E^2

#Q4)n = 16, xbar = 60,139.7, s = 3645.94, 95% CI, a = 0.05
##a)
4_ME <- t[a]*sqrt(s^2 / n)
4_CL <- P((s - ME) < u < (s + ME))

#Q5) x = c(2216, 2237, 2249, 2204, 2225, 2301, 2281, 2263, 2318, 2255, 2275,
2295)
##a) a = 0.05 n =12
5a_xbar <- sum(x) / n
5a_s <-sd(X)
5a_LC <- xbar - Z[a/2] * sqrt(s/sqrt(n))
5a_UC <- xbar + Z[a/2] * sqrt(s/sqrt(n))
##b)
5b <- xbar - t[a] * s / sqrt(n)

#Q6) s = 0.37, n = 51, a = 0.05
##a)
6_LC <- sqrt(((n-1) * s^2) / (a * t[]))
6_UC <- sqrt(((n-1) * s^2) / ((1-a) * t[]))

#Q7) a = 0.05, n =41
#x = c(101, 104, 104, 77, 89, 88, 104, 96, 82, 70, 89, 91, 39, 103, 93, 85,
104,
#######104, 81, 67,104, 104, 104, 87, 104, 89, 78, 104, 86, 76, 103, 102, 80,
#######45, 94, 104, 104, 76, 80,72, 73)
7_xbar <- sum(x) / n
7_s <- sd(x)
7_LC <- xbar - Z[a/2] * sqrt(s/sqrt(n))
7_UC <- xbar + Z[a/2] * sqrt(s/sqrt(n))

#Q8) xbar = 50, v = 5, n = c(16, 30, 71)
```

```
8a_LP <- P(X^2 >= ((n-1) * 7.44)/v)
8a_UP <- P(X^2 <= ((n-1) * 2.56)/v)

#Q9)  n = 40
# x<- c(0.6248 0.6237 0.6118 0.6159 0.6298 0.6192
####...0.6520 0.6368 0.6220 0.6151 0.6121 0.6548
####...0.6226 0.6280 0.6096 0.6300 0.6107 0.6392
####...0.6230 0.6131 0.6223 0.6297 0.6435 0.5978
####...0.6351 0.6275 0.6261 0.6262 0.6262 0.6314
####...0.6128 0.6403 0.6521 0.6049 0.6170
####...0.6134 0.6310 0.6065 0.6214 0.6141
9_xbar <- sum(x) / n
9_s <- sd(x)
9_LC <- xbar - (t[a] * s) / sqrt(n)
9_UC <- xbar + (t[a] * s) / sqrt(n)
```