We are entering the later part of the course where a larger number of concepts and terms are introduced all of which are important and will be examined in quizzes, projects and the final exam. We will not spend so much time on carrying out detailed calculations of the kind already studied. If you need further detail ask about it when the need arises (for you) in lecture.

1-21 A day’s production of $N = 3,600$ spools of fishing line is being inspected for the score $x =$ weight of finished product (spool with line and label). The day’s production has mean $\mu = 74$ and standard deviation $\sigma = 8.4$. These true population parameter values are not known to the quality inspector who has selected an i.i.d. sample of $n = 46$ from the day’s production. The inspector finds that $\overline{x}$ (the mean of this sample) is equal to $75.2$ and $s$ (the sample s.d. of this sample) is equal to $7.9$.

The parameter to be estimated from the sample is $\mu$. The other parameter $\sigma$ is unavoidably involved but is not of itself of interest so it is termed a "nuisance" parameter.

1. What is the true value of the population parameter $\mu$ ?

2. What is the naive choice for an estimator of $\mu$ ?

3. Is the estimator of (2) unbiased ?

4. Is the estimator of (2) consistent as $n \to \infty$ ?

5. What is the estimate of $\mu$ for this data ?

6. What is the true value of the nuisance parameter $\sigma$ ?

7. What is the commonly used estimator of $\sigma$ ?
8. What is the estimate of $\sigma$ from this data?

9. Is the estimator of (7) unbiased?

10. Is the estimator of (7) asymptotically unbiased as $n \to \infty$?

11. Is the estimator of (7) consistent as $n \to \infty$?

12. What is the true s.d. of xBAR?

13. What is the commonly used estimator of the s.d. of r.v. xBAR?

14. For the given data what is the estimate of the s.d. of r.v. xBAR?

15. Give the formula for the "95% confidence interval for $\mu$" based on a sample of n x-values.

16. Evaluate the formula of (14) for the given data.

17. Give the order of dependence upon n for the difference:
    $0.95 - P(\text{true value } \mu \text{ is covered by the random 0.95 confidence interval}).$

18. Does your interval from (15) actually cover the true value of $\mu$?

19. If thousands of people each independently carry out a sample of $n = 46$ from this same population, and each person then calculates their own 0.95 confidence interval from their own sample, what percentage of these thousands of people will have their confidence interval cover the true value of $\mu$?

20. Instead of the 0.95 c.i. in (15) give the 0.97 c.i..

21. What z score did you use in (19)?
22-44 A population has x-scores equal to 1 if a person is a carrier for a particular disease and 0 if they are not. The fraction of the population who are carriers is 0.32. Without knowing this, a health team samples \( n = 71 \) persons i.i.d. from this population (i.e. with replacement and with equal probability). Of this sample of 71 they find that 26 are carriers.

22. What is the true value of the population parameter \( \mu \) ?

23. What is the naive choice for an estimator of \( \mu \) ?

24. Is the estimator of (23) unbiased ?

25. Is the estimator of (23) consistent as \( n \to \infty \) ?

26. What is the estimate of \( \mu \) for this data ?

27. What is the true value of the nuisance parameter \( \sigma \) ?

28. What is the commonly used estimator of \( \sigma \) ?

29. What is the estimate of \( \sigma \) from this data ?

30. Is the estimator of (28) unbiased ?

31. Is the estimator of (28) asymptotically unbiased as \( n \to \infty \) ?

32. Is the estimator of (28) consistent as \( n \to \infty \) ?

33. What is the true s.d. of \( x_{\text{BAR}} \) ?

34. What is the commonly used estimator of the s.d. of r.v. \( x_{\text{BAR}} \) (express in terms of \( p \)) ?

35. For the given data what is the estimate of the s.d. of r.v. \( x_{\text{BAR}} \) ?

36. Give the formula for the "95% confidence interval for \( \mu \)" based on a
sample of \( n \) x-values (express in terms of \( p \)).

37. Evaluate the formula of (36) for the given data.

38. Give the order of dependence upon \( n \) for:
\[
0.95 - \Pr(\text{true value } \mu \text{ is covered by the random 0.95 confidence interval}).
\]

39. Does your interval from (37) actually cover the true value of \( \mu \)?

40. Instead of the 0.95 c.i. in (37) give the 0.99 c.i..

41. What \( z \) score did you use in (40)?

42. If thousands of people each independently carry out a sample of \( n = 71 \) from this same population, and each person then calculates their own 0.99 confidence interval from their own sample, what percentage of these thousands of people will have their confidence interval cover the true value of \( \mu \)?

43. What is the connection between \( \mu \) and \( p \) in this example?

44. What is the connection between \( \bar{x} \) and \( \hat{p} \) in this example?

45-48 A population is assumed to be NORMALLY DISTRIBUTED (very special!). An i.i.d. sample of \( n = 5 \) is selected from this NORMAL population. The sample mean is \( \bar{x} = 23.6 \) and \( s = 3.8 \).

45. Give the 0.95 c.i. for \( \mu \) that one would ordinarily use for a large sample size (not recommended for the measly \( n = 5 \) we have!).

46. Give the 0.95 c.i. for \( \mu \) based upon our sample of \( n = 5 \) (t-table).

47. Since our population is NORMAL what is:
\[
0.95 - \Pr(\mu \text{ is covered by 0.95 c.i.})\]

48. What does the \( t \) method not apply to \( x \)-scores of the 0, 1 type only?

49-51 An initial sample of \( n_0 = 40 \) service stations is selected (i.i.d. sample). We desire a 0.95 c.i. for \( \mu = \) the mean selling price of regular
unleaded gasoline, self serve, that is accurate to within one penny. That is, we desire a 0.95 c.i. of the form (in dollars) xBAR + \{-1, 1\} 0.01. The formula for Stein’s 0.95 large sample c.i. for \(\mu\) differs from the usual

\[
\text{xBAR} + \{-1, 1\} \frac{1.96 \text{ S}}{\sqrt{n}}
\]

in that the eventual total sample size \(n\) is random, determined from an initial sample of \(n_0\). Stein’s c.i. is

\[
\text{xBAR} + \{-1, 1\} \frac{1.96 \text{ S}_0}{\sqrt{n}}
\]

where \(s_0\) is the s.d. of the INITIAL sample of \(n_0\) and one continues sampling up from the initial sample of \(n_0\) to a final total sample size (including the initial sample) of \(n = (1.96 \text{ S}_0 / \Delta)^2\) (if this is less than or equal to \(n_0\) there is no need to increase the initial sample size at all). By choosing the total sample size in this way Stein’s method forces (at least approximately) that the half-width of Stein’s 0.95 c.i. is

\[
1.96 \text{ S}_0 / \sqrt{n} \approx \Delta.
\]

so Stein’s c.i. is (approximately)

\[
\text{xBAR} + \{-1, 1\} \Delta.
\]

49. Using Stein’s method, to what TOTAL sample size \(n\) should we continue to sample in order that we achieve a 0.95 c.i. of approximately \(\text{xBAR} + \{-1, 1\} 0.01\) ?

50. What is the downside to Stein’s method?

51. If thousands of persons each independently select initial samples of various \(n_0\) (large) and each applies Stein’s 0.95 c.i. as above, around what percentage of these persons will fail to cover the true \(\mu\) with their Stein c.i.?

**Reminder:** Usual c.i. (n large) rests on \((\text{xBAR} - \mu) / (s / \sqrt{n}) \approx Z\). Student’s t -method (NORMAL population, arbitrary \(n > 1\)) on \((\text{xBAR} - \mu) / (s / \sqrt{n}) \approx t\) with \(n-1\) degrees of freedom. Stein’s method on \((\text{xBAR} - \mu) / (S_0 / \sqrt{n}) \approx Z\) provided \(n\) (random) is chosen as per Stein’s method and \((\text{xBAR} - \mu) / (S_0 / \sqrt{n_0}) \approx Z\) (i.e. that the CLT approximation would have been good for the initial sample of \(n_0\)).