Chapter 7: Interval Estimation

7.1 Confidence Intervals for Means

Given a random sample $X_1, X_2, ..., X_n$ from a normal distribution $N(\mu, \sigma^2)$, we shall now consider the closeness of $\bar{X}$, the unbiased estimator of μ, to the unknown mean μ. To do this, we use the distribution of $\bar{X}$, where $\bar{X} \sim N(\mu, \sigma^2/n)$.

Confidence interval for the unknown parameter μ when the variance σ^2 is *known*:

For the probability α, we can find a number $z_{\alpha/2}$ such that

Example 1. 1. If $1 - \alpha$ =0.95 then $z_{\alpha/2}$ =

2. If $1 - \alpha$ =0.90 then $z_{\alpha/2}$ =

Recall that $\sigma > 0$ we see that the following inequalities are equivalent:

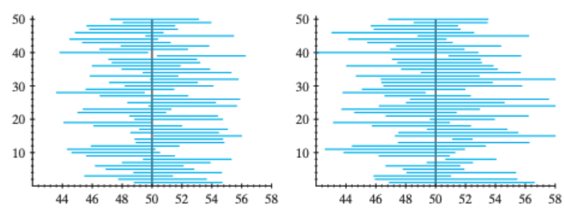
- Once the sample is observed and the sample mean computed to equal______, the interval __________________ becomes known.
- Since the probability that the random interval covers ______ before the sample is drawn is equal to ______, we now call the computed interval, __________________, **a 100(1 - α)% confidence interval for the unknown mean μ.**
- For example,__________________ is a 95% confidence interval for μ.
- The number ______, or equivalently, ______, is called the________________________.

Note:

1. The confidence interval for ______ is centered at the point estimate ______ and is completed by subtracting and adding the quantity __________.
2. As ______ increases, ____________ decreases, resulting in a shorter confidence interval with the same confidence coefficient $1-\alpha$.
3. A shorter confidence interval gives a more precise estimate of ______ , regardless of the confidence we have in the estimate of μ.
4. For a fixed sample size n, the length of the confidence interval can also be shortened by decreasing the confidence coefficient $1-\alpha$. But if this is done, we achieve a shorter confidence interval at the expense of losing some confidence.

Example 2. Let X equal the length of life of a 60-watt light bulb marketed by a certain manufacturer. Assume that the distribution of X is $N(\mu, 1296)$. If a random sample of $n=27$ bulbs is tested until they burn out, yielding a sample mean of $\bar{x}=1478$ hours, Find a 95% confidence interval for μ.

Example 3. Let $\bar{x}$ be the observed sample mean of five observations of a random sample from the normal distribution $N(\mu, 16)$. Find a 90% confidence interval for the unknown mean μ.



(a) 90% confidence intervals, σ known (b) 90% confidence intervals, σ unknown

For a particular sample, this interval either does or does not contain the mean μ. However, if many such intervals were calculated, about 90% of them should contain the mean μ.

Fifty random samples of size 5 from the normal distribution N(50, 16) were simulated on a computer. A 90% confidence interval was calculated for each random sample, as if the mean were unknown. Figure (a) depicts each of these 50 intervals as a line segment. Note that 45 (or 90%) of them contain the mean, $\mu=50$.

1. If **we cannot assume that the distribution from which the sample arose is normal**, we can still obtain an approximate confidence interval for μ.

2. By the *central limit theorem*, provided that n is large enough, the ratio ________________ has the approximate normal distribution $N(0,1)$ when the underlying distribution is not normal. In this case,

3. The closeness of the approximate probability $1 - \alpha$ to the exact probability depends on both the ________________ and the ________________.

4. When the underlying distribution is ________________ (has only one mode), symmetric, and continuous, the approximation is usually quite good even for small n, such as ________________.

5. As the underlying distribution becomes "less normal" (i.e., badly skewed or discrete), a larger sample size might be required to keep a reasonably accurate approximation. But, in almost all cases, an n of ________________ is usually adequate.

Example 4. Let X equal the amount of orange juice (in grams per day) consumed by an American. Suppose it is known that the standard deviation of X is $\sigma = 96$. To estimate the mean μ of X, an orange growers' association took a random sample of $n = 576$ Americans and found that they consumed, on the average, $\bar{x} = 133$ grams of orange juice per day.

Find an approximate 90% confidence interval for μ.

1. If σ^2 **is unknown** and the sample size n is 30 or greater, we shall use the fact that the ratio ________________ has an approximate normal distribution $N(0, 1)$.

2. This statement is true whether or not the underlying distribution is normal.

3. If the **underlying distribution is badly skewed or contaminated with occasional outliers**, most statisticians would prefer to have a larger sample size—say, 50 or more—and even that might not produce good results.

Example 5. Lake Macatawa, an inlet lake on the east side of Lake Michigan, is divided into an east basin and a west basin. To measure the effect on the lake of salting city streets in the winter, students took 32 samples of water from the west basin and measured the amount of sodium in parts per million in order to make a statistical inference about the unknown mean μ. They obtained the following data:

13.0	18.5	16.4	14.8	19.4	17.3	23.2	24.9
20.8	19.3	18.8	23.1	15.2	19.9	19.1	18.1
25.1	16.8	20.4	17.4	25.2	23.1	15.3	19.4
16.0	21.7	15.2	21.3	21.5	16.8	15.6	17.6

For these data, $\bar{x} = 19.07$ and $s^2 = 10.60$.

Find an approximate 95% confidence interval for μ.

Summary: we have found a confidence interval for the mean μ of a normal distribution,assuming that the value of the standard deviation σ is known or assuming that σ is unknown but the sample size is large.

How to proceed under the circumstances that the sample sizes are small and we do not know the value of the standard deviation?

If the random sample arises from a normal distribution, we use the fact that

Example 6. Let X equal the amount of butterfat in pounds produced by a typical cow during a 305-day milk production period between her first and second calves. Assume that the distribution of X is $N(\mu, \sigma^2)$. To estimate μ, a farmer measured the butterfat production for n = 20 cows and obtained the following data:

481 537 513 583 453 510 570 500 457 555
618 327 350 643 499 421 505 637 599 392

For these data, $\bar{x} = 507.50$ and s = 89.75. Find a 90% confidence interval for μ.

Note: The Interval ________________ is shorter than ________________.

So far, we have created only what are called **two-sided confidence intervals** for the mean μ. Sometimes, however, we might want only a lower (or upper) bound on μ. We proceed as follows. Say $\bar{X}$ is the mean of a random sample of size n from the normal distribution $N(\mu, \sigma^2)$, where, for the moment, assume that σ^2 is known. Then

Once $\bar{X}$ is observed to be equal to $\bar{x}$, it follows that ________________, is a $100(1-\alpha)\%$ **one-sided confidence interval for** μ.

That is, with the confidence coefficient $1-\alpha$, ________________ is a lower bound for μ. Similarly, ________________ is a one-sided confidence interval forμ and ________________ provides an upper bound for μ with confidence coefficient $1-\alpha$.

When σ is unknown, we would use ________________ to find the corresponding lower or upper bounds for μ, namely,