MODULE 13

CONFIDENCE REGIONS

Contents

T HE FINAL RESULT FROM A HYPOTHESIS TEST (Module 12) can feel uneventful – i.e., either conclude that the parameter may be equal to or different from the hypothesized value.[1] If the parameter is thought to be different from the hypothesized value we might then say that our best guess at the parameter is the observed statistic. However, as shown in Module 10, a statistic is an imperfect estimate of the unknown parameter because of sampling variability. This imperfectness can be recognized by computing a range of values that is likely to contain the parameter. For example, we may make a statement such as this – "Our best guess for the true population mean length of fish in Square Lake is the sample mean of 98.5 mm; however, we are 95% confident that the mean of ALL fish in the lake is between 95.9 and 101.1 mm." The range in the last phrase acknowledges sampling variability and is called a confidence region. In this module, the concept, calculation, and interpretation of confidence regions is explored.

13.1 Confidence Concept

An understanding of what it means to be "95% confident" requires examination of multiple samples from a population, as was done in Module 10 when considering sampling variability. In this initial discussion, only 95% confidence intervals (CI), where a range (i.e., bounded on both ends) is computed, are considered. These simplifying restrictions and unrealistically knowing population values are used here only so that the **concept** of confidence intervals can be more easily explained. General methods for constructing other types of confidence regions with other levels of confidence are in Section 13.2.

---

[1]Depending on the H_A it may be known if the parameter is more or less than the hypothesized value.

In the Square Lake example (introduced in Module 1), it was known that μ=98.06 and σ=31.49 (Table 2.1). Additionally, $\bar{x} = 100.04$ was obtained from the first sample of n=50 (Table 2.2). A 95% CI for μ is defined as $\bar{x} \pm 2SE_{\bar{x}}$. The 95% CI for the mean total length for the Square Lake population, computed from this one sample, is $100.04 \pm 2\frac{31.49}{\sqrt{50}}$, 100.04 ± 8.91, or (91.13,108.95). This interval contains μ (i.e., 98.06 is between 91.13 and 108.95). In other words, this particular CI accomplished what it was intended to do; i.e., provide a range that contains μ.

Despite the success observed in this first sample, not all confidence intervals will contain μ. For example, four of 100 95% confidence intervals shown in Figure 13.1 did not contain μ. Thus, the researcher would have concluded that μ was in an incorrect interval four times in these 100 samples. The concept of "confidence" in confidence regions is related to determining how often the intervals correctly contain the parameter.

Figure 13.1. Sampling distribution of the sample mean (top) and 100 95% confidence intervals (horizontal lines) from samples of n=50 from the Square Lake population. Confidence intervals that do NOT contain μ=98.06 are shown in red and with an open circle. The green shaded area represents 95% of the sample means. See text for more explanation.

From the Central Limit Theorem, the sampling distribution of $\bar{x}$ for samples of n=50 is $N(98.06, \frac{31.49}{\sqrt{50}})$ or $N(98.06, 4.45)$ for this known population. According to the 68-95-99.7% Rule, it is known that 95% of the sample means in this sampling distribution will be between $\mu \pm 2SE$ or, in this specific case, between $98.06 \pm 2(4.45)$. The sampling distribution and this range of expected sample means is shown at the top of Figure 13.1. Note that any sample that produced a mean (dot on the CI line) inside the expected range of sample means (light green area) also produced a 95% CI that contained μ (i.e., black CI line with a solid circle). Because 95% of the sample means will be within the expected range of sample means, 95% of the 95% CIs will contain μ. So, "95% confident" means that 95% of all 95% CIs will contain the parameter and 5% will not. In other words, the mistake identified above will be made with 5% of all 95% CIs.

The specifics for constructing confidence regions with different levels of confidence is described below. However, at this point, it should be noted that the number of CIs expected to contain the parameter of interest is set by the level of confidence used to construct the CI. For example, 80% of 80% CIs and 90% of 90% CIs will contain the parameter of interest. In either case, a particular CI either does or does not contain the interval and, in real-life, we will never know whether it does or does not (i.e., we won't know the value of the parameter). However, we do know that the technique (i.e., the construction of the CI) will "work" (i.e., contain the parameter) a set percentage of the time. To reiterate this point, examine the 100 90% CIs (Figure 13.2-Left) and 100 80% CIs (Figure 13.2-Right) for the Square Lake fish length data.
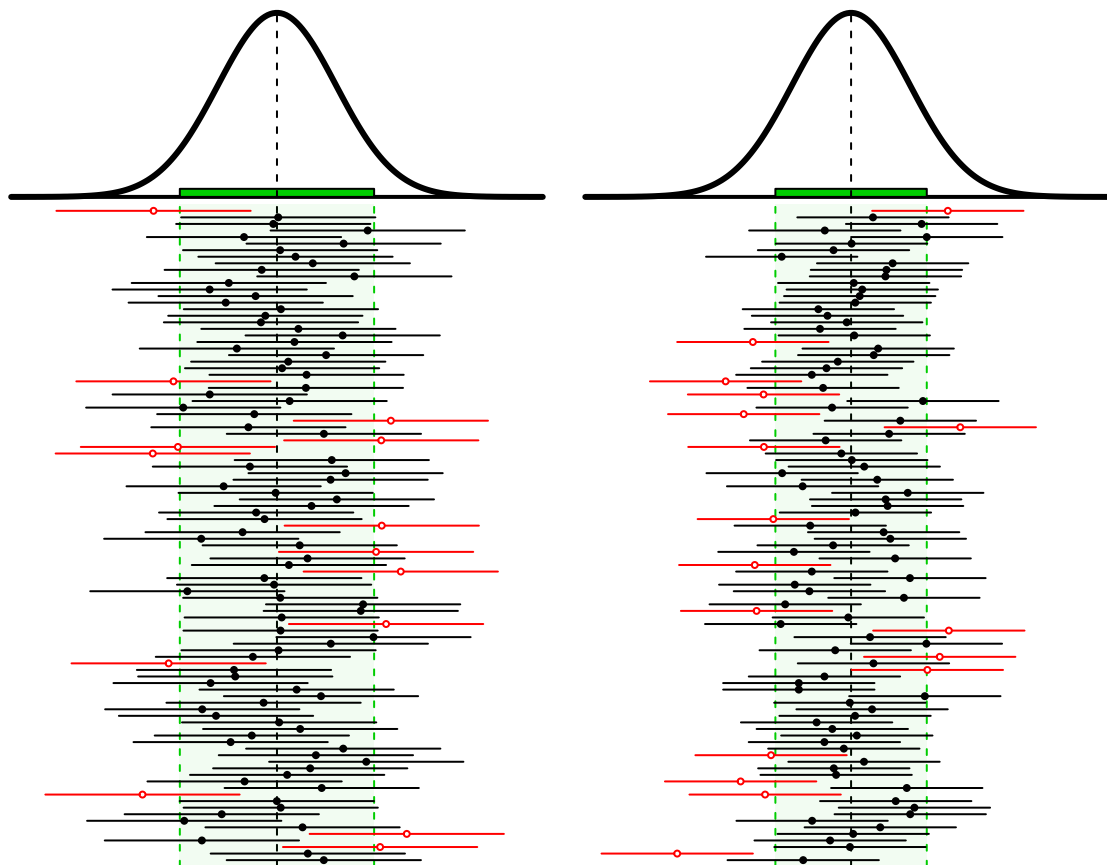

Figure 13.2. Sampling distribution of the sample mean (**tops**) and 100 random 90% (**Left**) and 80% (**Right**) confidence intervals (horizontal lines) from samples of n=50 from the Square Lake population. Confidence intervals that do NOT contain μ are shown in red.

The concept of confidence regions can be difficult to grasp at first. Thus, one should consider the following subtleties about the concept of a confidence region:

- A CI is a random variable like any other statistic. That is, each sample results in a different 95% CI (see CI lines in Figures 13.1 and 13.2) just like it results in a different $\bar{x}$ (see dots on CI lines in Figures 13.1 and 13.2).
- Any CI either contains the parameter (e.g., μ) or not. However, on average, 95% of 95% CIs will contain the parameter and 5% will not. That is, 95% of all possible 95% CIs wil contain the parameter.
- A 95% CI is a technique that "works correctly" 95% of the time. In other words, 95% of all 95% CI "capture" the unknown parameter.

Because of these subtleties, confidence regions are often misinterpreted. Common misinterpretations are listed below with an explanation for the misinterpretation in parentheses. These misinterpretations should be studied, compared to the interpretations discussed above, and avoided.

1. "There is a 95% probability that the population mean is contained in the confidence interval." [*This is incorrect because the population mean is constant (not random), it either is or is not in a particular CI, and it will never change whether it is or is not in that CI. The CI, not the parameter, is random.*]
2. "There is a 95% probability that the sample mean is contained in the confidence interval." [*This is incorrect for the simple fact that CI are not used to estimate sample means (or, generally, statistics); they are used to estimate population means (or parameters). Furthermore, the sample mean has to be exactly in the middle of the CI (see next section).*]
3. "95% of all 95% confidence intervals are contained in the confidence interval." [*First, this is physically impossible because each CI is the same width (if n and the level of confidence stay constant). Second, it is not important how many CI are contained in a CI; interest is in whether the parameter is in the interval or not.*]

◇ **Confidence intervals are constructed for parameters, not statistics.**

◇ **Care and specificity must be used when interpreting and describing confidence intervals.**

13.2 Constructing Confidence Regions

Not all confidence regions are designed to contain the parameter 95% "of the time," are intervals, or are computed to contain μ. Confidence regions can be constructed for any level of confidence, as intervals or bounds, and for nearly all **parameters**.

The level of confidence (C) used will be determined by the α chosen for the hypothesis test; specifically, $C = 100(1 - \alpha)\%$. For example, if α is set at 0.05, then the level of confidence will be $100(1 - 0.05)\%$ (Table 13.1). Thus, if α is decreased such that fewer Type I errors are made, then the confidence level will increase and more of the confidence regions will contain the parameter of interest (i.e., fewer errors). In this manner the proportion of Type I errors in hypothesis testing is linked to the proportion of errors made with confidence regions.

Table 13.1. Several common confidence levels (C) and the corresponding probability of a Type I error (α).

α	C
0.01	99%
0.05	95%
0.10	90%

The type of confidence region depends on the type of alternative hypothesis (Table 13.2). If the alternative hypothesis is two-tailed (i.e., $\neq$), then the confidence region will be an interval (i.e., a range will be computed, as in Section 13.1). However, if the alternative hypothesis is one-tailed, then a confidence bound is used. For example, if the alternative hypothesis is a "less than", then interest lies in determining what is the "largest possible value" for the parameter (rather than a range of possible values). In other words, if the alternative hypothesis is a "less than", then an upper confidence bound for the parameter is constructed. In contrast, if the alternative hypothesis is a "greater than", then a lower confidence bound is constructed to estimate the "smallest possible value" for the parameter.

Table 13.2. Confidence regions and their interpretation in relation to alternative hypotheses (H_A) types.

H_A	Confidence Region	Interpretation
$\neq$	Interval	Parameter in interval
$<$	Upper Bound	Parameter less than upper bound
$>$	Lower Bound	Parameter greater than lower bound

Fortunately, most confidence regions follow the same basic form of

$$\text{“Statistic”} + \text{“scaling factor”} * SE_{statistic}$$

where "Statistic" represents the statistic used to estimate the parameter, $SE_{statistic}$ is the standard error of that statistic, and "scaling factor" $* SE_{statistic}$ is called the margin-of-error. The scaling factor is computed from a known distribution. When σ is known, the scaling factor is computed from a $N(0,1)$ and is called Z^*. Thus, in the case when a confidence interval is being constructed for μ and σ is known, the specific formula for the confidence region is

$$\bar{x} + z^* \frac{\sigma}{\sqrt{n}}$$

The "scaling factor" serves to control the width and type of confidence region. The magnitude of the scaling factor controls the relative width of the region such that the parameter is contained in the region at a rate according to the level of confidence. For example, the scaling factor for a 99% confidence region will be set such that 99% of the confidence regions will contain the parameter.

The sign of the scaling factor controls whether an interval, upper bound, or lower bound is computed. For example, if the alternative hypothesis is two-tailed, then Z^* is the two values such that an area equal to the level of confidence is contained between them (Figure 13.3-Left). The two values that delineate these boundaries will be the same value but with different signs because the $N(0,1)$ is symmetric about zero. Thus, a confidence interval is computed with a scaling factor of $\pm Z^*$.
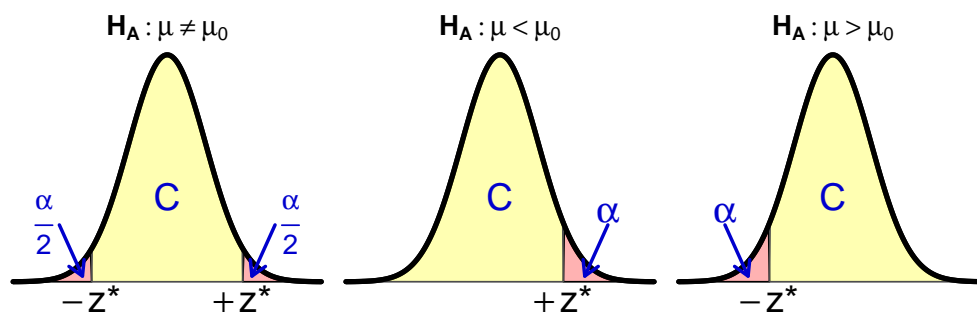
Figure 13.3. Areas (yellow) that define Z^* for confidence regions of a parameter in a hypothesis test.

In contrast, if the alternative hypothesis is a "less than", then an upper confidence bound is desired and Z^* has an area equal to the level of confidence LESS THAN it (Figure 13.3-Middle). As the level of confidence will always be greater than 50%, this definition will produce a positive Z^* so that the scaling factor will be $+Z^*$. Similarly, if the alternative hypothesis is a "greater than", then a lower confidence bound is desired and Z^* has an area equal to the level of confidence GREATER THAN it (Figure 13.3-Right). This definition produces a negative Z^* so that the scaling factor will be $-Z^*$.

> ◇ **When finding Z^* for a confidence bound, the level of confidence always represents an area shaded in the same direction as the sign in H_A.**

Constructing a proper confidence region should follow the five steps below. These steps are illustrated in three examples further below.

1. Identify the level of confidence (i.e., $C = 100(1 - \alpha)\%$; Table 13.1).
2. Identify the type of confidence regions – interval, lower bound, or upperbound (Table 13.2).
3. Determine the scaling factor.
4. Compute the actual confidence region.
5. Interpret the confidence region.

Consider the Square Lake example where $H_A : \mu < 105$, $\alpha = 0.05$, and $\bar{x}{=}100.04$ from $n = 50$ (Table 2.2).

1. $C = 95\%$ $(100(1 - 0.05))$.
2. Upper confidence bound because H_A is a "less than."
3. $Z^* = +1.645$ as found with [Note that `mean=0` and `sd=1` are the default settings for `distrib()` and can, thus, be omitted when finding a Z^*.]

```
> ( distrib(0.95,type="q") )
[1] 1.644854
```

4. $100.04 + 1.645\frac{31.49}{\sqrt{50}}$, $100.04 + 7.33$, or 107.37.
5. One is 95% confident that the mean total length of ALL fish in Square Lake is less than 107.4 mm. By confident, it is meant that 95% of all 95% confidence regions will contain μ.

Second, suppose that the Lake Superior ice cover data from Table 5.2 (note that $\bar{x}{=}107.8$ and $n{=}42$) was tested with $H_A : \mu \neq 100$, $\sigma = 22$, and $\alpha = 0.01$.

1. $C = 99\%$ $(100(1 - 0.01))$.
2. Confidence interval because H_A is a "not equals."
3. $Z^* = \pm 2.576$ as found with

```
> ( distrib(0.995,type="q") )
[1] 2.575829
```

4. $107.8 \pm 2.576\frac{22}{\sqrt{42}}$, 107.8 ± 8.74, or $(99.06, 116.54)$.
5. One is 95% confident that the mean annual number of days of ice cover on Lake Superior is between 99.1 and 116.5 days. By confident, it is meant that 95% of all 95% confidence regions will contain μ.

Finally, suppose that the second example hypothesis test in Module 12 about battery life (i.e., $H_A : \mu > 36$ vs $H_0 : \mu = 36$) is being tested with $\alpha = 0.10$. Further suppose that $\sigma = 7$ and that $\bar{x} = 45$ from $n = 40$.

1. $C = 90\%$ $(100(1 - 0.10))$.
2. Lower confidence bound because H_A is a "greater than."
3. $Z^* = {-}1.282$ as found with

```
> ( distrib(0.90,type="q",lower.tail=FALSE) )
[1] -1.281552
```

4. $45 - 1.282\frac{7}{\sqrt{40}}$, $45 - 1.42$, or 43.58
5. One is 90% confident that the mean life for ALL batteries with the additive is more than 43.58 months. By confident, it is meant that 95% of all 95% confidence regions will contain μ.

13.3 Hypothesis Tests and Confidence Region Relationship

An alternative conceptualization of confidence intervals can show how confidence regions and hypothesis tests are related. This conceptualization rests on considering the sample means that would be "reasonable to see" from populations with various values of μ. A graphic is contructed below using the Square Lake population as an example and assuming that σ is known (=31.49), $n = 50$, and 95% CIs are used.

First, compute the most common 95% of sample means assuming that $\mu = 70$; i.e., $70 \pm 1.960\frac{31.49}{\sqrt{50}}$ or $(61.27, 78.73)$. This range is plotted as a vertical rectangle centered on $\mu = 70$ (left-most rectangle) in Figure 13.4-Left. Next, compute and plot the same range for a slightly larger μ (e.g., with $\mu = 71$, plot $(62.27, 78.73)$). Then repeat these steps for sequentially larger values of μ until a plot similar to Figure 13.4-Left is constructed.

Consider very carefully what Figure 13.4-Left represents. The vertical rectangles represent the ranges of the most common 95% of sample means (values read from the y-axis) that will be produced for a particular population mean (value read from the x-axis). In essence, each vertical line represents the sample means that are likely to be observed from a population with a given population mean (x-axis).

Now suppose that $\bar{x}$=100.04 (Table 2.2). Draw a horizontal line across Figure 13.4 at this value and then draw vertical lines down from where the horizontal line first enters and last leaves the band of possible sample means (Figure 13.4-Right). The x-axis values that these vertical lines intercept are an approximate 95% CI for μ. The approximation is only as close as the intervals used to construct the rectangles (i.e., 1.0 mm were used here). However, the results from this graphical approach (i.e., $(92, 108)$) compare favorably to the previous results from using the CI formula (i.e., $(91.27, 108.73)$).
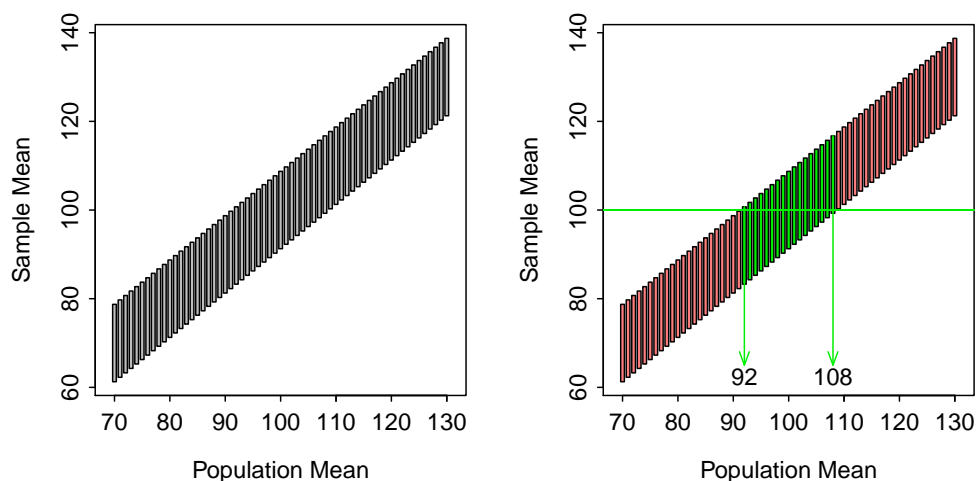

Figure 13.4. Range (95%) of sample means that would be produced by particular population means in the Square Lake fish length example (**Left**) and the ranges intercepted by $\bar{x} = 100.04$ mm (**Right**).

Surely, the CI formula (Section 13.2) is a more efficient and precise way to construct confidence intervals. However, this conceptualization illustrates that a confidence interval (or region, more generally) consists of population means that are likely to produce the observed sample mean. Thus, a confidence region represents possible null hypothesized population means that WOULD NOT BE rejected during hypothesis testing.

⋄ **A confidence region represents null hypothesized values that would NOT be rejected.**

13.4   Precision and Sample Size

The width of a confidence interval explains how precisely the parameter is estimated. For example, narrow intervals represent precise estimates of the parameter. The width of a confidence interval is directly related to the margin-of-error which depends on (1) the standard error and (2) the scaling factor. As either of these two items gets smaller (while holding the other constant), the width of the confidence interval gets smaller.

A small standard error means that sampling variability is low and the parameter is precisely estimated by the statistic. Smaller standard errors are obtained only by increasing the sample size. A smaller standard deviation would also result in a smaller SE, but the standard deviation cannot be made smaller (i.e., it is an inherent characteristic of the population).

A smaller scaling factor is obtained by reducing the level of confidence. For example, a 90% confidence interval uses a $Z^* = \pm 1.645$ whereas a 95% confidence interval uses a $Z^* = \pm 1.960$. Thus, decreasing the confidence level narrows the CI. However, reducing the level of confidence will also increase the number of confidence intervals that do not contain the parameter. Thus, reducing the level of confidence may not be the best choice for narrowing the confidence interval.

The margin-of-error formula can be solved for n.

$$m.e. = Z^* \frac{\sigma}{\sqrt{n}}$$
$$\sqrt{n} = \frac{Z^* \sigma}{m.e.}$$
$$n = \left(\frac{Z^* \sigma}{m.e} \right)^2$$

This formula can be used to find the n required to estimate μ within $\pm m.e.$ units with C% confidence assuming that σ is known. For example, suppose that one wants to determine n required to estimate the mean length of fish in Square Lake to within 5 mm with 90% confidence knowing that the population standard deviation is 34.91. From this, $m.e.$=5, σ=34.91, and Z^*=1.645 (found previously for 90% confidence).[2] Thus, $n = \left(\frac{1.645*34.91}{5} \right)^2 = 131.91$. Therefore, a sample of at least 132 fish from Square Lake should be taken to meet these constraints. Note that sample size calculations are always rounded up to the next integer because rounding down would produce a sample size that does not meet the desired criteria.

⋄ **Always round sample size calculations up to the next integer.**

The margin-of-error and confidence level in these calculations need to come from the researcher's beliefs about how much error they can live with (i.e., chance that a confidence interval does not contain the parameter) and how precise their estimate of the mean needs to be. Values for σ are rarely known in practice (because it is a parameter) and estimates from preliminary studies, previous similar studies, similar populations, or best guesses are often used instead. In practice, a researcher will often prepare a graph with varying values of σ to make an informed decision of what sample size to choose.

---

[2]Strictly, $Z^*\pm$=1.645, but the sign is inconsequential due to squaring in the sample size formula.