6.9.3 Parametric Estimation with a Confidence Interval

Parameter estimation with a confidence interval[1] uses sample data to estimate population parameters. The random sample in Table 6.1 can be used to estimate the population mean (average) weight at the time. Because the sample exhibits mean weight 62.07 kg, sample standard deviation[2] of the observations 7.05 kg, and standard deviation of the sample mean 1.82 kg, and because this mean comes from a sum of weights we assume to be independent and identically distributed, the central limit theorem tells us this sample mean is asymptotically normal (i.e., more and more normal as sample size increases) with these parameters. A normally-distributed random variable falls within 1.960 standard deviations of its mean 95% of the time, and within 2.576 standard deviations 99% of the time. Thus, we might conclude from this sample with 95% confidence that the population mean lies within 58.50 and 65.64 kg. However, because our sample size is quite small and the observations themselves are not necessarily normally distributed, a more-cautious nonparametric (i.e., lacking this rigid assumption of normality) assumption uses the more-conservative t-distribution[3]. With sample size-induced 14 degrees of freedom, the t distributed random variable falls within 2.145 standard deviations of this mean 95% of the time, and within 2.977 standard deviations 99% of the time. So, a more cautious (and robust) confidence interval is that 95% of the time the population means lies within 58.17 and 65.97 kg.

To make this prediction more precise, we need to increase our random sample size. Unfortunately, to double our precision (that is, halve the confidence interval width) we would expect to have to quadruple our sample size.

-break-

[1] https://en.wikipedia.org/wiki/Confidence_interval
[2] https://en.wikipedia.org/wiki/Standard_deviation
[3] https://en.wikipedia.org/wiki/Student%27s_t-distribution

6.10 Inferential Statistics

Inferential statistics[4] assesses how much some state can be expected to vary, making statements in terms of probabilities of exceedance of a given threshold. Inferential statistics also uses probability to make decisions about whether or not some relationship exists between two types of state. A null hypothesis states there is no relationship. Based on probabilistic modeling, the null hypothesis may be accepted incorrectly, a *Type I*, or *false positive error*, or the null hypothesis may be rejected incorrectly, a *Type II,* or *false negative error*. Each type of error has its cost, and a test is designed to recognize these costs in setting the limits governing a decision. Reducing the risk of committing an error can be achieved by coarsening the decision rule, or gathering larger samples upon which to base statistics.

We have been told that in 2014 the average population weight of US females over 20 years old was 76 kg (about 169 lb) (CDCP 2017). We wonder, has this population significantly changed average weight since the 1975 random sample was collected?

Suppose the data in Table 6.1 has been hidden from us (more on this in a moment).

Statistical Hypothesis Test (Deterministic, Descriptive)

Null Hypothesis H_0: 2014 average weight of 76 kg is no different than it was when the 1975 sample was collected.
Alternative Hypothesis H_1: average weight is different from 76 kg.
Significance level, or probability of making a Type I error: rejecting H_0 even though it's true: 99% (that is, a *critical probability* $\alpha = 0.01$).
Test statistic: mean of a random sample of 15 1975 observations.
Decision Rule: If sample mean (62.07 kg) is within 2.977 sample mean standard deviations (1.82 kg) from the hypothetic mean 76 kg (in the interval 70.58 to 81.42 kg), do not reject H_0, otherwise reject it.

Now we open the data envelope. A sample mean of 62.07 kg is not within our interval, so we reject H_0.

[4] https://en.wikipedia.org/wiki/Statistical_hypothesis_testing

We create the hypothesis test significance level[5] *before* we see the test statistic - otherwise, the test statistic is not a statistic, but a known constant, and then we can rig our significance level to conclude whatever we please.

“But, wait, obviously this 2014 average population weight has increased. It isn’t even within the range of the 1975 random sample.” This may seem intuitive, but the hypothesis test we have conducted gives us statistical confidence [sic] in our common sense.

When you read about experimental results (especially, for some reason, in medical literature), you should be suspicious when you find statements such as “this result has 99.99% significance” (as might have been claimed with our example). This is a symptom of misuse: a likely explanation is that “this is as far as we could push our realized statistic to support our preferred hypothesis test conclusion.” If the scientific method is properly employed, a test is designed *before* the sample data is viewed. The potential costs of committing a Type I or Type II error are assessed, and the significance level is fixed. The decision and its potential risks of having made an error are known once the sample data are known. Otherwise, are you saying that the decision might be different if we change our prior estimates of the costs of making an error? That’s an entirely different analysis. It is human nature to seek certainty, and scientists are human; there is continuing debate on the misuse of statistical methods (Baker, 2016).

We might also be criticized for unstated assumptions. For instance, the 1975 sample is from adult U.S. females aged 30-39, while the 2014 statistic applies to U.S. females aged 20 years and over. This may or may not be a serious problem, and should certainly be part of the documentation accompanying the statistical work.

-break-

[5] https://en.wikipedia.org/wiki/Statistical_significance

6.33 Where to go from here

The INFORMS journal ***INTERFACES*** (renamed in 2019 the ***INFORMS Journal on Applied Analytics***) is aggressively edited for clarity of exposition and includes many modeling examples explained with care; refreshingly, some recounting of false starts and lessons learned also appear. The INFORMS newsletter ***OR/MS Today*** features some articles about modeling, including the shared experiences of clients and modelers, some analysis puzzles, and entertaining features about the operations research (and modeling) craft. To access more of our huge open literature of articles and textbooks, some mathematical preparation will be necessary, including at least algebra, probability, statistics, and elementary modeling. An introductory modeling course with a lot of homework drills does wonders: *the best way to learn how to model is to try modeling*.

> ”Anyone who has never made a mistake
> has never tried anything new.” Albert Einstein