Partial Synopsis of 6.1a - One Mean

Background

First, recall from Chapter 5 that

- if a **random variable** X is **normally distributed** with mean μ and standard deviation σ in a population, then the **sampling distribution** of $\overline{X}$ is also normal with the same mean μ, but with standard deviation $\dfrac{\sigma}{\sqrt{n}}$ (called the **standard error**), for <u>*any*</u> sample size n.

Furthermore, via the much more general **Central Limit Theorem**, it is also true that

- as long as X has finite mean μ and standard deviation σ, the same result will *approximately* hold for "large" n (say ≥ 30), and
- the same is true even if σ is unknown and replaced by its estimate, the sample standard deviation s. [[<u>CAUTION</u>: This last fact is *not* true if n is small! Later…]]

Also recall that **parameters** in general – such as the population mean μ and population standard deviation σ of a numerical **random variable** X, or population probability π of "Success" of a variable X having **binary outcomes** – are by definition numerical **characteristics of a *population***. One goal of Statistics (as a field) is **"parameter estimation"** (e.g., by $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\pi}$, respectively) via numerical **characteristics of a random *sample***, i.e., **statistics**, such as sample mean $\overline{x}$, sample standard deviation s, and sample binomial proportion p, respectively.

Parameter Estimation

For the sake of simplicity, we *temporarily* confine our discussion to the population *mean* μ of a ***normally-distributed*** random variable, with a value of $\overline{x}$ calculated from a single sample. This **point estimate** of μ can be improved to an **interval estimate**, i.e., an interval centered at $\overline{x}$, that contains μ with a high "probability" (or more precisely, **confidence level**), say 95%; the complementary 5% is the **significance level**. To compute it,

- multiply the **±.025 critical values** (that is, the symmetric positive and negative z-scores $\pm z_{.025}$ that divide the region under the **standard normal distribution** $N(0, 1)$ into a central area of 0.95, and symmetric left and right tails areas of .025 each; these critical values turn out to be ±1.96), times the **standard error** $\dfrac{\sigma}{\sqrt{n}}$. The resulting (positive) product is known as the **95% margin of error**.

Extending the point estimate $\overline{x}$ by this margin of error symmetrically in both directions gives the corresponding **95% confidence interval** estimate of μ. That is, a *sample-based* interval $(\overline{x} -$ margin of error, $\overline{x} +$ margin of error$)$ that contains μ with 95% "confidence." (More correctly, from among an arbitrarily large number of such samples, each with its own $\overline{x}$ and confidence interval, the *probability* of the event that "a *random* confidence interval contains the true value of μ" approaches 95%, "in the limit.")

Hypothesis Testing

Confidence intervals can be used in a formal test of a **null hypothesis** H_0: $\mu = \mu_0$ versus the complementary two-sided **alternative hypothesis** H_A: $\mu \neq \mu_0$, where the **null value** μ_0 is usually some standard reference amount. For example, if testing for mean pH on a scale of 0 (acid) to 14 (alkaline), the null value might be taken as 7 (neutral). By definition, the 95% confidence interval contains the true value of μ with 95% confidence. Thus, at the 5% significance level,

- if the interval (is one of the 5% that) *does not* contain μ_0, it then follows that the sample serves as evidence to *refute* the null hypothesis, and it can be ***rejected***, i.e., the difference between the true mean value μ and the hypothesized null value μ_0 is **statistically significant**. In other words, it is a "genuine" difference, beyond what can be expected simply by random chance.
- Otherwise, if it *does* contain μ_0, the evidence is "too close to call," and the null hypothesis cannot be rejected, i.e., it must be ***retained***, and the result is **not statistically significant**.*

Note: Changing the significance level from .05 to some other value α (equivalently, changing the confidence level from .95 to $1 - \alpha$) will change the critical values *(but not the standard error)*, and hence the margin of error, so that the confidence interval will become wider (if α decreases) or narrower (if α decreases). Hence, the corresponding hypothesis test will be either more or less **conservative**, respectively.

A second way to test a null hypothesis is to determine its **acceptance interval** and complementary **rejection (or critical) region**. That is, a symmetric interval centered at the null value μ_0, inside of which a random sample mean $\bar{x}$ is theoretically expected to fall with a high probability (again, say 95%)… *IF the null hypothesis is in fact true*. To find it, surround μ_0 with the *same* margin of error as above, both left and right. By its construction, a sample mean value $\bar{x}$ that happens to land inside indicates that the small difference between it and μ_0 could be due to random chance, and is therefore not statistically significant. However, if it lands outside – in the rejection region – the difference is indeed statistically significant.

The ***p*-value** is a way to quantify the strength of the rejection (or non-rejection). It measures the *probability* (hence it is a number between 0 and 1) of finding a *random* sample mean value $\bar{X}$ that is *at least* as far away from the null value μ_0 (in both directions, since the alternative is two-sided) as the $\bar{x}$ actually obtained from the sample… again, *IF the null hypothesis is in fact true*. Hence, a "small" *p*-value (i.e., close to 0) would indicate a low probability that the obtained sample agrees with the null hypothesis, thus yielding evidence to reject it. In particular…

- If the *p*-value of the sample is *less* than the significance level $\alpha = .05$, then ***reject*** the null hypothesis; the difference is **statistically significant** at that level. Moreover, the *smaller* the *p*-value, the *stronger* the rejection, and the *more* statistically significant the finding. See the notes for details.

---

* Note that this is really not the same as "accepting" the null hypothesis, although that term is commonly used in practice. For instance, if a particular study fails to show that a drug works, that does not necessarily mean that the drug is ineffective. Similarly, in the US criminal justice system, if a prosecuting attorney fails to gather enough evidence to convince a jury to reject the hypothesis that "the defendant is innocent" beyond a "shadow of a doubt" (i.e., at some level of significance), it does not necessarily follow that the defendant is truly innocent. If he/she is indeed guilty, then failing to reject is what is known as a **Type 2 error**. Rejecting a null hypothesis that is indeed true (i.e., "innocent") is a **Type 1 error**.