Chapter 11 Optional Material

**THE STANDARD ERROR OF THE MEAN DIFFERENCE FOR INDEPENDENT SAMPLES: A MORE COMPLETE ACCOUNT (OPTIONAL)**

Although, as noted earlier, it is unrealistic to expect σ to be known, for heuristic reasons we describe the shape of the sampling distribution of sample-mean differences when σis known, but we do not expand on this case. We then consider in detail the case when σ is not known.

**Case 1: σ known**

Suppose we have two normally distributed populations with equal variances σ2 and with population means μ1 and μ2, respectively. If we intend to select random samples of size N1 and N2 from these two populations, respectively, then the sampling distribution of all possible sample-mean differences -, as shown in Figure 9.2, will be normally distributed with mean μ1-μ2 and standard deviation

.

Normal Curve

Mean = μ1 - μ2

StDev =

μ1-μ2

Figure 11.3. The sampling distribution of-’s when σ is known.

Note that we can compute, the standard deviation of the sampling distribution of sample-mean differences only if we know σ. When σ ­is *not* known an alternative procedure is needed, just as an alternative procedure was needed when σwas not known in the one-sample case.

**Case 2**: **σ not known**

Suppose we again assume that we have two normally distributed populations with equal variances σ2 and that their population means are μ1 and μ2, respectively. But now assume that the value of σ is not known. Because, as noted in case 1, σ2 is necessary for computing, it is tempting to try to approximate the value of σ2 by using the variance estimators  and  from our two samples. As we saw in the one-sample situation, however, the resulting sampling distribution when such an estimation procedure is used need not be normal. By using a standardization procedure, however, we were able to obtain a new statistic whose sampling distribution could be described in a simple mathematical way as a t distribution. We will again obtain a statistic that has a t distribution using this same step-by-step procedure as in the previous section.

**Step 1: Estimating σ2** **using the variance estimators**  **and** 

In the one-group case of the previous section, we simply used  to estimate σ2, because is known to be an unbiased estimator of *σ2***.** Here we have two unbiased estimators of *σ2*, one from each sample, and we would like to know the best way to use both to obtain an estimate of σ2. Should we use alone, or  alone, or should we combine the two variance estimators in some way and obtain our estimate of σ2 from both? Because our estimate will be more accurate if it is based on a larger sample size, it makes sense to use both samples (both variance estimators) instead of either one alone.

The question we face is *how* to combine  and  to obtain an estimate. Clearly, we would like the estimator that we expect to be more accurate to weigh more heavily in the estimation process, and we would believe in general that the estimator from a larger sample is the more accurate. But as we saw earlier, it is not the sample size that determines the accuracy of a variance estimator. Rather, it is the number of independent pieces of data contained in the sample - that is, the degrees of freedom ν. Therefore, we will weight the individual variance estimators and  not by their sample sizes *N1* and *N2*, respectively, but by their degrees of freedom ν1 = *N1*-1 and ν2 = *N2*-1, respectively. The equation for pooling and using the weights ν1 = *N1*-1 and ν2 = *N2*-1 is

Estimated σ2 =  (11.17)

**Step 2: Estimating the standard error of the mean difference,using σ2**

From the equation for presented in case 1 for σ known,

= .

Using the estimated σ2 derived in step 1 in place of the unknown σ2 in this equation, we obtain

Estimated = 

As in the one-sample t-statistic situation, we will denote estimated as for simplicity. Using the equation for estimated σ2 obtained in step 1, we can write the complete equation for determining as

=  (11.18)

This is the denominator of the two-population, unrelated-samples t statistic with equal variances assumed.

The number of degrees of freedom for this t statistic may be determined by recalling that we are using both sample variance estimators and  to obtain the denominator of this statistic and that has *N1* - 1 degrees of freedom associated with it, whereashas *N2* - 1 degrees of freedom associated with it. Because and come from independent samples, the total number of independent pieces of information contained in the data, and therefore the number of degrees of freedom for this t statistic, is *df* = *df1* + *df2* = (*N1* – 1) + (*N2* – 1) = *N1* + *N2* – 2