Using Statistics to Identify Thieves

Methods of statistics can be used to determine that an employee is stealing, and they can also be used to estimate the amount stolen. The following are some of the indicators that have been used. For comparable time periods, samples of sales have means that are significantly different. The mean sale amount decreases significantly. There is a significant increase in the proportion of "no sale" register openings. There is a significant decrease in the ratio of cash receipts to checks. Methods of hypothesis testing can be used to identify such indicators. (See "How to Catch a Thief," by Manly and Thomson, *Chance*, Vol. 11, No. 4.)

Part 1: Independent Samples with σ_1 and σ_2 Unknown and Not Assumed Equal

This section involves two *independent* samples, and the following section deals with samples that are *dependent*. It is important to know the difference between independent samples and dependent samples.

DEFINITIONS

Two samples are **independent** if the sample values from one population are not related to or somehow naturally paired or matched with the sample values from the other population.

Two samples are **dependent** (or consist of **matched pairs**) if the sample values are somehow matched, where the matching is based on some inherent relationship. (That is, each pair of sample values consists of two measurements from the same subject—such as before/after data—or each pair of sample values consists of matched pairs—such as husband/wife data—where the matching is based on some meaningful relationship.)

Example 1

***Independent Samples:* Proctored Tests and Nonproctored Tests in Online Courses** Researchers Michael Flesch and Elliot Ostler investigated the reliability of test assessment. One group consisted of 30 students who took proctored tests. A second group consisted of 32 students who took tests online without a proctor. The two samples are independent, because the subjects were not paired or matched in any way.

***Independent Samples:* Weights of M&Ms** Data Set 20 in Appendix B includes the following weights (grams) of a sample of yellow M&Ms and a sample of brown M&Ms. The yellow and brown weights might appear to be paired because of the way that they are listed, but they are not matched according to some inherent relationship. They are actually two independent samples that just happen to be listed in a way that might cause us to incorrectly think that they are matched.

Yellow	0.883	0.769	0.859	0.784	0.824	0.858	0.848	0.851
Brown	0.696	0.876	0.855	0.806	0.840	0.868	0.859	0.982

***Dependent Samples:* Heights of Husbands and Wives** Students of the author collected data consisting of the heights (cm) of husbands and the heights (cm) of their wives. Five of those pairs of heights are listed below. These two samples are dependent, because the height of each husband is matched with the height of his wife.

Height of Husband	175	180	173	176	178
Height of Wife	160	165	163	162	166

For inferences about means from two independent populations, the following box summarizes key elements of a hypothesis test and a confidence interval estimate of the difference between the population means.