Two Independent Samples

W. Robert Stephenson
Department of Statistics
Iowa State University

One of the most commonly used statistical techniques is the comparison of two independent samples of measurement data. More specifically, the comparison of the means of two independent samples. This is often the basis for making a decision to go with a particular method, process or supplier. The validity of the procedure hinges on the random selection of items for each sample. The correctness of the subsequent decision rests with the stability of the processes that produce items that can be selected for the samples.

Example: In the handout "Display and Summary of Data," data on the temperatures of electric irons set at 450^oF are given. These data are reproduced below along with similar data from irons with thermostats from a second supplier.

Supplier 1				Supplier 2			
445.0	438.0	441.8	450.7	444.0	454.9	450.0	459.1
453.0	435.1	459.7	448.4	454.7	430.3	458.8	465.4
451.1	448.7	464.6	448.3	451.7	443.3	456.9	454.9
434.7	453.8	454.5	455.0	436.7	451.1	449.9	459.3
463.1	454.8	444.6	444.8	469.1	455.0	449.4	438.6
452.9	460.3	458.4		466.3	438.8	436.1	

One can use dot plots or box plots to make a visual comparison of the data from the two suppliers. Side-by-side box plots are given at the end of this handout. From that graph, the central values for the two samples look quite similar. The thermostats from Supplier 2 show slightly more variation than those from Supplier 1. Both data distributions appear to be fairly symmetric.

To formalize the comparison, one can summarize the data in terms of sample means and sample standard deviations. This is done in the below (summary statistics are rounded).

Supplier 1	Supplier 2
$n_1 = 23$	$n_2 = 23$
$\overline{Y}_1 = 450.5^oF$	$\overline{Y}_2 = 451.1^oF$
$s_1 = 8.28$	$s_2 = 10.29$

These summaries verify that the sample of thermostats from Supplier 2 is slightly more variable ($s_2 = 10.29 > s_1 = 8.28$). Also, the sample of thermostats from Supplier 2 has a slightly higher (more off target) mean ($\overline{Y}_2 = 451.1 > \overline{Y}_1 = 450.5$). The statistical question becomes: Is the difference in the two sample means an indication of a true difference in suppliers or can such a difference be explained by random sampling error?

The comparison of interest is $\overline{Y}_1 - \overline{Y}_2$. This difference, -0.6^oF in this example, is compared to the standard error of the difference of two sample means. This standard error is given by:

$$se(\overline{Y}_1 - \overline{Y}_2) = s_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

For our example,

$$s_p = \sqrt{\frac{22(8.28)^2 + 22(10.29)^2}{44}} = \sqrt{87.22} = 9.34$$

and

$$se(\overline{Y}_1 - \overline{Y}_2) = 2.754$$

The difference in sample means can be evaluated in two ways.

1. **Confidence Interval**

   $$(\overline{Y}_1 - \overline{Y}_2) \pm t_{df,\frac{\alpha}{2}} * se(\overline{Y}_1 - \overline{Y}_2)$$

   where for all except very small samples, $t_{df,\frac{\alpha}{2}} \doteq 2$ for 95% confidence.

   The 95% confidence interval for the difference between the means for Supplier 1 and Supplier 2 is: $-0.6 \pm 2(2.754)$ or (–6.1, 4.9). Since this interval contains zero, no difference between the mean temperatures for the two suppliers is inferred. If the entire interval is on one side of zero, then the difference between the two suppliers' means is said to be statistically significant.

2. **Test of Hypothesis**

   In a formal statistical test of hypothesis, the difference in sample means is standardized by dividing by the standard error to produce a test statistic. This test statistic is then compared to a value from the tabulation of the t-distribution in order to assess statistical significance. The form of the test statistic is:

   $$t = \frac{(\overline{Y}_1 - \overline{Y}_2)}{se(\overline{Y}_1 - \overline{Y}_2)}$$

   If the absolute value of this test statistic is greater than the tabulated value from a t-distribution, the difference in sample means is said to be statistically significant. Otherwise, the difference is attributed to random sampling error. The following rule of thumb can be used when a t-distribution is unavailable.

   - If $|t| < 2$, there is no statistically significant difference between the means of the two samples.
   - If $|t| > 3$, there is a statistically significant difference, that is the difference is so large it cannot be explained by random variation alone.

- If $2 \leq |t| \leq 3$, statistical significance depends on the number of observations and the chance of making an error.

Computer programs, like JMP, convert the t-test statistic into a probability value, P-value. This is a measure of how likely it is to get a difference in sample means larger than the one observed when random sampling from identical frames. The smaller the P-value, the less likely random sampling can explain the difference. Thus small P-values lead one to declaring the difference in sample means to be statistically significant.

Below is the output of JMP: **Basic Stats** → **Oneway** with Temp as the Y, Response and Supplier as the X, Grouping. Choose the Means/Anova/t Test from the red triangle pull down.

t Test

Assuming equal variances

	Difference	t Test	DF	Prob > $\|t\|$
Estimate	−0.565	−0.205	44	0.8383
Std Err	2.754			
Lower 95%	−6.12			
Upper 95%	4.99			

Note that there are slight differences in the calculated values since less rounding is done in JMP. Also the high P-value (P=0.84) indicates that it is very likely that the frames from which the samples were taken are identical (same mean, standard deviation and shape). Implicit in the formal statistical analysis presented above is an assumption that data are normally distributed. If this is not true, then the true P-value and the true confidence level will be different from what is reported.

A note on enumerative and analytic purposes

The comparison of two independent samples can be enumerative in that the difference, or lack of difference, seen in the samples can be inferred to the frames from which the samples are randomly selected. The standard error of the difference in sample means quantifies the uncertainty introduced by using random samples instead of complete coverage. Most comparisons do not stop at a simple description of the samples or even with inference to the frames. Instead, based on a comparison like the one above, decisions are often made to keep or change suppliers. This has an analytic purpose since future production will be affected. Information on the stability of the processes producing the frames from which the random samples are taken is essential. The standard error does **NOT** quantify the uncertainty introduced by unstable processes.