In: Encyclopedia of Biostatistics. P. Armitage and T. Colton (eds). New York: Wiley, 3244-3246.

Student, who used the divisor n in computing the variance. Using the unpaired t-test we obtain a $t = 1.86$ with 18 df and a ***P* value** of 0.0792. By pairing the observations, as we should, the result is $t = 1.58/(0.39/\sqrt{10}) = 4.06$ with 9 df and a P value of 0.0028. It is better to report the mean of the differences and its standard deviation rather than showing only the t statistic or the P value (or worse, NS for "not significant", or some number of asterisks!).

Robustness

The **robustness** properties correspond to those of the one-sample t test. The effect of nonnormality is fairly small if n is at least 30, since the distribution of d will be close to normal in that case. If one difference (or a few) appear to be quite large (i.e. **outliers**) the results can be affected. Outliers can be considered a form of nonnormality. They affect the variance of the observations, and can also affect the **skewness** of the distributions. The P values reported from an analysis are often given as $P < 0.05$ or $P < 0.01$. Since the P value depends on the behavior of the distribution in its tails, nonnormality generally means that statements such as $P < 0.001$ are rarely accurate (the quoted P value for the example thus should be regarded as $P < 0.01$). Lack of independence among the pairs (which might arise if multiple members of a litter or a family were included in a study, i.e. clustering) can seriously affect the level of the test. If the correlation between any pair of differences is γ, the variance of the differences is $\sigma^2(1 + 2\gamma)$. Thus, the estimated variance is biased. If $\gamma > 0$, the denominator of the t statistic is too small, and the significance levels are incorrect. If the correlation holds only among certain pairs (e.g. independent clusters would lead to a block diagonal **covariance matrix**), the analysis is more complex, but the estimated variance is still biased. Lack of common variance in X and Y does not formally affect the analysis. However, unless the primary interest is in the difference between the observations, the lack of common variance indicates that X and Y do not have the same distribution, although they might have the same mean. If the variance differs over the pairs, heteroscedasticity concerns arise (*see* **Scedasticity**). Rosner [4] has suggested a **random effects** model which accounts for this. Missing values can create problems. Usually, only one member of the pair is missing. If the missing value is related to the mean value within the pair, the missingness is not random, and the t test is affected (*see* **Missing Data**). For further discussion of these points, see Miller [2] or Madansky [1].

Several alternatives exist to the paired t test. These are useful if the distribution is not normal and there is concern that this may affect the performance of the test. The **sign test** uses the number of positive (or negative) signs as a **binomial** variable with probability parameter 1/2 under H_0, and, for large samples, computes the **standard normal deviate**, z, to test H_0. The **asymptotic relative efficiency** (ARE) of this test is 0.637 when the differences are normal. The **signed-rank** test ranks the absolute values of the differences and sums the ranks corresponding to the positive (or negative) signs. The ARE of this test is 0.955. The **normal scores** test replaces the ranks of the differences by their expected values under normality and computes a t test on these. Its ARE is 1.0. For observations from nonnormal distributions, the efficiencies of the **nonparametric** procedures may be higher than indicated here and the t test can be very inefficient.

References

[1] Madansky, A. (1988). *Prescriptions for Working Statisticians*. Springer-Verlag, New York.

[2] Miller, R.G. (1985). *Beyond Anova*. Wiley, New York.

[3] Moses, L. (1985). Matched pairs t-tests, in *Encyclopedia of Statistical Sciences*, Vol. 5, S. Kotz & N.L. Johnson, eds. Wiley, New York, pp. 289-203.

[4] Rosner, B. (1982). A generalization of the paired t-test, *Applied Statistics* **31**, 9-13.

[5] Senn, S. & Richardson, W. (1994). The first t-test, *Statistics in Medicine* **13**, 785-803.

HENRY HSU & PETER A. LACHENBRUCH

Pairwise Concordance *see* Twin Concordance

Pairwise Independence

Suppose that two variables are considered as potentially **explanatory** for a further variable, called the

Table 1 Dependence reversal because of strongly associated explanatory variables

	C, site				Overall; that is, summed over sites	
	$k = 1$		$k = 2$			
	B, treatment		B, treatment		B, treatment	
A, outcome	$j = 1$	$j = 2$	$j = 1$	$j = 2$	$j = 1$	$j = 2$
$i = 1$ (success)	96 (96%)	600 (60%)	400 (40%)	4 (4%)	496 (45%)	604 (55%)
$i = 2$	4	400	600	96	604	496
Sum	100	1000	1000	100	1100	1100
Odds ratio	16		16		0.67	

response, and that the dependence of the response on each of the variables taken alone and on both acting jointly is of main interest. For an appropriately chosen scale and measure of dependence, suppose furthermore that the effects of both variables turn out to be (essentially) additive (*see* **Additive Model**). This means that the effect of one of them on the response is (nearly) the same no matter at which level the other explanatory variable is fixed. Often, this is described as the absence of an **interaction** but the presence of two main effects.

An important role of pairwise **independence** of explanatory variables is then as follows: it is certain that no dependence reversal can occur for the (nearly) additive effect of one of the explanatory variables in comparison with the effect of this variable taken alone. To put it differently, if the explanatory variables are nearly independent, and have essentially additive effects on the response, then the overall effect of just one of them coincides at least qualitatively with the corresponding effects considered conditional given the other variable. A strong reversal of treatment success as related to variable B occurs instead in the 2^3 contingency table displayed in Table 1, since the explanatory variable pair B, C is highly dependent.

For both discrete and for continuous responses, further discussions of dependence reversal in spite of essentially additive main effects are to be found, for instance, in Snedecor & Cochran [5, p. 472], Good & Mittal [1], Wermuth [6, 7], and Guo & Geng [2]. In a **contingency table** context early insights are due to Yule [8] and Simpson [4], (*see* **Simpson's Paradox**).

Mutual Dependence in Spite of Pairwise Independences

In general, no mutual independence results even if several variables are all pairwise independent. Instead, more complicated types of dependencies may still exist, which are often called higher-order interactions. An important implication is that methods of analysis relying completely on pairwise associations, **correlation**-based techniques or **correspondence analysis**, will overlook the existing dependencies in such situations and are therefore likely to lead to misleading interpretations.

Table 2 Symptoms after LSD intake: mutual dependence and pairwise independence

	C, dimming of consciousness			
	$k = 1$ (yes)		$k = 2$	
	B, distorted thinking		B, distorted thinking	
A, distorted affective behavior	$j = 1$, (yes)	$j = 2$	$j = 1$, (yes)	$j = 2$
$i = 1$ (yes)	21	5	4	16
$i = 2$	2	13	11	1
Odds ratio	27.30		0.023	

An empirical example with four **binary** variables is due to Lienert [3]. He reported on symptoms after LSD intake. The 2^3 contingency table shown in Table 2 is an adaptation of his results. The three transient symptoms, recorded to be present (level 1) or absent (level 2), are distortions in affective behavior (A), distortions in thinking (B), and dimming of consciousness (C). There is a strong three-way interaction, as reflected for instance in the quite distinct **odds ratios** at the two levels of C; at the same time, the frequencies in the three marginal tables show all three symptom pairs as being close to independence.

With completely randomized designs (*see* **Randomization**) it will typically occur that – at the time a study starts – not only observed variables but also unobserved variables will essentially be both pairwise, and mutually, independent. Note, however, that even with this technique it is not possible to avoid dependencies with unobserved intermediate variables, i.e. with unrecorded variables related to both treatment and outcome, but occurring unnoticed before observing outcome. Typical examples are noncompliance of some patients (*see* **Compliance Assessment in Clinical Trials**) or, more generally, unrecorded treatment effects or changes in measurement devices before treatment outcome is established.

References

[1] Good, I.J. & Mittal, Y. (1987). The amalgamation and geometry of two-by-two contingency tables, *Annals of Statistics* **15**, 694–711.

[2] Guo, J. & Geng Z. (1995). Collapsibility of logistic regression coefficients, *Journal of the Royal Statistical Society, Series B* **57**, 263–267.

[3] Lienert, G.A. (1970). Konfigurationsfrequenzanalyse einiger Lysergsäurediäthylamid-Wirkungen, *Arzneimittelforschung* **20**, 912–913.

[4] Simpson, E.H. (1951). The interpretation of interactions in contingency tables, *Journal of the Royal Statistical Society, Series B* **13**, 238–241.

[5] Snedecor, G.W. & Cochran, W.G. (1967). *Statistical Methods*, 6th Ed. Iowa State University Press, Ames.

[6] Wermuth, N. (1987). Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable, *Journal of the Royal Statistical Society, Series B* **49**, 353–364.

[7] Wermuth, N. (1993). Association structures with few variables: characteristics and examples, in *Population Health Research*, K. Dean, ed. Sage, London, pp. 181–203.

[8] Yule, G.U. (1903). Notes on the theory of association of attributes in statistics, *Biometrika* **2**, 121–134.

(*See also* **Statistical Dependence and Independence**)

NANNY WERMUTH

Panel Study

In a panel study a number of individuals are followed for a given period of time. At each of a predetermined set of time points several measurements on each individual are taken. Data obtained from a panel study are called *panel data*. A panel study designed to have observations at k time points is called a k-wave panel design. Under this definition the term *panel study* could be used to refer to a large range of studies in biostatistics, particularly in epidemiology and **clinical trials**, although in many cases the term panel study is not used. The main advantage of a panel study is that individual changes over time can be modeled and the unobserved heterogeneity across individuals and over time can be taken into account.

In many panel studies, especially those lasting for a long period, attrition or loss to follow-up is an important issue. To keep the study population at a proper size during the study, two variations of the simple panel study can be used [3]. One is the *rotating panel study*, which replaces a part of the previous panel by a new panel at some time points and each individual only stays in the study for a certain period. Another is the *split panel design*, which recruits a new panel at some time points and keeps following all the panels until the end of the study.

Two important design issues are the calculation of sample size and the choice of time points to take the observations. For studies with continuous or categorical outcomes, standard sample size calculation procedures for repeated measurement models can be used. For panel studies that measure time to event (*see* **Survival Analysis, Overview**) the exact time