EEOS 601, Prob & Stats
Handout 2
Revised: 1/24/11
EDG Homepage

STATISTICAL DEFINITIONS

TABLE OF CONTENTS

List of Figures

Mathematical Symbols

See **http://en.wikipedia.org/wiki/Table_of_mathematical_symbols**

A’ transpose of a matrix or vector A’ or, more rarely, the **complement** of an event A

A^c the complement of Event A $P(A)=1-P(A^c)$

$\binom{n}{r}$ binomial coefficient $\frac{n!}{(n-r)!\,r!}$

exp(x) e^x, where **e** is the base of natural logarithms

iid identically independently distributed

ln(x) **Natural logarithm** of x, to the base **e**. It may also be represented as log(x), but log(x) can be a logarithm to any base. Base 10 and 2 are also common.

$\in$ ‘is a member of’, ‘in’ See **membership**

$\cap$ **intersection**

$\wedge$ Logical conjunction; The statement $A \wedge B$ is true if A and B are both true; else it is false.

$\vee$ Logical disjunction; The statement $A \vee B$ is true if A or B (or both) are true; if both are false, the statement is false.

$\varnothing$ **the null set**

$\cup$ **union**

$\therefore$ therefore

Definitions

***A priori* contrast** A planned comparison, specified before the experiment was conducted, with implications for the interpretation of tests in **ANOVA** and other statistical tests.

***A posteriori* contrast** Any unplanned comparison carried out after collecting and examining patterns in the data. These statistical tests usually require an adjustment of the alpha level for the test decision. See **multiple comparison tests**, **family-wise error rate**.

Absorbing Markov chain **Roberts** (1976, Theorem 5.3) A **chain** is absorbing if and only if it has at least one **absorbing state**, and from every nonabsorbing (**transient**) state it is possible to reach some absorbing state. *cf.,* **Markov chain**, **fundamental matrix**.

Accuracy refers to the difference between the measured or computed value and the true value; it is also called the **systematic error** (*cf.* **precision**)

ACE Abundance-based coverage estimators, a species richness method reviewed by **Colwell & Coddington (1994)** and **Hughes et al. (2001)**.

Adjusted R-squared *cf.*, **R-squared**

Akaike Information Criterion (AIC) A measure of goodness of fit of a regression models with a strong penalty for the number of parameters in the model. See Ripley on model choice. AIC is intended for nested models (with one model a proper subset of another): **http://www.stats.ox.ac.uk/~ripley/ModelChoice.pdf** *cf.*, **BIC**

Algorithm A set of well-defined steps designed to produce an outcome.

Alpha level The probability of **Type I error**. An alpha level of 0.05 is the pragmatic cutoff adopted both by Fisher & Neyman and Pearson to decide whether a result is significant or not, but the significant/not significant dichotomy is not recommended in current statistical parlance.

Alternative or alternate hypothesis A hypothesis that is often complementary to the **null hypothesis**. For example, the null hypothesis might be $\mu_i = \mu_j$, and the two-tailed (=two-sided) alternate hypothesis might be $\mu_i \neq \mu_j$. There might also be a one-tailed (one-sided) alternative hypothesis that $\mu_i > \mu_j$. The alternate hypothesis usually must be specified to calculate **Type II error (β)** and the **power** (1-β) of a statistical test.

Analysis of covariance (**ANCOVA**)

ANOVA Analysis of Variance. Invented by **Fisher**. A partitioning of sums of squared deviations from means that allows tests for differences in means and differences in variance. ANOVA is a form of the **general linear model** with explanatory variables (formerly called independent variables) or factors that are categorical. Most ANOVA problems can also be analyzed as regression problems with categories coded as indicator or dummy variables, but regression also allows continuous explanatory variables to be included in the design.

Assumptions A) Equal variances among subgroups (also called homogeneity of variance or homoscedasticity) B) Normally, identically independently distributed errors. For specific ANOVA models, there are further assumptions. For example, in an unreplicated randomized block design, to test the main effects the test is based on the assumption that block and treatment effects are additive (*i.e.*, no interaction)

Which assumptions matter? Unequal variance is a major problem for ANOVA, but the results can be **robust** if sample sizes are equal. **Winer *et al.* (1991, Table 3.8, p 102)** provide an example of why equal sample size is important. The table, adapted from **Glass et al. (1972)**, states that with equal sample sizes alpha levels are unaffected, 'Effect on α: 'Very slight effect on α, which is seldom distorted by more than a few hundreths. Actual α seems always to be slightly increased over the nominal α'. [But this has been documented by others to not be the case by Wilcox and others] With unequal n's, the alpha levels can be affected, with α increased if the smaller group has the larger variance. **Sleuth Display 5.13** (top row) indicates that with unequal variance Type I error can be much higher than nominal (7.1% vs. the nominal 5%) or much less than nominal level (0.4%). **Quinn & Keough (2002, p. 193)** review a study by Wilcox documenting that unequal variances can affect Type I error, **and** the problem is much worse with unequal sample sizes. If the smaller group has the larger variance, the probability of Type I error could be one in four or larger.

Blocked ANOVA (**randomized block ANOVA**)

Model I (Model 1) Fixed effects model. Each level of each explanatory factor is assumed to add a fixed amount to the mean.

Model II (Model 2) Random-effects model. ANOVA is used to assess whether different levels of a factor contribute to the variance. For example, in assessing plant height, there could be among plant variance in height within a local patch, an additional variance component due to patch-to-patch variability within an area, and finally an additional variance component due different areas. In factorial models, the appropriate statistical tests depend on whether the factors are fixed or random. **Ramsey & Schafer 1997 p. 130** "(1) Is inference desired to a larger set from which these groups are a sample, in which case one must also be concerned about (2) are the groups (operators) truly a random sample from the larger set? A yes to (1) would indicate that a random effects model should be used, but could only be justified if the answer to (2) was yes. See also **random effects**

Mixed model (Model III) A model including both **fixed** and **random effects**. A **nested or hierarchical ANOVA** can be an example of a mixed model, with treatment effects being fixed and the variability among experimental or survey units being a random factor. Mixed models can be treated as a **general linear models**, assuming normally and independently distributed errors, with model parameters estimated through minimization of sums of squares. Mixed models can also be analyzed using **generalized linear models**, which usually estimate parameters through **maximum likelihood**. See **http://www.statsoft.com/textbook/stvarcom.html** for a brief discussion of generalized linear modeling approaches to mixed models. SPSS's GLM (UNIANOVA) can be used for a general linear mixed model, and SPSS's program 'mixed' can be used for a generalized linear model. The generalized linear model allows a number of different ways of handling the variance-covariance estimates.

Nested (Hierarchical) Involves more than one observation per experimental unit. The **degrees of freedom** must be partitioned into error and 'experimental unit within treatment' sources of variation. Note that some nested ANOVA models treat both experimental or survey units and treatment levels as fixed factors. A mixed model nested ANOVA treats units as random factors and treatment levels as fixed factors. The main effect of the fixed factor is tested over the experimental unit within treatment mean square.

One-way One explanatory factor or category

Two-way Two explanatory factor or categories

Factorial Two or more explanatory categories. A full factorial model is sometimes called a crossed ANOVA.

Randomized block Each level of treatment is included randomly allocated within each block. To test the full randomized block model, including block x treatment interaction, requires that there be replicates of each treatment within each block.

Repeated measures The same experimental units (e.g., patients, quadrats) are sampled more than once (e.g., clinical trials in which a patient is given a placebo and a test drug). Student's paired *t* test would be appropriate if there were just two variables measured on each subject.

Split plot Multiple treatment levels are nested within a larger treatment level. For example, an entire field could receive a given level of fertilizer, and different watering levels could be used on different portions of the field. Or, different greenhouses could be used to control temperature for a large number of trays of plants, and then different watering levels and fertilizer levels could be used within different areas or blocks of each greenhouse. The ANOVA table is often split, with tests of the main plot being based on a partition of the degrees of freedom of the main plots (*e.g.,* fields or greenhouses), whereas the factors being assessed in the subplots (*e.g.,* water or fertilizer level) can be assessed with error terms incorporating a much larger number of degrees of freedom. **Cochran & Cox (1957, p. 296-297)** compare split plot and randomized blocks design with **A** being the main factor and **B** being the split-plot factor:

1) **B** and **AB** effects estimated more precisely than **A** effects in the split-plot design
2) Overall experimental error is the same between designs: increased precision on **B** and **AB** effects are at the expense of precision for tests of **A** effects,
3) The chief advantage of the split plot over the factorial is combining factors that are expensive to create (the **A** or main plot factors) with relatively inexpensive subplot factors.

Consider the use of a split plot design when **B** and **AB** effects of more interest than **A**, or if the **A** effects can not be fully replicated with small amounts of resources.

Analytical error In measurement, there is usually sampling error, and there is also analytical error. Even if a sample had a known value for a variable (sampling error is zero), some analytical methods introduce error. The **expected value** of this analytical error, if the instrument is properly calibrated, should be zero so that **precision** is affected but not **accuracy** (*c.f.*, **systematic error**). [Note added 5/15/09: I just did a web search on analytical error and found that in chemical analysis, **Total analytical error (TAE)** is defined as the sum of both the random and systematic error, so that TAE affects both **accuracy** and **precision**].

Arcsine square root transformation For some frequency data, √arcsin(x) when x ranges between 0 and 1 will sometimes expand truncated tails in a distribution. The residual vs. predicted value plot indicating the need for a transform looks football-shaped: thick in the middle thin at the tails. The **logit transform** often works better and is easier to interpret.

ARIMA autoregressive integrated moving-average *cf.*, **CAR**, **SAR**, **SARIMA**

Asymptotic relative efficiency [Pitman efficiency] “Suppose that, ..., the sample size m=n has been determined for which the Wilcoxon test will achieve a specified power ... One would then wish to know what sample size m’=n’ is required by the t-test to achieve the same power against the same alternative. The ratio n’/n is called the efficiency of the Wilcoxon test relative to the t-test ... the limiting efficiency, which turns out to be independent not only of Π [power] but also of α is called the Pitman efficiency (or asymptotic relative efficiency) of the Wilcoxon test to the t-test." **Lehmann (2006, p. 78-80)**. The asymptotic relative efficiency of the sign test is 62% relative to the *t* test and 66% relative to the Wilcoxon signed rank test. The Wilcoxon signed rank test has a 94% asymptotic relative efficiency relative to the t test. **Lehmann (2006, p. 172)**.

Axiomatic probability see **probability**

Bartlett's test A test for homoscedasticity or equality of variances, not used much now since it is sensitive to normality. **Levene's test** and graphical methods are preferred.

Bayes, Thomas (1702(?)-1761). Protestant minister who described **Bayes theorem**.

Bayes theorem. A theorem named after the English minister **Thomas Bayes**, published posthumously in 1763. The first explicit statement of the theorem is due to **Laplace**.

$$\textit{Let } \{ A_i \}_{n}^{i=1} \textit{ be a set of } n \textit{ events,}$$
$$\textit{each with positive probability,}$$
$$\textit{that partitions } S \textit{ in such a way that}$$
$$\bigcup_{i=1}^{n} A_i = S$$
$$\textit{and } A_i \cap A_j = \varnothing \textit{ for } i \neq j.$$
$$\textit{For any event } B \textit{ (also defined on } S\textit{),}$$
$$\textit{where } P(B) > 0,$$
$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$
$$\textit{for any } i \leq j \leq n.$$

Or from **Robert & Casella (1999)**:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$$
$$\textit{where } m(x) = \int f(x|\theta)\pi(\theta)d\theta \textit{ is the marginal density of } X.$$

Bayesian inference A school of statistics based on **Bayes theorem**. Every analyst has a prior belief about the probability of a given hypothesis & its alternatives. After evaluating data, these prior probabilities can be combined with the data to produce posterior probabilities. Bayesian probability estimates usually converge with p values from statistical tests used in the **frequentist school of statistics**. Bayesians argue that their methods are more general, and that Bayesian methods are more suitable for evaluating one-shot events, where long run probabilities have little meaning. *cf.,* **probability**

Bayesian information criterion (BIC) BIC statistic used to choose a parsimonious multiple **regression** equation *cf.*, **Mallow's Cp**, **Aikake information criterion**

Behrens-Fisher problem Testing the difference between means or central tendency of populations with unequal variances. *Cf.*, **Welch's t test**, **Satterthwaite approximation**, **Fligner-Policello test**

Bernoulli trial **Hogg & Tanis (1977, p. 66)** A **Bernoulli experiment** is a random experiment, the outcome of which can be classified in but one of two mutually exclusive and exhaustive ways, say success or failure … A sequence of **Bernoulli trials** occurs when a Bernoulli experiment is performed several independent times so that the probability of success remains the same from trial to trial.

Beta distribution **http://mathworld.wolfram.com/BetaDistribution.html**

Bias The difference between the **expected value** and the true value of a parameter *cf.*, **unbiased estimator**

BIC **Bayesian information criterion**

Binomial coefficient Used in the **binomial expansion** and in calculating the number of **combinations**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Binomial distribution (**Larsen & Marx 2001** **Theorem 3.3.2, p. 136)** Consider a series of n independent trials, each resulting in one of two possible outcomes, "success" or "failure." Let p=P (success occurs at any given trial) and assume that p remains constant from trial to trial. Let the variable X denote the total number of successes in the n trials. Then X is said to have a **binomial distribution** and the **binomial mass function** is

$$p_x(k) = P(X=k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

See also the **Poisson approximation to the binomial**

Binomial expansion

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} b^r a^{n-r}.$$

Binomial test

One-sample binomial test:
http://www.math.bcit.ca/faculty/david_sabo/apples/math2441/section9/singpoppropsht/singpoppropht.htm

Binomial theorem Invented by Newton

Binomial variable

Biometry

Birthday problem **http://www.math.uah.edu/stat/urn/urn7.html**

Bivariate normal distribution

Blocking Experimental design involves assigning treatments to experimental units. When groups of experimental units may be more similar than others, the experimenter often creates blocks of similar experimental units with replicates of treatments applied within each block. A common example might be the agricultural experiment in which the experimental units are agricultural plots, arrayed in space. Blocks can be created based on spatial location, and treatments allocated to plots within spatial blocks.

Bonferroni A conservative **multiple comparisons test**: test p value=Experimentwise alpha/number of tests.

Bootstrap A Monte Carlo simulation in which n samples are drawn from a finite set of samples a large number of times *cf.*, **jackknife**

Box-Cox family of transformations. **Box & Cox (1964)** developed a **maximum likelihood** method to estimate which transformation of the response variable, Y, provided the best fit to the linear model W=Xβ + ε, given that ε ~ N(0,Iσ^2). The major transformations (square root, log, inverse) can be specified by one parameter, λ , in the following transformation equation:

$$W = \begin{cases} \dfrac{(Y^{\lambda} - 1)}{\lambda}, & for\ \lambda \neq 0, \\ ln(Y), & for\ \lambda = 0. \end{cases}$$

To perform the Box-Cox transformation, values of λ are chosen in the range -1 to 1 and the value of the **likelihood function** is plotted vs. lambda. The maximum likelihood estimate of lambda is found. Following **Draper & Smith (1998)**, an approximate 100 (1 - α)% confidence interval for λ which satisfy the inequality:

$$L_{\max}(\hat{\lambda}) - L_{\max}(\lambda) \leq \frac{1}{2}\,\chi^2{}_1\ (1 - \alpha).$$

where $\chi^2{}_1\ (1 - \alpha)$ is the percentage point of the chi-squared distribution with 1 **df** (3.84 for the 95% CI).

One half this value can be used graphically in a plot of $L_{\max}(\hat{\lambda})$ to find the upper and lower 95% confidence intervals for lambda, as shown in Figure 1.

Box's M A test of homogeneity of variance-covariance matrices.

Boxplot Invented by Tukey and displaying an approximate interquartile range, **median**, range and extreme data points. A box marks the the **interquartile range (IQR)** with lower and upper limits approximately equal to the 1st and 3rd quartiles. Tukey didn't define the boxes in terms of quartiles, but used the term **hinges**, to eliminate ambiguity. There are a number of different ways of defining the 1st and 3rd quartiles, which mark the 25th and 75th % of the cumulative frequency distribution. Hinges are simply the medians of the lower and upper half of the data points. Whiskers extend to the adjacent values, which are actual data outside the IQR but within 1.5 IQR's from the median. Points more than 1.5 IQR's from the IQR are outliers. Points more than 3 IQR's from the box are extreme outliers. See also **http://mathworld.wolfram.com/Box-and-WhiskerPlot.html**

Brown-Forsythe test A test for equal variance uses test using an ANOVA on the absolute deviation from group medians (Available in SPSS Oneway). *Cf.*, **Levene's test**.

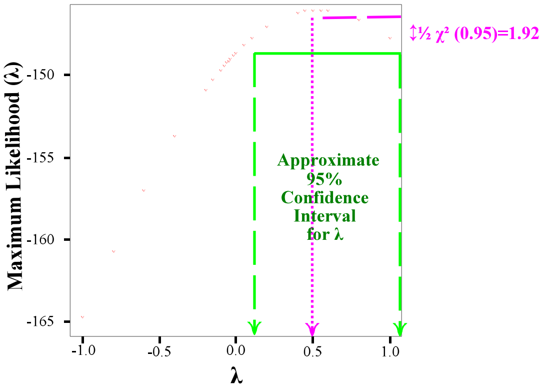
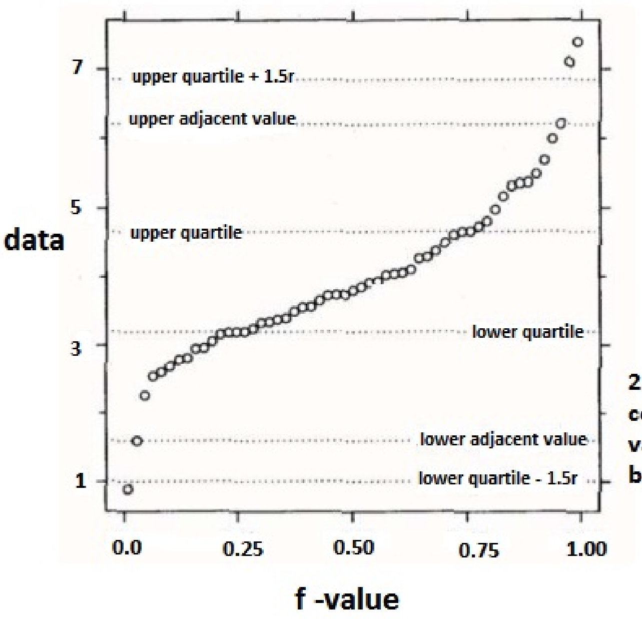


Figure 1.A graphical display showing how to identify the the Box-Cox transformation parameter, λ , with data from Draper & Smith. $L_{\max}(\hat{\lambda})$ is plotted vs λ. A horizontal line is drawn 1.92 units below the maximum likelihood value to find the lower and upper confidence limits for λ (0.01 and 1.05 here). λ =0.5 indicates that a √Y transform is appropriate, but the 95% CI includes λ = 1, indicating no transformation of Y. The 95% CI does not include λ = 0, indicating the ln transform is not appropriate. This analysis was performed with an SPSS macro on benthic infauna data from MA Bay fit to an equal means general linear model.

Buffon's needle A problem in geometric probability (**http://www.mste.uiuc.edu/reese/buffon/buffon.html**)



2.7 The diagram illustrates the computation of the adjacent values, which are used in the box plot display method.

Canonical correlation analysis

Canonical correspondence analysis cf., **redundancy analysis**

Capture-recapture experiment

CAR Conditional autoregression model *cf.*, **SAR**

Cauchy distribution

Causation

Census *cf.*, **quota sampling**, **survey design**

Central Limit Theorem Discovered by **Laplace (1811)** [see **Stigler 1986, p. 146**] See this brief synopsis: **http://mathworld.wolfram.com/CentralLimitTheorem.html**

Chain Suppose G = (V, E) is a graph: A **chain** in G is a sequence $u_1,e_1,u_2,e_2,\ldots,u_t,e_t,u_{t+1}$, where $t \geq 0$, so that each u_i is a member of V and each e_i is a member of E and e_i is always the edge $\{u_i,u_{i+1}\}$. The chain is usually written $u_1,u_2,\ldots,u_t,u_{t+1}$.

Change score analysis As **Campbell & Kenny (1999)** discuss, there are several ways to measure the effect of an intervention, say a change in test scores as the result of a change in teaching method: 1) The outcomes can be compared directly, 2) Change score analysis in which the pretest is subtracted from the post test, 3) Regressing the post test score on the pre test score (this can create an artifact).

Chao1 A diversity index to estimate species richness. Reviewed by **Hughes et al. (2001)** and **Colwell & Coddington (1994)** *Cf.*, ACE

$$\begin{aligned} Chao_1 &= S_{obs} \times \frac{n_1^2}{n_2}, \quad where \\ S_{obs} &= Observed\ species. \\ n_1 &= Species\ observed\ only\ once. \\ n_2 &= Species\ observed\ twice. \end{aligned}$$

Chebyshev's inequality (**Hogg & Tanis 1997**) If the random variable X has a finite mean μ and finite variance σ^2, then for every $k \geq 1$,

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Chi square distribution

Chi-squared statistic, invented by **Pearson** in 1900 (**Stigler 1986, p. 348**)

Climate field reconstuction CFR Approach to reconstructing a target large-scale climate field from predictors employing multivariate regression methods. CFR methods have been applied both to filling spatial gaps in early instrumental climate data sets, and to the problem of reconstructing past climate patterns from 'climate proxy' data. **http://www.realclimate.org/index.php?p=29**

Cluster effect Replicate samples are not independent due to samples being collected in subgroups such as pigs in a litter (**Ramsey & Schafer 2002 p. 62**)

Cluster sampling

Multistage cluster sampling

Coefficient of determination R^2 See **R squared**

Coefficient of multiple determination the amount of variation in a response variable explained by a regression with more than one explanatory variable

Coefficient of variation The standard deviation, *s*, divided by the mean.

Collinearity see **multicollinearity**

Combinations The number of combinations of n objects taken r at a time is

$$\begin{aligned} C(n,r) &= \binom{n}{r}. \\ &= \frac{n!}{r!(n-r)!}. \end{aligned}$$

Combinatorics

Complement Let *A* be any event defined on a **sample space** *S*. The complement of *A*, written A^c or A’, is the event consisting of all the outcomes of *S* other than those contained in *A*. (**Larsen & Marx 2001, Definition 2.2.10**) **Concordant**

Conditional independence

Conditional probability The symbol P(A|B) — read “the probability of A given B”--- is used to denote a **conditional probability**. Specifically (P|A) refers to the probability that *A* will occur given that *B* has already occurred.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Confidence interval **Kendall & Stuart (1979, p. 199)** state that the ideas of confidence interval estimation are due to Neyman, especially Neyman (1937).

Confidence limits

for a proportion

$$P\left(-z_{\alpha/2} < \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}\ (1-\hat{p})}{n}}} < z_{\alpha/2} \right) \doteq 1-\alpha. \qquad \textbf{(19)}$$

Confounding variables A **variable** related both to group membership and to the outcome. Its presence makes it hard to establish the outcome as being a direct consequence of group membership.” **Ramsey & Schafer 1997**. A confounding variable has no relation to the response, but an **effect modifier** does.

Consistent estimator see **estimators**

Contingency table

Cook’s D A diagnostic statistic for outliers that matter in regression. Essentially, the change in regression parameters resulting from the deletion of individual cases.

Corner test

Correlation Introduced by **Galton (1888)** (**Stigler, 1986, p. 297**) The correlation is a standardized form of covariance obtained by dividing the covariance of two variables by the product of the standard deviations of x and y. [*cf.*, **Pearson's r**, Spearman's ρ, **Kendall's τ**]

biserial correlation coefficient the correlation between an artificial dichotomy (made by imposing a cut-point on a "continuous" variable) and a "continuous" variable [Burrill on sci.stat.edu]

part correlation From SPSS regression user's guide. The correlation between the dependent variable and an independent variable when the linear effects of the other independent variables in the model have been removed from the independent variable. It is related to the change in R-squared when a variable is added to an equation. Sometimes called the semipartial correlation.

partial correlation From SPSS regression user's guide. The correlation that remains between two variables after removing the correlation that is due to their mutual association with the other variables. The correlation between the dependent variable and an independent variable when the linear effects of the other independent variables in the model have been removed from both.

point biserial correlation the correlation between a dichotomy and a quasi-continuous variable [Burrill], or "The product-moment correlation between a dichotomous correlation and a continuous (scale) variable." **Cohen et al. (2003)**

polychoric correlation "This measure of association is based on the assumption that the ordered, categorical variables of the frequency table have an underlying bivariate normal distribution. For 2 ×2 tables, the polychoric correlation is also known as the **tetrachoric correlation**. ...the polychoric correlation coefficient is the maximum likelihood estimate of the product-moment correlation between the normal variables, estimating thresholds from the observed table frequencies. The range of the polychoric correlation is from -1 to 1." **http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap28/sect20.htm** The tetrachoric correlation is a special case of the polychoric. **http://ourworld.compuserve.com/homepages/jsuebersax/tetra.htm**

Phi coefficient: correlation between two dichotomies

tetrachoric correlation coefficient Used when both variables are dichotomies which are assumed to represent underlying bivariate normal distributions **http://www2.chass.ncsu.edu/garson/pa765/correl.htm#tetrachoric** and **http://ourworld.compuserve.com/homepages/jsuebersax/tetra.htm**

Correspondence analysis, also known are reciprocal averaging. A form of **principal components analysis** designed to partition and display the variation of a chi-square metric. There are at least 5 different ways of scaling the displays (see **Greenacre 1984**, **Legendre & Gallagher 2001**[especially notes to **Gallagher's Matlab programs** that accompany the paper])

Countably infinite (Larsen & Marx 2001, p 37 footnote). A set of outcomes is **countably infinite** if it can be put in one-to-one correspondence with the positive integers.

Covariance a measure of association between two variables; covariance is the mean of the cross products of the **centered data**. It can also be defined as the expected value of the sum of cross products between two variables expressed as deviations from their respective mean. The covariance between **z-transformed** variables is also known as the **correlation**.

Cox proportional hazard model See **Cox regression**

Cox regression "Cox regression offers the possibility of a multivariate comparison of **hazard rates** (**Hazard ratios**). However, this procedure does not estimate a "baseline rate"; it only provides information whether this 'unknown' rate is influenced in a positive or a negative way by the independent variable(s) (or covariates)."
http://www.lrz-muenchen.de/~wlm/wlmscox.htm
From the SPSS help file: "Like Life Tables and **Kaplan-Meier survival analysis**, Cox Regression is a method for modeling time-to-event data in the presence of censored cases. However, Cox Regression allows you to include predictor variables (covariates) in your models. For example, you could construct a model of length of employment based on educational level and job category. **Cox Regression** will handle the censored cases correctly, and it will provide estimated coefficients for each of the covariates, allowing you to assess the impact of multiple covariates in the same model. You can also use **Cox Regression** to examine the effect of continuous covariates." See also:
http://www.statsoft.com/textbook/stsurvan.html

Craps The most popular game played only with dice. The shooter makes a bet, called the center bet, and other players 'fade' the bet or bet against the shooter. The shooter rolls a pair of dice. If the sum of the dice is 7 or 11, called a *natural*, the shooter wins immediately, if 2, 3, or 12 is rolled, called *craps*, the shooter immediately loses. If the shooter rolls a 4, 5, 6, 8, 9, or 10, that number becomes his point. He rolls the dice again until he shoots the same number again, called making one's point or he rolls a seven or *craps out* or *sevens out*. There are dozens of side bets that can be made on the eventual outcome or the outcome of a single roll. In a casino, all bets are against the house with bets being placed on a craps table (See Fig. 3) *c.f.* **odds**

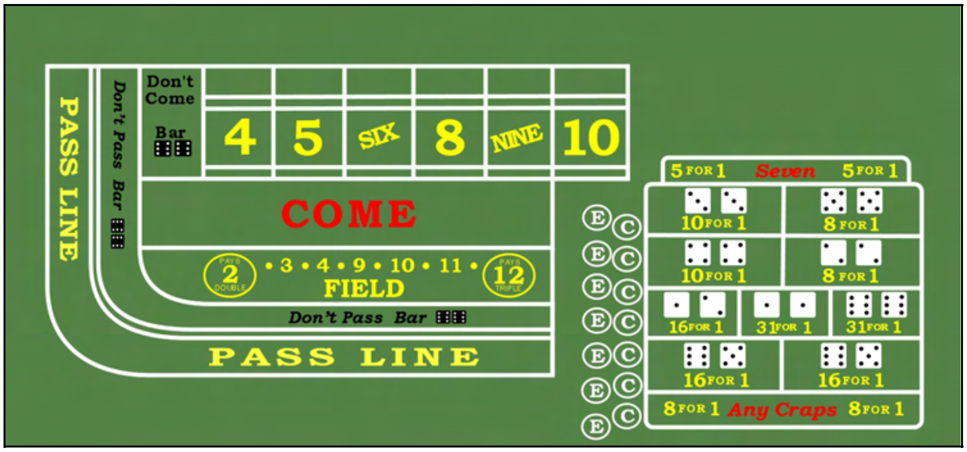


Figure 3. Craps table with odds, from Wikipedia. Note that odds for rolling seven on the next roll, 5 for 1, equals 4-to-1 odds.

Critical region **Hogg & Tanis (1977, p. 255)**. The critical region C is the set of points in the sample space which leads to the rejection of the **null hypothesis** H_o. The rejection region for a null hypothesis is called the critical region and the cutoff is called the **critical value**. These concepts are associated with the **Neyman-Pearson school** of statistical inference *cf.*, test statistic

Crossover design Each subject receives more than one treatment level, and the order of the treatment is usually randomly assigned. Crossover designs can be analyzed with either univariate or multivariate repeated measures analyses with treatment order as a between subject factor. **Cochran & Cox (1957, p. 127-142)** describe modifications of Latin Square analysis appropriate for several different types of crossover design (using cow milk production as the response and diet as the treatment factor). **Neter et al. (1996, p. 1225)** refer to crossover designs as latin square changeover designs, 'often useful when a latin square is to be used in a repeated measures study to balance the order positions of treatments, yet more subjects are required than called for by a single lain square.' **Neter et al. (1996)** provide the model and expected mean square ANOVA table.

> ... a relatively simple model can be developed ... ρ_i denotes the effect of the *i*th treatment order pattern, κ_j denotes the effect of the *j*th order position, τ_k denotes the effect of the *k*th treatment, and $\eta_{m(I)}$ denotes the effect of subject *m* which is nested within the *i*th treatment order pattern:

$$Y_{ijkm} = \mu_{...} + \rho_i + \kappa_j + \tau_k + \eta_{m(i)} + \varepsilon_{ijkm}$$

where:

$\mu_{...}$ *is a constant*

ρ_i, κ_j, *and* τ_k *are subject to the restrictions* $\sum \rho_i = \sum \kappa_j = \sum \tau_k = 0$

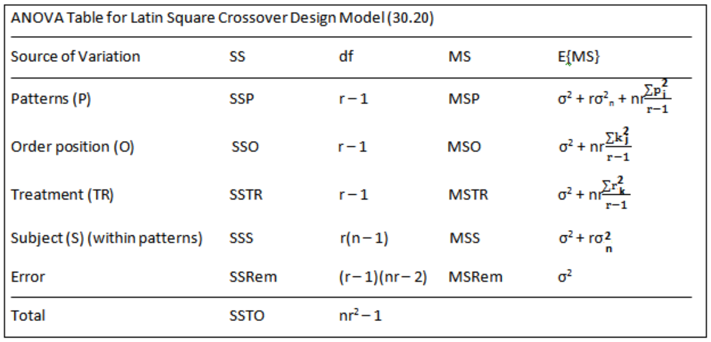
$\eta_{m(i)}$ *are independent* $N(0,\sigma^2_\eta)$

ε_{ijkm} *are independent* $N(0,\sigma^2)$ *and independent of the* $\eta_{m(i)}$

$i = 1,\ldots,r;\ j = 1,\ldots,r;\ k = 1,\ldots,r;\ m = 1,\ldots,n$

Neter et. al (1996, p. 1226) show how to test a crossover design with an ANOVA model to distinguish treatment, order and pattern effects (see Figure 4).

Dallal (**http://www.tufts.edu/~gdallal/crossovr.htm**) reviews the strengths and limitations of crossover designs. The increased precision of crossover design, due to each subject serving as its own control, is vitiated by the need to have subjects participate a longer period of time and the need to account for carryover effects. **Mead (1988, p 197)**, like **Neter et al. (1996)** describes crossover experiments as a form



ANOVA Table for Latin Square Crossover Design Model (30.20)

Source of Variation	SS	df	MS	E{MS}
Patterns (P)	SSP	r − 1	MSP	$\sigma^2 + r\sigma^2_n + nr\frac{\sum p_i^2}{r-1}$
Order position (O)	SSO	r − 1	MSO	$\sigma^2 + nr\frac{\sum k_j^2}{r-1}$
Treatment (TR)	SSTR	r − 1	MSTR	$\sigma^2 + nr\frac{\sum r_k^2}{r-1}$
Subject (S) (within patterns)	SSS	r(n − 1)	MSS	$\sigma^2 + r\sigma^2_n$
Error	SSRem	(r − 1)(nr − 2)	MSRem	σ^2
Total	SSTO	nr² − 1		

Figure 4. showing the ANOVA mean squares corresponding to the model shown in the previous equation.

of **Latin square** with time as a **blocking factor**. Crossover designs allow a tremendous gain in precision by reducing the effects of among patient variability while also requiring fewer subjects than completely randomized designs to attain similar **relative power efficiency**. **Mead (1988, p 198)** discusses the difficulty in relevance,

> *The difficulty with the cross-over design is that the conclusions are appropirate to unis similar to those in the experiment; that is , to subjects for a short time period in the context of a sequence of different treatments. We have to ask if the observed difference between two treatments would be expected to be the same if a treatment is applied consistently to each subject to which it is allocated. This is a problem of interpretation of results from experiment to subsequent use, and it is a problem which must be considered in all experiments. It is particularly acute in cross-over designs, because the experiment is so different from subsequent use. After all, no farmer is going to continually swap the diets for his cattle!*

Data mining Looking for pattern in gigantic datasets

Degrees of freedom The number of true replicates minus the number of model parameters that must be estimated from the data. **Stigler (1986, p. 348)** states that the term degrees of freedom was. not formally introduced until 1922 when **Fisher** introduced the term. Here is a pdf of Walker (1940), describing the history and geometric interpretation: http://courses.ncssm.edu/math/Stat_Inst/PDFS/DFWalker.pdf

DeMorgan's Laws (**Larsen & Marx 2001, p. 29**) Let A and B be any two events. The complement of their intersection is the union of their complements:

$$(A \cap B)^c = A^c \cup B^c.$$

the complement of their union is the intersection of their complements:

$$(A \cup B)^c = A^c \cap B^c.$$

Deviance Calculated from the log likelihood statistic in genarlized linear models. Change in deviance can be used to test the goodness of fit of a **generalized linear model** (with the chi-square distribution) and the change in deviance permits a test between full & reduced hierarchical generalized linear models. **Agresti (1996, p 96)**: Let L_M denate the maximized log-likelihood value for the model of interest. Let L_S denote the maximized log-likelihood value for the moxt complex midel, which has a separate parameter at each explanatory setting: that model is said to be saturated. The deviance of a model is defined to be: Deviance = $-2(L_M - L_S)$.

DFFITS

Discriminant analysis

Disjoint

Distributions

- **beta**
- **binomial**
- **bivariate normal**
- **Cauchy**
- **chi-square**

empirical
exponential
F
gamma
geometric
Gompertz
hypergeometric
lognormal
multinomial
negative binomial
normal
Poisson
posterior
Student's t
Weibull

Doubly multivariate designs A form of **profile analysis** in which several different response variables are measured at several different times (**Tabachnick & Fidell 2001, p 423**)

Duncan's test A multiple comparisons test

Dummy variables Also called **indicator variables**. Variables made up of zeros and ones. Dummy variables play a key role in ANOVA analysis using regression. A discrete (or categorical) variable with 8 levels can be coded for with 8 dummy variables. In least squares regression, one of these dummy variables is left out of the regression equation and becomes the **reference level**. There are two common ways to code dummy variables for regression, the first using 0's and 1's and the second using 0's, 1's and -1's. The former approach is the most common.

Dunn's test A **multiple comparison procedure** [MCP] "The Dunn multiple comparison procedure is based on the use of the ***t*** **distribution** with C comparisons that are planned. Not only do you know the *number* of comparisons before the research is done, you also know *which* comparisons will be computed." **Toothaker (1993, p. 31)**

Dunnet's t test *A posteriori* comparison of control vs. treatments.

Durbin-Watson test A test for **serial correlation**

e (mathematical constant) **http://www.answers.com/topic/e-mathematical-constant**, *cf.*, **natural logarithms**

Ecological fallacy [**Ecological inference problem]** Error in predicting individual behavior from aggregate data. Introduced by **Robinson (1950)** and perhaps solved by **King (1997)**. King describes the problem as often involving trying to estimate the cell frequencies of an r x c contingency table, knowing only the marginal totals. The problem *cf.*, **Simpson's paradox**

Edge In a **graph**, the line connecting two **vertices**. It can be represented as an unordered pair of vertices {u,v}. If G is the matrix representation of the graph, there is an edge connecting two vertices u and v if the G_{uv} and G_{vu} elements are 1.

$E(S_n)$ (**Sanders 1968**, **Hurlbert 1971**) Hurlbert-Sanders expected number of species $E(S_n)$. Hurlbert, using formulae for the **hypergeometric probability distribution**, corrected the algorithm described by Sanders for estimating the number of species found in a random subsample of size n from a sample.

$$E(S_n) = \sum_{k=1}^{S} 1 - \frac{\binom{N-N_k}{n}}{\binom{N}{n}}.$$

$$\begin{aligned} \textit{where,} \quad n &= \textit{random sample size.} \\ \binom{N}{n} &= \textit{binomial coefficient.} \\ &= \textit{No. of ways to sample } N \textit{ objects, } n \textit{ at a time.} \\ &= \frac{N!}{(N-n)! * n!} \\ N &= \textit{Total individuals in sample.} \\ N_k &= \textit{Individuals of species } k. \\ S &= \textit{Number of species.} \end{aligned}$$

Effect modication A factor, Z, is said to be an **effect modifier** of a relationship between a risk factor, X, and an outcome measure, Y, if the strength of the relationship between the risk factor, X, and the outcome, Y, varies among the levels of Z. A factor, Z, is said to **confound** a relationship between a risk factor, X, and an outcome, Y, if it is not an effect modifier and the unadjusted strength of the relationship between X and Y differs from the common strength of the relationship between X and Y for each level of Z. More complicated definitions allow for a factor to be both an effect modifier and a counfounder. If Z is an effect modifier, then it is important to report the strength of the X-Y relationship for specific values of Z. If the strength of the X-Y relationship does not vary greatly among the levels of Z, it may not be important to account for the effect modification. If Z is a confounder, then it is common to report both the strength of the unadjusted X-Y relationship and the strength of the adjusted X-Y relationship. If the adjusted and unadjusted strengths do not differ greatly, then it may not be important to report both.... "Effect modification: ... The effect of High dose cyclosporin (cs) on transplant failure is modified by type of transplant." "Confounding: The effect of treatment on patient survival is confounded by age."

http://www-personal.umich.edu/~bobwolfe/560/review/kkm13confoundeffectmodify.txt

Cf., **mediation**

Efficient estimator see **Estimators**

EM algorithm expectation-maximization (EM) algorithm, *cf.*, **maximum likelihood**

http://www.mathdaily.com/lessons/Expectation-maximization_algorithm

EMAP EPA's Environmental monitoring and assessment program

Empirical distribution function **Hogg & Tanis (1977, p86)** Let $x_1, x_2, \dots , x_n$ denote the observed values of the random sample $X_1, X_2, \dots , X_n$ from a distribution. Let $N(\{x_i:\ x_i \le x\})$ equal the number of these observed values that are less than or equal to x. Then the function

$$F_n(x) = \frac{N(x_i:\ x_i \le x)}{n}.$$

defined for each real number x, is called the **empirical distribution function**.

Empirical orthogonal function analysis (EOF) A modification of **principal components analysis** that is widely used in physical oceanography & meteorology. **Ramsey & Schafer (2002 p. 519-520)** provide a too-brief description. Spatial pattern tied to a particular mode of time/space variance in a spatiotemporal data set (see also **Principal Components Analysis**).**http://www.realclimate.org/index.php?p=25**

Empirical rule Larsen & Marx

Ergodic Markov chain A Markov chain is called **ergodic** if its transition digraph is **strongly connected** (*i.e.*, every state can reach every other state). The chain is a **regular ergodic Markov chain** if there is a number k such that every state can reach every other state in exactly k steps (**Roberts 1976, 289-290**). A Markov chain is **regular** if and only if it is possible to be in any state after some number N of steps, no matter what the starting state, That is, if and only if P^N has no zero entries for some N (**Kemeny & Snell, 1976**). **Gondran & Minoux (1984, p. 20)** provide a graph-theoretic definition. Each Markov chain can be associated with a **transition graph** which consists of N vertices corresponding to the states, two vertices i and j being linked by an **arc** (i,j) if and only if $P_{ij}>0$. If the transition graph is connected and not periodic (*i.e.*, the largest common factor of the lengths of all the circuits passing through a vertex equals 1). If G_R is the **reduced graph** and a **strong component** has outdegree greater than zero, then that component is a transient subset of the graph. If the outdegree of a strong component is 0, then that strong component is a recurrent (ergodic) subset. If the graph is not periodic and contains only 1 recurrent (ergodic) class (*i.e.*, the **reduced graph** is **strongly connected**), then the system is completely ergodic.

Error

Estimate A statistic used as a guess for the value of a parameter. Estimates can be calculated, but parameters remain unknown (**Ramsey & Schafer 1997, p. 20**)

Estimators This section is from **Harman (1976)**.

consistent estimator An estimator $\hat{\theta}$ is said to be **consistent** if it converges (in a probabilistic sense) to the true parameter as the sample increases without limit, *i.e.*, $limit_{N\to\infty}\ \hat{\theta} \Rightarrow \theta$.

Hogg & Tanis (1977 Definition 7.5-2) The statistic Y =*u* (X_1, X_2, …, X_n) is a **consistent estimator** of θ if, for each positive number ε,

$$\lim_{n\to\infty} P(|Y-\theta| \geq \varepsilon) = 0.$$

$$or,\ equivalently,$$

$$\lim_{n\to\infty} P(|Y-\theta| \leq \varepsilon) = 1.$$

efficient estimator An estimator is said to be **efficient** if it has the smallest limiting variance. When an estimator is efficient it is also consistent.

minimum variance unbiased estimator Given a choice between two unbiased estimators, the one with minimum variance is preferred. For example, **Draper & Smith (1998)** note that while **OLS** and **WLS** regression provide **unbiased estimators** of the regression parameters, the variance of the WLS estimators will be lower if the variance of the regression areas are **heteroscedastic**.

sufficient estimator An estimator is said to be **sufficient** if it utilizes all the information in the sample concerning the parameter.

unbiased estimator If the expected value of the estimator is the true parameter, *i.e.*, $E(\hat{\theta}) = \theta$, then the estimator is unbiased.

*"While it is of some advantage to devise an **unbiased estimate**, it is not a very critical requirement. The method of **maximum likelihood** is a well established and popular statistical method for estimating the unknown population parameters because such estimators satisfy the first three of the above standards. Not all parameters have sufficient estimators, but if one exists the maximum likelihood estimator is such a **sufficient estimator** (Mood and Graybill 1963, p. 185). However, a **maximum-likelihood estimator** will generally not be unbiased. (By getting the expected value of such an estimator, an unbiased statistic can be derived). This method yields values of the estimators which maximize the likelihood function of a sample."* **Harman (1967, p. 212-213)**

Expected value **Hogg & Tanis (1977, p 53)** *If f(x) is the **probability density function** of the random variable X of the discrete type with space R and if the summation*

$$\sum_R u(x)f(x) \ = \ \sum_{x\in R} u(x)f(x).$$

*exists, then the sum is called the mathematical expectation or the **expected value** of the function u(x), and it is denoted by E[u(X)]. That is,*

$$E[u(X)] \ = \ \sum_R u(x)f(x).$$

Theorem When it exists, mathematical expectation E satisfies the following properties: i) If c is a constant E(c)=c, ii) If c is a constant and u is a function, E[cu(X)]=cE[u(X)], iii) If c_1 and c_2 are constants and u_1 and u_2 are functions, then $E[c_1u_1(X)+c_2u_2(X) = c_1E[u_1(X)] + c_2E[u_2(X)]$, and

$$iii') \; E\left[\sum_{i=1}^{k} c_i u_i(X)\right] = \sum_{i=1}^{k} c_i E[u_i(X)].$$

Experiment The fundamental difference between a **survey** (**observational study**, **census**) and an experiment is that the sampling units in an experiment can be regarded as being drawn from an infinite population:

"The distinction between the design of experiments and the design of ***sample surveys*** *is fairly clear-cut, and may be expressed by saying that in* ***surveys*** *we make observations on a sample taken from a finite population of individuals, whereas in* ***experiments*** *we make observations which are in principle generated by a hypothetical infinite population, in exactly the same way that the tosses of a coin are. Of course, we may sometimes experiment on the members of a sample resulting from a survey, or even make a sample survey of the results of an (extensive) experiment, but the essential distinction between the two fields should be clear."* **Kendall & Stuart 1979**

"By experiment we will mean any procedure that (1) can be repeated, theoretically, an infinite number of times; and (2) has a well-defined set of possible outcomes" (**Larsen & Marx 2001 p. 21**)

"A cornerstone of the scientific process is the **experiment**. Ecologists in particular use a wide variety of types of experiments. We use the term "**experiment**" here in its broadest sense: a test of an idea. Ecological experiments can be classified into three broad types: manipulative, natural, and observational. Manipulative, or controlled, experiments are what most of us think of as experiments: A person manipulates the world in some way and looks for a pattern in the response..... Natural experiments are "manipulations" caused by some natural occurrence..... Observational experiments consist of the systematic study of natural variation." Gurevitch, J., S. M. Scheiner and G. A. Fox. 2002. The Ecology of Plants. Sinauer Associates, Sunderland, Massachusetts.

Experimental design *cf.*, **orthogonal arrays**

Experimentwise error (or **experiment-wise** or **family-wise error**) The error associated with rejecting one or more true null hypotheses in an experiment. If alpha [α] is the probability of **Type I error** for a single test and n tests are performed on the results of the experiment, then $Experimentwise \; \alpha = 1 - (1 - \alpha_{test})^n$. For example, the experimentwise error level if each of 10 independent tests is performed at alpha = 0.05 is 40.1%. Various **multiple comparisons tests** have been designed, some of which control for experimentwise alpha level. **Family-wise error rate**.

Explanatory variable A variable used to predict the value of a response variable, usually in a **regression model**. Sometimes called an **independent variable**, but this is a poor term, since these variables are rarely independent of the response variable or other explanatory variables. *cf.*, **response variable**

Exponential distribution Waiting times for a process that has Poisson distributed rates **http://mathworld.wolfram.com/ExponentialDistribution.html**

Extra sum of squares principle **http://www.tufts.edu/~gdallal/extra.htm**

False positive See **sensitivity**

F distribution named by **George Snedecor** in honor of **R. A. Fisher**

F-test A ratio of variances or mean squares with expected value of unity, tested with the F distribution given numerator and denominator degrees of freedom.

Factor Analysis (FA) A term coined by **Spearman (1904)**. The goal of **PCA** is to account for as much variance in the data as possible, whereas the goal of FA is to account for the covariance between descriptors (variables). Factor analysis assumes that the observed descriptors are linear combinations of hypothetical underlying variables (or factors). Factor analysis can be divided into two types: **Exploratory factor analysis** and **Confirmatory factor analysis**. Factor Analysis is primarily directed towards the analysis of covariation among descriptors, so that with most models, the relative positions of samples can not be readily determined, but methods do exist to estimate the orientation of samples (termed **factor scales** or **factor scores**) in factor space. In **confirmatory factor analysis**, specific expectations about the number of factors and their loadings can be tested.

Confirmatory Factor Analysis **factor analysis** in which specific expectations regarding the number of factors and their loadings are tested on sample data. (**Kim & Mueller (1978)**, **Legendre & Legendre (1998)**)

Exploratory Factor Analysis no *a priori* specification of the number of factors or **loadings**

Factor score **R mode**: the estimate for a case on an underlying factor formed by a linear combination of observed variables. **Q mode**: the estimate for a variable on an underlying factor formed by a linear combination of observed cases.

Factor loadings The elements of the eigenvectors are also the weights or **loadings** of the various original descriptors. If the eigenvectors have been normalized to unit length (*i.e.*, the sum of the squared loadings for a variable across factors equals 1.0), then the elements of the eigenvector matrix (the loadings) are direction cosines of the angles between the original descriptors and the principal axes. So that if the element of the U vector (the **loading** for a variable) is .8944, the angle is $\cos^{-1}$ (.8944)=arc cos(.8944)=26° (**Legendre and Legendre 1983**). The principal component axis is rotated 26° from the original axis. For this reason, the factor loadings are sometimes called **directional cosines**.

oblique factor rotation In **orthogonal** rotations the causal underlying factors are not permitted to be correlated, while in oblique rotations the factors can be rotated. The geometric relationships among variables in ordination 2-space is greatly altered with oblique rotations. For example, one can no longer assume that descriptors plotted at right angles relative to the origin are uncorrelated. The main virtue of oblique rotations is in naming axes or factors and using factors as explanations rather than as descriptions. **Legendre & Legendre (1983; Fig 8.13; p. 308)** describe the relationship between oblique factor rotations and path analysis. Oblique rotations may be needed when common factors are correlated.

orthogonal factors factors that are not correlated with each other.

Factorial n! is pronounced “n factorial”

$$\begin{aligned} n! &= n\times(n-1)\times\ldots 2\times 1.\\ 0! &= 1.\\ n! &= \Gamma(n+1) \end{aligned}$$

In calculations with large n, the natural log of the **gamma distribution (Γ)** is usually used: nfactorial=exp(gammaln(n+1))

Family-wise error The error associated with rejecting one or more true null hypotheses in an observation study or experiment. If alpha (α_{test}) is the probability of **Type I error** for a single test and n independent tests are performed on the results of the experiment, then $Familywise\ \alpha = 1-(1-\alpha_{test})^n$. For example, the experimentwise error level if each of 10 independent tests is performed at alpha = 0.05 is 40.1%. Various **multiple comparisons tests** have been designed, some of which control for family-wise or experimentwise alpha level, synonymous with **Experimentwise error rate**.

Fixed effects *cf.*, **random effects**

Fixed point probability vector (= **stationary vector**, **Limiting vector**) The **left eigenvector** (if the transition matrix is in “from rows to columns form”) associated with the **dominant eigenvalue** of an **ergodic Markov chain** process. The dominant eigenvalue is 1.0 for ergodic Markov chains.

Pierre de Fermat (1601-1665) With **Pascal**, the father of the mathematical theory of **probability** (**Bell 1937, p. 87**).

Fisher, Sir Ronald A., discoverer of maximum likelihood, discriminant analysis (with Burt), and **ANOVA**. His Genetics of Natural Populations laid the foundation for quantitative population genetics. His statistics for experimenters laid the foundation for experimental design. Fisher is one of the fathers of the frequentist school of statistics, the others being Jerzy Neyman and Egon Pearson. Fisher introduced many of the test statistics and advocated the use of p values in judging the significance of results. Neyman & Pearson introduced critical values and confidence limits. The frequentist philosophy of statistics differs from the **Bayesian** philosophy of statistics.

Figure 5. RA Fisher from http://www.thefullwiki.org/R._A._Fisher

Fisher’s exact test A test for 2x2 tables, designed for hypergeometric distributions, but widely applicable to other 2x2 problems.

Fisher’s sign test A distribution-free test for paired data, analogous to the paired t test. The number of positive (or negative) differences is compared to expectations from the binomial distribution.

Fligner-Policello test A rank-based test for differences in central tendency for two independent samples with unequal variances. Note, that the **Wilcoxon rank sum test** assumes equal variances. On Matlab Central file exchange, Trujillo-Ortiz et al. have posted fptest.m, which implements the **Fligner & Policello (1981)** test, which is described in **Hollander & Wolfe (1999)**. *Cf.*, **Behrens-Fisher problem**, **Wilcoxon rank-sum test**

Forward selection One of many automatic selection procedures in multivariate regression. The explanatory variable with the highest correlation with the response variable is entered in the equation first, and the explanatory variable with the highest partial correlation with the response variable is entered next, and so on.

Friedman's test A non-parametric 2-way ANOVA, specifically designed for repeated measures problems. *cf.*, **Kruskal-Wallis ANOVA**

Frequentist theory of statistical inference This is the traditional model of statistical inference, developed in the 20th century by **RA Fisher** and **Neyman & Pearson**. Statistical tests are performed with an assumed probability model. The p value is the probability that an observed even or one more extreme would have been observed if the assumed probability model and associated null hypothesis was true. Neyman & Pearson introduced the use of critical values and confidence limits for describing the results of statistical tests *cf.*, **Bayesian inference**

Fundamental matrix For an **absorbing Markov chain**, N is the fundamental matrix and is found as the inverse of the identity matrix minus the transitions among the non-absorbing states (the Q submatrix): $N = (I\text{-}Q)^{-1}$. See **Kemeny & Snell 1976**

Galton, Francis (1822-1911) Darwin's brother in law who coined the term **regression** in the context of describing **regression to the mean** **http://en.wikipedia.org/wiki/Francis_Galton**

Gamma A measure of association "The estimator of gamma is based only on the number of concordant and discordant pairs of observations. It ignores tied pairs (that is, pairs of observations that have equal values of X or equal values of Y). Gamma is appropriate only when both variables lie on an ordinal scale. It has the range . If the two variables are independent, then the estimator of gamma tends to be close to zero." **http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap28/sect20.htm**

Gamma distribution

Gamma function

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y}\, dy, \; 0 < 2.$$

GAMS Generalized additive models

Gauss, Carl Friedrich Arguably the most influential mathematician of all time, but Stigler argues that he should not be given credit for the normal curve, which is sometimes called the Gaussian curve *cf.*, **Stigler's law of eponymy**

Gaussian curve see **normal curve**

Generalized additive models (GAMS) According to **Leathwick & Austin (2001, p 2562)**, GAMS are an extension of **generalized linear models** ' which offer a more realistic approach to the analysis of ecological data in that complex relationships between preditor and response variables can be accommodated in a nonparametric manner through use of scatter-plot smoothers, rather than using more inflexible parametric terms as in GLM's.'

Generalized least squares (GLS) The general linear model is

$$\mathbf{y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}.$$

where **y** is an *n* x 1 matrix of observations, **β** is a *p* x 1 matrix of regression parameters, **X** is an *n* x *p* matrix of explanatory variables, and ε is an *n* x 1 vector of residuals. General least squares modeling assumes that the errors are independently, identically, normally distributed ($\boldsymbol{\varepsilon} \sim N(0,\sigma^2 \mathbf{I})$), leading to the normal equation solution for **β**

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X'X})^{-1}\mathbf{X'}\,\mathbf{y}.$$

Generalized least squares permits a broader array of variance-covariance matrices for the error ($\boldsymbol{\varepsilon} \sim N(0, \Sigma)$), where Σ is a symmetric, positive-definite variance-covariance matrix.

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = \left(\mathbf{X'}\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X'}\Sigma^{-1}\mathbf{y}.$$

One simple form of generalized least squares analysis is **weighted least squares regression**.

General linear model Regression, especially regression using indicator or 'dummy' variables for categorical explanatory variables is one form of the general linear model. **McCulloch & Searle (2001, p. 1)** describe the essence of the general linear model, "...the mean of each datum [is] taken as a linear combination of unknown parameters ..., and the data [are] deemed to have come from a normal distribution... The model is linear in the parameters, so 'linearity' also includes being of the form $\mu_{ij} = b_o + b_1 X_{1ij} + b_2 x_{2ij}^2$, where the *x*s are known and there can be (and often are) more than two of them." ANOVA is a subset of the general linear model and regression, and almost all ANOVA problems, can be analyzed as regression problems, although some problems (such as **Model II ANOVA** or **mixed model ANOVA**) are better handled as ANOVA problems since it is often difficult to determine the appropriate error term for testing hypotheses regarding random effects in a regression context.

Generalized linear model "A Generalized Linear Model (GLM) is a probability model in which the mean of a response variable, or a function of the mean, is related to explanatory variables through a regression equation." **Ramsey & Schafer 2002 p 584 (or 1997 p. 568)** The data are not necessarily assumed to be normally distributed. **Probit analysis** (appropriate for types of survival data), **tobit analysis** (for censored data), and **logistic regression** (binary and binomial), and Poisson loglinear regression are regarded as types of generalized linear model.

Geometric distribution **http://mathworld.wolfram.com/GeometricDistribution.html** The probability function is

$$\begin{aligned} P(n) &= p(1-p)^n \\ &= pq^n \end{aligned}$$

Geometric series Sum of a geometric series, where 0 < x <1:

$$\begin{gathered} \textit{Sum of a geometric series}: \\ \textit{if } 0 < x < 1 \\ \sum_{k=0}^{\infty} x^k = \frac{1}{1-x}. \end{gathered}$$

or (from **Abromowitz & Stegun, 1965**)

$$\begin{gathered} \textit{Sum of a geometric progression to n terms}: \\ s_n = a + ar + ar^2 + \ldots + a^{n-1} = \frac{a(1-r^n)}{1-r}. \\ \lim_{n\to\infty} s_n = \frac{a}{1-r} \qquad (-1 < r < 1) \end{gathered}$$

Note, **Nahin (2002)** solves several problems in probability, including 'The Duelling Idiots' problem of the title using the sum of convergent geometric series. Often, absorbing Markov chains can be used to model these problems.

Gompertz distribution

Goodness of fit

Gosset, W. S. Developed **Student's t distribution** in 1908

GLM Generalized linear model

GLS Generalized least squares

Hazard rate, hazard ratios
http://www.weibull.com/AccelTestWeb/proportional_hazards_model.htm

Heteroscedasticity Unequal variance or unequal spread *cf.*, **homoscedasticty**

Homoscedasticity Equal variance or equal spread *cf.*, heteroscedasticity

Hotelling's T^2 A multivariate test for the difference in location. It is a generalization of the univariate **Student's t** test. SPSS prints Hotelling's trace, which is $T^2/(N-1)$.

Hurlbert's $E(S_n)$ The expected number of species from a random draw of n individuals from a sample (**Hurlbert 1971**)

$$E(S_n) = \sum_{k=1}^{S} 1 - \frac{\begin{pmatrix} N-N_k \\ n \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}.$$

$$\begin{aligned} where, \quad n &= random\ sample\ size. \\ \begin{pmatrix} N \\ n \end{pmatrix} &= binomial\ coefficient. \\ &= No.\ of\ ways\ to\ sample\ N\ objects,\ n\ at\ a\ time. \\ &= \frac{N!}{(N-n)!\cdot n!} \\ N &= Total\ individuals\ in\ sample. \\ N_k &= Individuals\ of\ species\ k. \\ S &= Number\ of\ species. \end{aligned}$$

Hypergeometric distribution

Hypethetico-deductive method

ICA Independent components analysis, used to solve 'the cocktail party problem'

http://www.cis.hut.fi/projects/ica/cocktail/cocktail_en.cgi

http://www.cis.hut.fi/projects/ica/fastica/

See **http://ica2001.ucsd.edu/index_files/pdfs/115-hundley.pdf**

Independence (**Larsen & Marx 2001, p. 70 Definition 2.7.1**) Two events are said to be **independent** if and only if P(A ∩ B) = P (A) · P(B), otherwise A and B are dependent events. For more than two events: **(Larsen & Marx Definition 2.7.2**) Events A_1, A_2, …, A_n are said to be independent if for every set of indices i_1, i_2, …, i_k between 1 and n, inclusive,

$$P(A_{i1} \cap A_{i2} \cap \cdots \cap A_{ik}) = P(A_{i1}) \cdot P(A_{i2}) \cdot \cdots \cdot P(A_{ik})$$

Theorem from **Hogg & Tanis (1977, p. 42)** If A and B are independent events, then the following pairs of events are also independent: (i) A and B', (ii) A' and B, (iii) A' and B'. [A' is the **complement** of A]

Independent trials process The outcome of the process is unaffected by earlier events. *cf.*, **Bernoulli trial**.

Inference An **inference** is a conclusion that patterns in the data are present in some broader context A statistical inference is an inference justified by a **probability** model linking the data to the broader context **Ramsey & Schafer (1997)**

Inter-quartile range See **boxplots**

Intersection (Definition 2.2.1 from **Larsen & Marx 2001, p. 24**) Let *A* and *B* be any two events defined over the same **sample space** *S*. Then

- The **intersection** of *A* and *B*, written *A* ∩ *B*, is the event whose outcomes belong to both *A* and *B*.
- The **union** of *A* and *B*, written *A* ∪ *B*, is the event whose outcomes belong to either A or B or both.

Jackknife In a 1-sample bootstrap, 1 sample is dropped from a sample of n and the statistical test repeated. n samples produces n jackknifed samples. Usually, the value of the test statistic is subracted from the test statistic based on all n samples (with a scaling for sample size) to produce the Tukey jackknife pseudovalue *cf.*, **bootstrap**

Kaplan-Meier survival analysis Available as an SPSS advanced model (analyze\survival\kaplan-meier). From the SPSS help file: "There are many situations in which you would want to examine the distribution of times between two events, such as length of employment (time between being hired and leaving the company). However, this kind of data usually includes some censored cases. Censored cases are cases for which the second event isn't recorded (for example, people still working for the company at the end of the study). The Kaplan-Meier procedure is a method of estimating time-to-event models in the presence of censored cases. The Kaplan-Meier model is based on estimating conditional probabilities at each time point when an event occurs and taking the product limit of those probabilities to estimate the survival rate at each point in time." See also: **http://www.statsoft.com/textbook/stsurvan.html** **http://www.cmh.edu/stats/model/survival/kaplan.asp** *cf.*, **Cox regression**

Kendall's τ (**Kendall's tau**) The most non-parametric of correlation coefficients. The sign of the difference of all combinations of two observations in one vector of data are compared with observations holding the same positions in the 2nd list of data vector. If the signs in both sets are the same, the match is concordant. Kendall's tau is the ratio of {concordant - discordant ranks} to total possible ranks. Using Kendall's triangle, exact p values can be calculated if there are no tied ranks. *Cf.*, **Spearman's ρ**

Kendall's tau-b Stuart's tau-c makes an adjustment for table size in addition to a correction for ties. Tau-c is appropriate only when both variables lie on an ordinal scale. **http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap28/sect20.htm**

Kendall's tau-c Kendall's tau-b is similar to **gamma** except that tau-b uses a correction for ties. Tau-b is appropriate only when both variables lie on an ordinal scale. **http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap28/sect20.htm**

Kolmogorov-Smirnov test A test of whether one cumulative frequency distribution (cfd) differs from another. A one-sample test compares a known cfd with an observed cfd. A two-sample test compares two cfds. The test statistics is the maximum difference between cfds.

Kruskal-Wallis test Nonparametric one-way ANOVA. This is a k-independent samples extension of **Wilcoxon rank sum** test *cf.*, **Friedman's ANOVA**

Kurtosis The fourth moment about the mean (**Larsen & Marx, 2001 p. 233**). The peakedness of a distribution. a flat pdf is callyed **platykurtic**, while a peaked distribution is called **leptokurtic**. cf., **skewness**

Lack of fit In a regression analysis with one explanatory factor, if there are true replicate observations at one or more values of the explanatory variable, then the residual variation from a simple **least squares** regression can be partitioned into pure error and lack of fit components (**Ramsey & Schafer 1997, p. 212**):

$$\begin{array}{ccccc} Y_{ij}-(\hat{\beta}_o+\hat{\beta}_1 X_{ij}) & = & \left[Y_{ij}-\bar{Y}_i\right] & + & \left[\bar{Y}_i-(\hat{\beta}_o+\hat{\beta}_1 X_{ij})\right]. \\ \textit{Residual} & = & \textit{Pure error} & + & \textit{Lack of fit.} \end{array} \qquad \textbf{(46)}$$

The pure error sum of squares can be obtained by performing a one-way ANOVA to test for differences in means among replicated groups. This one-way ANOVA will also produce the among replicated means sum of squares. If there are n groups of replicated means, there are n-1 df for this among means SS. The lack of fit SS is the among means sum of squares minus the regression sum of squares (with 1 df). Therefor, the lack of fit MS has n-2 df, where n is the number of replicated groups. The Lack of Fit F-test uses the ratio of the Lack of Fit MS over the Pure error MS. The former measures the departure of the mean of replicated observations from the line and the latter the within group variation. **Draper & Smith (1998)** recommend performing a lack of fit F test and only pooling the two sources of residual variation into the regression error sum of squares if the lack of fit test is not significant (p>0.05).

Laplace, Pierre Simon (1749-1827) Described the **central limit theory**

Latent class analysis **http://ourworld.compuserve.com/homepages/jsuebersax/index.htm** or **http://www2.chass.ncsu.edu/garson/pa765/latclass.htm**

Latent trait models for rater agreement **http://ourworld.compuserve.com/homepages/jsuebersax/ltrait.htm**

Latent variables Unmeasured variables or factors estimated from measured variables and used in **structural equation models**

Latin hypersquare sampling A **Monte Carlo method** used to assess the variance of model predictions *cf.*, Monte Carlo method

Latin squares A method of arranging experimental units in a 2-factor (or more rarely a 3- or 4-factor) ANOVA

Least significant difference LSD Sometimes called Tukey's LSD. A pair of means is tested using the ANOVA error mean square and df for testing differences among means. Contrasts tested with the LSD must be established **a priori**, because the test offers no protection against the inflation of Type I error due to multiple hypothesis testing. Indeed the LSD test is more powerful than the independent samples **t test** if there are more than 2 groups.

Least squares Adrien Marie **Legendre (1805)** clearly described the method of least squares (**Stigler 1986, p. 13**), an improvement of **Laplace's** earlier work in minimizing the sum of absolute deviations:

> ***On the method of least squares***
>
> *In most investigations where the object is to deduce the most accurate possible result from observational measurements, we are led to a system of equations of the form:*
>
> $$E = a + bx + cy + fz + \&c.,$$
>
> *in which a, b, c, f, &c. are known coefficients varying from one equation to the other, and x, y, z, &c. are unknown quantities, to be determined by the condition that each value of E is reduced either to zero or to a very small quantity … Of all the principles that can be proposed for this purpose, I think there is none more general, more exact, or easier to apply, than that which we have used in this work; it consists of* making the sum of the squares of the errors a minimum. *By this method, a kind of equilibrium is established among the errors which, since it prevents the extremes from dominating, is appropriate for revealing the state of the system which most nearly approaches the truth.*

… We see therefore, that the method of least squares reveals, in a manner of speaking, the center around which the results of observations arrange themselves, so that the deviations from the center are as small as possible." (Adrien Marie Legendre, 1805).
Legendre's paper can be read in translation at:
http://www.stat.ucla.edu/history/legendre.pdf

Least squares regression Solving the **regression model** by minimizing the sum of squares of residuals of the response variable to the regression line.*cf.*, **regression**

Leibniz, Gottfried Wilhelm (1646-1716) (**Larsen & Marx 2001 p 86**) The 1666 treatise, "*Dissertatio de arte cominatoria*" was perhaps the first monograph written on **combinatorics**

Levene's test There are at least 4 different versions of Levene's test for equality of variance, an assumption of general linear models (e.g., ANOVA, regression, 2-sample t tests). They are all ANOVAs of deviations from the mean (squared deviation from mean, squared deviation from median, absolute deviation from the mean or absolute deviation from the median). All test the equal-variance assumption in ANOVA or regression. SPSS calculates the Levene's test based on absolute deviations from the group mean. A **Brown-Forsythe test** for equal variance performs an ANOVA on the absolute deviation from group medians. *Cf.*, **homoscedasticity**, **heteroscedasticity**

Leverage The deviation of an individual case from the range of explanatory variables. Cases with high leverage have the potential for being outliers, with outliers being more readily detected by **Cook's D**. See **http://www.j.org/v02/i05/pirls/node15.html**

Likelihood from **Mathworld** Likelihood is the hypothetical probability that an event that has already occurred would yield a specific outcome. The concept differs from that of a probability in that a probability refers to the occurrence of future events, while a likelihood refers to past events with known outcomes. *Cf.*, **Maximum likelihood**

Likelihood function invented by **Fisher** 1922 A definition from **Mathworld**: A likelihood function L(a) is the probability or probability density for the occurrence of a sample configuration , ..., given that the probability density with parameter *a* is known

$$L(a) = f(x_1|a) \ldots f(x_n|a)$$

Likelihood ratio Hogg & Tanis 1977, p. 394 The likelihood ratio is the quotient $\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$,

where $L(\hat{\omega})$is the maximum of the **likelihood function** with respect to θ when $\theta \in \omega$ and $L(\hat{\Omega})$is the maximum likelihood function with respect to θ when $\theta \in \Omega$. *Cf.*, **Maximum likelihood**

Linear combination An estimated value obtained by a linear equation in which the coefficients need not add to zero. If the sum of coefficients is zero, then the linear combination is, by definition, a **linear contrast**. The variance of a linear combination and contrast can be readily calculated using formulae for the **propagation of error**. See:
http://www.itl.nist.gov/div898/handbook/prc/section4/prc426.htm

Linear contrast A contrast is a linear combination of 2 or more factor level means with coefficients that sum to zero. *Cf.*, **linear combination**, **orthogonal contrast** **http://www.itl.nist.gov/div898/handbook/prc/section4/prc426.htm**

Log-linear model "Log-linear modeling is an analog to multiple regression for categorical variables. When used in contrast to log-linear regression models like logit and logistic regression, log-linear modeling refers to analysis of tables without necessarily specifying a dependent. Rather the focus is in accounting for the observed frequencies." **http://www2.chass.ncsu.edu/garson/pa765/logit.htm**

Logistic regression A form of **generalized linear model**, with the logit link function, $\ln\left(\frac{P}{1-P}\right)$. There are several forms of logistic regression, including binary logistic regression (=bivariate logistic regression), with the response variable taking only two states (*e.g.,* dead or alive), and binomial logistic regression with the response variable taking on discrete values along the interval (0,1). See **http://www2.chass.ncsu.edu/garson/pa765/logistic.htm**

logit is the log odds function ln(p/(1-p)). Logits can be converted to probabilities, frequencies or proportions using p = 1-1/(1+Exp(Logits)), or p = exp(logits)/(1+exp(logits))

logit link function is used to convert probabilities or frequency data in logistic regression.

logit transform If x ranges between 0 & 1, then log [(x)/(1-x)] often expands the tail of a distribution. **Ramsey & Schafer (1997, 2002)** apply this transform to percentage cover data (scaled to range from 0 to 1) and call it the **regeneration ratio**.

Lognormal distribution

Longitudinal data The same subjects or experimental units are followed through time. Often analyzed with repeated measures designs

LSD Tukey's **Least significant difference** is a test of means using the error mean square from the overall ANOVA as the estimate of pooled **standard error**. It does not protect against inflation of the **experimentwise error**. It is one of the least conservative of 20 or more **multiple comparison tests**.

Mallow's C_p *cf.*, **Bayesian information content**, **AIC**

Mann-Whitney U Test Algebraically identical to **Wilcoxon rank sum test**

Markov, Andrei Andreevich (1856-1922) **http://www-history.mcs.st-andrews.ac.uk/Mathematicians/Markov.html**

Markov chain A **stochastic model** in which the future state of the system can be predicted from the probability matrix and the state of the system on the previous time step (see also **absorbing Markov chain** and **ergodic Markov chain**).

http://en.wikipedia.org/wiki/File:AAMarkov.jpg

Figure 6. A.A. Markov.

Markov chain Monte Carlo (MCMC) A search method used to estimate model parameters.

Markov property (process) A Markov process is defined as a stochastic process with the property for any set of *n successive* times (*i.e.*, $t_1<t_2<\ldots<t_n$) one has

$$P_{1|n-1}(y_n,t_n|y_1,t_1;\ldots;y_{n-1},t_{n-1})=P_{1|1}(y_n,t_n|y_{n-1},t_{n-1}). \qquad (1.1)$$

In other words, the conditional probability density at t_n, given the value y_{n-1} at t_{n-1}, is uniquely determined and is not affected by any knowledge of the values at earlier times. $P_{1|1}$ is called the **transition probability**. **Van Kampen (1981, p. 76)**

Maxmium likelihood "The **maximum likelihood estimate** of a parameter is defined to be the parameter value for which the **probability** of the observed data takes its greatest value." **Agresti 1996, p. 9** From **Mathworld**: Maximum likelihood, also called the maximum likelihood method, is the procedure of finding the value of one or more parameters for a given statistic which makes the known likelihood distribution a maximum. The maximum likelihood estimate for a parameter μ is denoted $\hat{\mu}$. See: **http://www.mathdaily.com/lessons/Maximum_likelihood** and **http://socserv.socsci.mcmaster.ca/jfox/Courses/SPIDA/MLE-basic-ideas.pdf**

Maximum likelihood estimator From **Mathworld**: A maximum likelihood estimator is a value of the parameter such that the **likelihood function** is a maximum. see **estimators**

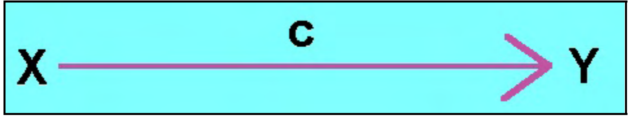
McNemar's test A test of proportions in a 2 x 2 classification for repeated measures (paired) data. **http://www.amstat.org/publications/jse/secure/v8n2/levin.cfm**

MDS Multidimensional scaling (sometimes called NMDS, for nonmetric multidimensional scaling). See **http://forrest.psych.unc.edu/teaching/p208a/mds/mds.html**

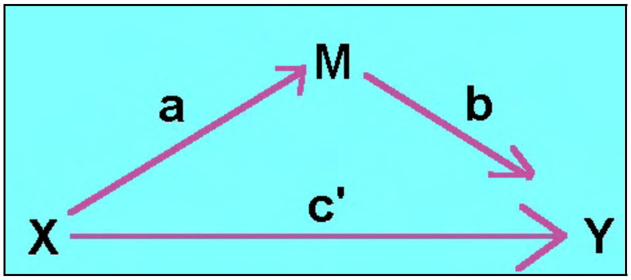
Measurement scales Variables can be classified into nominal, ordinal, interval and ratio scales of measurement. **Stevens (1951)** and **Roberts (1976)** show that some mathematical operations require at least an interval or even ratio scale of measurment. For example, temperature on the Fahrenheit scale is an interval measure and the ratio of interval measurments is meaingless. Some procedures, like **Factor Analysis**, assume at least an interval scale of measurement. **Vellman & Wilkinson (1993)** review 40 years of research indicating that Stevens' proscriptions may have been too severe, e.g., it is valid to calculate average GPA from an ordinal measure.

Median The median is the sample item that is in the middle in magnitude, or if the average of the two middle items if the number of items is even.

Mediation **http://davidakenny.net/cm/mediate.htm** Consider a variable X that is assumed to affect another variable Y. The variable X is called the initial variable and the variable that it causes or Y is called the outcome. In diagrammatic form, the unmediated model is



The effect of X on Y may be mediated by a process or mediating variable M, and the variable X may still affect Y. The mediated model is



The mediator has been called an intervening or process variable. Complete mediation is the case in which variable X no longer affects Y after M has been controlled and so path c' is zero. Partial mediation is the case in which the path from X to Y is reduced in absolute size but is still different from zero when the mediator is controlled. When a mediational model involves latent constructs, structural equation modeling or SEM provides the basic data analysis strategy. If the mediational model involves only measured variables, however, the basic analysis approach is multiple regression or OLS. Regardless of which data analytic method is used, the steps necessary for testing mediation are the same.

- Step 1: Show that the initial variable is correlated with the outcome. Use Y as the criterion variable in a regression equation and X as a predictor (estimate and test path c). This step establishes that there is an effect that may be mediated.
- Step 2: Show that the initial variable is correlated with the mediator. Use M as the criterion variable in the regression equation and X as a predictor (estimate and test path a). This step essentially involves treating the mediator as if it were an outcome variable.
- Step 3: Show that the mediator affects the outcome variable. Use Y as the criterion variable in a regression equation and X and M as predictors (estimate and test path b). It is not sufficient just to correlate the mediator with the outcome; the mediator and the outcome may be correlated because they are both caused by the initial variable X. Thus, the initial variable must be controlled in establishing the effect of the mediator on the outcome.
- Step4: To establish that M completely mediates the X-Y relationship, the effect of X on Y controlling for M should be zero (estimate and test path c'). The effects in both Steps 3 and 4 are estimated in the same regression equation.

See also **http://www.public.asu.edu/~davidpm/ripl/mediate.htm**

Metaanalysis From **Wikipedia** A meta-analysis is a statistical practice of combining the results of a number of studies that address a set of related research hypotheses. The first meta-analysis was performed by Karl Pearson in 1904, in an attempt to overcome the problem of reduced statistical power in studies with small sample sizes; analyzing the results from a group of studies can allow more accurate estimation of effects ...Modern meta-analysis does more than just combine the effect sizes of a set of studies. It tests if the studies' outcomes show more variation than the variation that is expected because of sampling different research participants. If that is the case, study characteristics such as measurement instrument used, population sampled, or aspects of the studies' design are coded. These characteristics are then used as predictor variables to analyze the excess variation in the effect sizes.

Metric A dissimilarity measure obeying the following four axioms, the last being the **triangular inequality axiom** (Legendre & Legendre 1983, p. 193):

1) if a = b, D(a,b)=0
2) if a ≠ b, D(a,b)>0
3) D(a,b)=D(b,a)
4) D(a,b)+D(b,c)≥D(a,c), as the sum of 2 sides of a triangle is necessarily equal to or larger than the third side (triangle inequality axiom)

see also **semimetric**, **Triangular inequality** & **ultrametric**.

Mill's cannon of the difference J. S. Mill's (1843) fifth cannon of experimental enquiry (The cannon of difference) "Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation" **Kendall & Stuart (1979)** find two major problems with basing an experimental or sampling design on Mill's 5th cannon: 1) the one-phenomenon (factor)-at-a-time approach does not work, and 2) "We can never be quite sure that all the important, or even the most important, causal factors have been incorporated in the structure of the experiment. Some may be quite unknown; others although known, may wrongly be considered to be of minor importance and deliberately neglected. We always need to guard against the perversion of the inferences within an experiment by adventitious outside effects."

Mixed model (linear mixed model) Mixed models are either general linear models or generalized linear models which contain both fixed and random factors. Generalized mixed models are a subset of generalized linear models, which include logistic, **probit** & log-linear models, in which random and fixed effects can affect the response variable. Linear mixed models are particularly useful for **longitudinal data**, in which the same subjects are followed through time. All such repeated measures designs, including paired t-tests, can be regarded as a subset of mixed models.. The standard mixed model is of the form

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

where, β contains population parameters describing average responses to external variables and b_i contains subject-specific parameters describing how the i'th subject deviates from the average population, and ε_i is a vector of error components. The matrices X_i and $Z_i$are covariates. Mixed models, including those in SPSS, allow a variety of error structures including analyses of autocorrelated errors and other features of repeated measures designs. There are a number of different methods for estimating model parameters, including restricted maximum likelihood, penalized likelihood, Bayesian techniques, and simulated maximum likelihood. Adequacy of models involve likelihood tests, with penalties for fitted parameters.

Minimum noise fraction (MNF) (or Maximum Noise Fraction) an eigendecomposition method used in satellite remote sensing. Examples: **http://www.earthsat.com/geo/oil&gas/hydrocarbon_MNF.html** or **http://www.eoc.csiro.au/hswww/oz_pi/svt_hilo/burke.pdf**

modus tollens The logical syllogism: "If A then B, not B implies not A" The ***modus tollens*** is the basis of **Popper's** method of falsificationism.

Monte Carlo method Any method, including bootstrap sampling, that uses randomly selected subsets of the data to estimate model parameters *cf.*, **bootstrap**, **permuation analysis**

Monty Hall Problem A contestant must choose one of three doors. Behind one door is a desirable prize and the other two contain goats. The contestant picks one door, and Monty Hall immediately opens one of the two remaining doors, revealing a goat. Monty then offers the remaining unopened door for the door you've chosen. Should you switch?

Multicollinearity (**Collinearity**) In multiple regression, if there is a strong correlation among explanatory variables, neither the sign nor the magnitude of the coefficient can be trusted. **Draper & Smith (1998, p. 369)** provide the following description of multicollinearity:

> Suppose we wish to fit the model $\mathbf{Y}=\mathbf{X\beta}+\boldsymbol{\varepsilon}$, The solution $\boldsymbol{b}=(\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$would usually be sought [b=X\Y in Matlab]. However, if **X'X** is singular, we cannot perform the inversion and the **normal equations** do not have a unique solution. (An infinity of solutions exists instead). … at least one column of **X** is linearly dependent on (i.e., is a linear combination of) the other columns. We would say that collinearity (or multicollinearity) exists among the columns of **X**.

Multicollinearity is tested with the **variance inflation factor** or **VIF** or the tolerance. VIF=1/tolerance. Tolerance = 1 -R_k^2, where R_k^2 is the amount of variation in one explanatory variable explained by the other explanatory variables.

Other links: UCLA Statistics page (Note that the implication that VIF's less than 20 aren't cause for concern is a dubious bit of advice. VIF's as low as 3 or 4 could create problems in a regression model

http://www.ats.ucla.edu/stat/stata/modules/reg/multico.htm

Multilevel models See **Singer & Willett (2003)** And **http://www2.chass.ncsu.edu/garson/pa765/multilevel.htm**

Multinomial distribution

Multiple comparison procedures (**multiple hypothesis tests**, ***a posteriori*** **tests**, post hoc tests) After an ANOVA indicates that all means are equal, an investigator may wish to know which pairs or groups of means differ. A variety of multiple comparisons tests, also known as ***a posteriori*** **contrasts**, have been developed to test for differences among means while considering the overall or **experiment-wise error**. These tests include **Bonferroni**, Duncan's, Dunn's, Dunnett's (for comparisons with a control group), Dunnett's C (for unequal variances), Dunnett's T3 (for unequal variances, Gabriel, Games-Howell (conservative for unequal n and unequal variance), Hochberg's GT2, Least significant difference or LSD (not conservative), **Scheffe's**, Sidak, Student-Newman Keuls (SNK), Tamhane's T2 (for unequal variances), **Tukey's HSD**, **Tukey-Kramer (a general form of Tukey's HSD)**, Waller (Bayesian approach for equal n). There are at least 20 such tests (see **Sokal & Rohlf (1995)** for a thorough discussion). The most important are perhaps the **Bonferroni**, **Tukey-Kramer**, and **Scheffé** tests, which adjust for the number of a posteriori contrasts. The Bonferroni adjustment for **experimentwise alpha level** can be used in any test. **Hotelling's T^2** can be used to adjust for multiple **correlation** (*cf.*, multiple hypothesis testing)

Multiple regression A **regression** with more than one explanatory variable.

Multiplication rule of combinatorics **http://www.math.uah.edu/stat/comb/comb1.html#Multiplication**

Multivariate hypergeometric distribution **http://www.math.uah.edu/stat/urn/urn4.html**

Mutually exclusive Events *A* and *B* defined over the same **sample space** are said to be **mutually exclusive** if they have no outcomes in common – that is, if $A \cap B = \varnothing$, where $\varnothing$ is the **null set** (**Larsen & Marx 2001, Definition 2.2.2**)

MVUE **Minimum variance unbiased estimator**

Natural logarithms [often inidicated with ln(x)] Logarithms were invented by the Scot John Napier in 1614 and natural logarithms are logarithms to the base e, but Napier didn't use the exponential function, another case of **Stigler's law of eponymy**.

Negative binomial distribution The random variable X is said to have a **negative binomial distribution** if

$$p_x(k) = P(X=k) = \left(\frac{r}{r+\lambda}\right)^r \frac{\Gamma(r+k)}{\Gamma(k+1)\,\Gamma(r)} \left(\frac{\lambda}{r+\lambda}\right)^k, \; k = 0,1,2,\ldots$$

$$\textit{where, } \lambda = \textit{population rate parameter} > 0.$$

The expected value E(X), like the Poisson distribution is λ but $\mathrm{Var}(X) = \lambda + \lambda^2/r$, where r is called the dispersion parameter. See **http://ehs.sph.berkeley.edu/hubbard/longdata/webfiles/poissonmarch2005.pdf**

Negative binomial regression In the analysis of count data, **Poisson regression** assumes the variance equals the mean. Overdispersion, or the variance exceeding the mean, may indicate the need for a negative binomial regression. **http://www.uky.edu/ComputingCenter/SSTARS/P_NB_3.htm**

Nested ANOVA An Analysis of variance in which the experimental units are a subset of treatment levels

Newton-Raphson method (**Newton's method**) Used to estimate the parameters of generalized linear models, as described **McCullagh & Nelder (1989, Ch 2, p. 40-41)**. **http://mathworld.wolfram.com/NewtonsMethod.html**

Neyman-Pearson school A **frequentist** statistical research program led by Jerzy Neyman and Egon Pearson. They introduced **critical values** and **confidence limits**

NIPALS Algorithm ("Nonlinear Iterative Partial Least Squares") In addition to solving **partial least squares** problems, can be used to find dominant eigenvalues & eigenvectors of a square matrix.

Nonparametric statistics An underlying parametric distribution, such as the normal distribution, is not assumed for the data. Normal and chi-square distributions are often used for calculating p values.

Nonsense correlation cf., **spurious correlation**

Normal distribution [**Gaussian distribution**, Normal curve, **error function**]. **Stigler (1986, p. 284)** traces this distribution to Abraham De Moivre (1733). Figure 7 shows a Matlab ezplot of a normal distribution with mean 5.3 and sd 1.3.

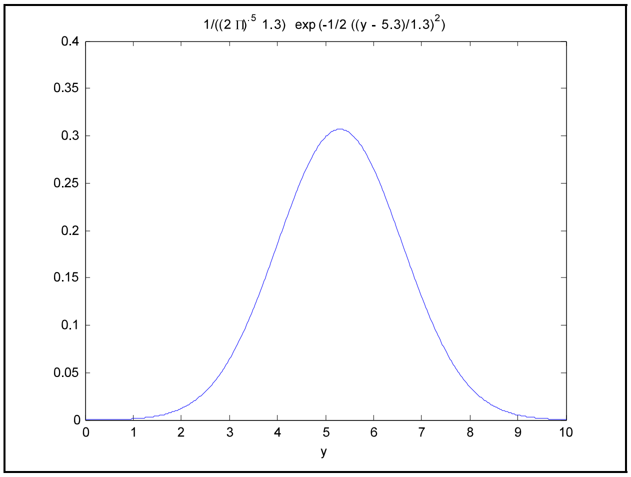


Figure 7. Matlab Ezplot of normal distribution, shown in above equation.

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}. \tag{59}$$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{x-\mu^2}{2\sigma^2}}. \tag{60}$$

$$f(y) = \frac{1}{\sqrt{2\pi}\ \sigma}\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]. \tag{61}$$

$$-\infty < y < \infty,\ \ -\infty < \mu < \infty,\ \ \sigma > 0$$

Normal equations According to **Stigler (1999, p. 415-420)**, a term introduced by **Gauss** (1822) ("normalgleichungen") to describe how the **least squares method** could be applied. **Stigler (1986, p. 14)** argues that the concept of the normal equations was used by **Legendre (1805)**. *Cf.*, **least squares**, **regression**

Null hypothesis In the frequentist school of statistics, the null hypothesis is the hypothesis that statistical tests are designed to reject (*cf.*, ***modus tollens***, **Type I error**, **Type II error**).

Observational study The sampling units are inherently finite *cf.*, **experiment**

Odds The odds are a different way of expressing the probability of an event occurring. If you know the probability of an event, then you can calculate the odds. If the probability of an event is p, then odds=p/(1-p):

$$Odds\ of\ an\ event = \frac{Probability\ of\ an\ event\ occurring}{Probability\ of\ an\ event\ \mathbf{not}\ occurring} = \frac{p}{(1-p)}.$$

If the probability of rain today is 50% or 0.5 then the odds of rain are 0.5/(1-0.5) = 1/1 = 1:1 or 1 to 1. This is also expressed as saying the odds of rain today are even. The odds of getting a six when rolling a die (the singular of dice) is $\frac{Probability\,of\,event\,occuring}{Probability\,of\,event\,not\,occuring} = \frac{\frac{1}{6}}{1-\frac{1}{6}} = \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{5} = 1:5$. This is sometimes expressed as saying the odds being 5 to 1 against getting a six. The odds of getting the King of Hearts when drawing a single card from a 52-card deck is $\frac{\frac{1}{52}}{1-\frac{1}{52}} = \frac{\frac{1}{52}}{\frac{51}{52}} = \frac{1}{51} = 1:51$. This is could be expressed as saying the odds are 51 to 1 against drawing a King of Hearts. Bets in horse races are set by the odds. If the probability of the favorite horse winning a race is 60% then the odds are $\frac{0.6}{1-0.6} = \frac{0.6}{0.4} = \frac{3}{2} = 3:2$. This horse would be listed as a 3:2 favorite. In order to win $2, you'd have to bet $3. If you bet $3, you would get $5 back if the horse won. A longshot in a horse race might have a probability of winning of 1%. The odds of that horse winning would be $\frac{0.01}{1-0.01} = \frac{0.01}{0.99} = \frac{1}{99} = 1:99$. This would be expressed as saying the odds are 99 to 1 against winning. If you bet $1 on this horse and it won, you'd win $99.

As shown at right, some casinos express odds using the notation '10 for 1'. This means that a $1 dollar bet at 9 to 1 odds returns $10. Richard Frey (1970, p 269) in his edition of 'According to Hoyle' describes the convention of reporting odds with 'for:'

> "On many [craps] layouts the actual odds being offered are disguised by the use of the word "for." If the house, for example, pays 4-to-1 odds, the winner of a bet receives his $1 bet back together with $4 paid by the house, a total of $5. Some houses quote these odds by offeinr "five-for-one," meaning that for every $1 the bettor puts up, he recieves, when he wins, $5 — including his own $1. This is equivalent ot odds of 4-to-1."

The odds, ω, can be converted to a probability by using the relation that if the odds of yes are ω, P(yes)=ω/(ω + 1). So 4-to-1 odds would have a probability of 1/(4+1) or 0.2. Odds reported as 5 for 1 would have a p value of 0.2

[cf., **craps**, **odds ratio**]

Odds ratio the ratio of two odds. If the odds of a person getting a cold taking vitamin C are 3 and the odds of a person getting a cold taking a placebo are 4.5, then the estimated odds ratio is 1.5. One can say that the odds of getting a cold are 50% greater if one doesn't take vitamin C. **Ramsey & Schafer (2002, p. 540)** prefer reporting the odds ratio to differences in proportions because: 1) the odds ratio tends to remain more nearly constant over levels of confounding variables, 2) the odds ratio is the only parameter that can describe the binary responses of two groups from a retrospective study, and 3) the comparison of odds extends nicely to logistic regression analysis.

OLS Ordinary least squares *cf.*, **WLS**

One-sided test also called **one-tailed test** *cf.*, **two-sided test**

orthogonal right angle

Orthogonal arrays In experimental design, you might want to test 1000 drugs on a mammalian cell culture, including the 2- and 3-way interactions. How can that be done with a relatively small number of experimental units. Orthogonal arrays and factorial designs provide ways of constructing choices of factors and levels of factor and the analyses that can be performed on them. See **http://support.sas.com/techsup/technote/ts723.html** or **http://support.sas.com/techsup/tnote/tnote_stat.html#market**. **Sleuth Chapter 24** provides a concise introduction to the concepts. *Cf.*, experimental design

Orthogonal contrast Two contrasts are orthogonal if the sum of the products of corresponding coefficients (*i.e.,* coefficients for the same means) adds to zero. *Cf.*, **linear contrast** **http://www.itl.nist.gov/div898/handbook/prc/section4/prc426.htm**

orthonormal basis **http://mathworld.wolfram.com/OrthonormalBasis.html**

Overdispersion In fitting binomial and Poisson logistic regression models, the variance of the observed data is greater than that predicted from the variance expected from the binomial or Poisson models. "The terms ***extra-binomial variation*** and ***overdispersion*** describe the inadequacy of the binomial model in these situations." **Ramsey & Schafer (2002, p. 621)**.

Overfitting "When a [regression] model is fitted that is too complex, that is, it has too many free parameters to estimate for the amount of information in the data, the worth of the model (*e.g.*, R^2) will be exaggerated and future observed values will not agree with the predicted values. In this situation, ***overfitting*** is said to be present, and some of the findings of the analysis come from fitting noise or finding spurious associations between X and Y." **Harrell (2002, p. 60)**

p value (from K Wuensch, edstat, 3/19/03) The probability of obtaining data as or more discrepant with the null hypothesis than are those in the present sample, assuming that the null hypothesis is absolutely correct. Jerry Dallal has a lengthy discussion on his web site: **http://www.tufts.edu/~gdallal/pval.htm**

Paired *t* test A form of Student's t test in which observations are paired and the null hypothesis is that the difference between paired observations is equal to some value (usually zero). This is a form of **repeated measures design**.

Parameter An unknown numerical value describing a feature of a probability model. Parameters are indicated by Greek letters (**Ramsey & Schafer 1997, p. 19**) *cf.*, **statistic**

Partial Least Squares (**PLS**) "In partial least squares regression, prediction functions are represented by factors extracted from the Y'XX'Y matrix. The number of such prediction functions that can be extracted typically will exceed the maximum of the number of Y and X variables. In short, partial least squares regression is probably the least restrictive of the various multivariate extensions of the multiple linear regression model. This flexibility allows it to be used in situations where the use of traditional multivariate methods is severely limited, such as when there are fewer observations than predictor variables. Furthermore, partial least squares regression can be used as an exploratory analysis tool to select suitable predictor variables and to identify outliers before classical linear regression. Partial least squares regression has been used in various disciplines such as chemistry, economics, medicine, psychology, and pharmaceutical science where predictive linear modeling, especially with a large number of predictors, is necessary. Especially in chemometrics, partial least squares regression has become a standard tool for modeling linear relations between multivariate measurements (de Jong, 1993)." **http://www.statsoft.com/textbook/stpls.html**, **http://www.vcclab.org/lab/pls/m_description.html**, See also **NIPALS**, **SIMPLS**, **WA-PLS**

Pascal, Blaise (b. 6/19/1623, Clermont Auvergne France d. 1662). With **Fermat**, the father of mathematical probability theory. (**Bell 1937, p. 86**)

Path analysis **Sewall Wright** invented this technique in 1921 in the paper "Causation and Correlation" to explain the causal basis of a set of partial correlations among a set of variables. Sometimes referred to as causal analysis. Path analysis is now a subset of **structural equation modeling**.

pdf Acronym for probability density function

Pearson, Karl 1857-1936

Pearson's r Pearson's product-moment **correlation coefficient**

Permutations (**Larsen & Marx 2001, p. 92**) Theorem 2.9.1 The number of **permutations** of length *k* that can be formed from a set of *n* distinct elements, repetitions not allowed, is denoted by the symbol $_n\mathrm{P}_k$, where

$${}_nP_k \;=\; n(n-1)(n-2)\cdots(n-k+1) \;=\; \frac{n!}{(n-k)!}.$$

Corollary The number of ways to permute an entire set of n distinct objects is n! The symbol n! is called n **factorial**

Phi Phi is a chi-square based measure of association, sometimes called Pearson's coefficient of mean-square contingency, though sometimes this term is applied to Pearson's contingency coefficient, discussed below, which is a modification of phi.. With dichotomized continuous data, **tetrachoric correlation** is preferred. phi = (bc-ad)/sqrt[(a+b)(c+d)(a+c)(b+d)], **http://www2.chass.ncsu.edu/garson/pa765/assocnominal.htm**

Poisson, Siméon Denis (1781-1840)

Poisson distribution (**Larsen & Marx 2001 Theorem 4.2.2, p. 251**) First described by Poisson as a limit theorem and then used in 1898 by Professor Ladislaus von Bortkiewicz to model the number of Prussian cavalry officers kicked to death by their horses.

The random variable X is said to have a **Poisson distribution** if

$$p_x(k) \;=\; P(X=k) \;=\; \frac{e^{-\lambda}\lambda^k}{k!}\;,\; k \;=\; 0,1,2,\ldots$$

$$\textit{where},\ \lambda \;=\; \textit{population rate parameter} \\ > 0.$$

The expected value E(X) and variance (Var(X)) are both λ.

Poisson limit theorem (**Larsen & Marx 2001 Theorem 4.2.2, p. 251**) If n →∞ and p →0 in such a way that λ = np remains constant, then for any nonnegative integer k,

$$\underset{n\to\infty}{LIM}\binom{n}{k}p^{k}(1-p)^{n-k} = \frac{e^{-np}(np)^{k}}{k!}.$$

The Poisson limit theorem justifies the use of the Poisson distribution to approximate the **binomial distribution**. The **Poisson approximation to the binomial** is quite accurate if n ≥ 20 and p ≤ 0.05 and is very good if n ≥100 and np ≤10,

Poisson process (**Hogg & Tanis 1977, p 78**) Let the number of changes that occur in a given continuous interval be counted. We have an approximate Poisson process with parameter λ > 0 if the following are satisfied: (i) the number of changes occurring in nonoverlapping intervals are independent, (ii) the probability of exactly one change in a sufficiently short interval of length h is approximately hλ, and (iii) The probability of two or more changes in a sufficiently short interval is essentially zero.

Poisson regression A generalized linear model for the analysis of count data, which meet the assumption that the variance of the counts equals the mean.

Popper, Sir Karl R. Austrian philosopher of science, who spent most of his career at the London School of Economics, who proposed his method of "conjectures of refutations" in his magnum opus *Logik der Forschung*, translated as **Logic of Scientific Discovery in 1959**. His scientific method is founded on the demarcation principle of falsificationism. Science is distinguished from pseudoscience because scientific principles are subject to falsification. His long career is profiled in the wonderful 2003 book, "Witgenstein's Poker."

Posterior distribution See **Bayes theorem**

Power The probability that a null hypothesis will be rejected, given that it is false. Power = 1 - P (**Type II error**) = 1-β. In order to calculate the statistical power, the **alternate hypothesis** must be specified. Bill Trochim's web page has a very nice discussion of statistical power (**http://trochim.human.cornell.edu/kb/power.htm**)

Power function

Precision indicates the random or chance variability about the mean of repeated observations (*cf.* **accuracy**)

PRESS Prediction error sum of squares.

principal component method = **Principal Components Analysis** **(PCA)** Developed by **Hotelling (1933)**. PCA is simply the rotation of the original system of axes in the multidimensional space. The principal axes are **orthogonal** and the **eigenvalues** measure the amount of variance associated with each principal axis. PCA is used to summarize in a few important dimensions the greatest part of the variability of a dispersion matrix of a large number of descriptors R-mode) or cases (Q-mode). *cf.*, **EOF**

principal component scores the value of a principal component for individual points, hence the new coordinates of data points measured along axes created by the principal component method. A principal component score can be regarded as an additional variable for each case, this variable is a linear function of the original variables.

Probability density function

Probit analysis A maximum likelihood regression procedure to estimate the proportion of a population that will be affected by a given treatment level. The method was pioneered by Bliss to analyze bioassay data from toxicology experiments. For example, a toxicologist might want to find the lethal dose required to kill 50% of a population of invertebrates in a beaker. Probit analysis is now regarded as one of many methods included among the **generalized linear models**. These generalized linear models are usually fit using the principle of maximum likelihood. In practice, the logistic regression modeling procedure often gives very similar results. See, **http://www2.chass.ncsu.edu/garson/pa765/logit.htm**

Probit link function **Agresti (1996, p 79)** writes, "The probit link applied to a probability $\pi(x)$ transforms it to the standard normal **z-score** at which the left-tail probability equals $\pi(x)$. For instance, probit (.05) = -1.645, probit (0.50) = 0, probit (.95) = 1.645, and probit (.975)=1.96. The probit model is a GLM with a random component and a probit link."

Probability **Larsen & Marx (2001)** provide four distinctly different definitions of probability:

- **Classical probability**, **Pascal** & **Fermat**
  - "Imagine an experiment, or game, having n possible outcomes—and suppose that those outcomes are equally likely. If some event A were satisfied by m out of those n, the probability of A [Written P(A)] should be set equal to m/n. This is the classical or *a priori* definition of probability.
- **Empirical probability** (Attributed to von Mises, but can be found at least a century earlier)
  - "Consider a sample space *S*, and any event *A*, defined on *A*. If our experiment were performed on them, either A or A^c would be the outcome. If it were performed n times, the resulting set of sample outcomes would be members of *A* on m occasions, m being some integer between o and n, inclusive. Hypothetically, we could continue the process an infinite number of times. As n gets large, the ratio *m/n* will fluctuate less and less. The number that *m/n* converges to is called the empirical probability of *A*, that is $P(A) = \lim_{n \to \infty} m/n.$
- **Axiomatic probability**. Andrei Kolmogorov
  - If S has a finite number of members, Kolmogorov showed that as few as three axioms are necessary and sufficient for characterizing the **probability function** P.
    - Axiom 1. Let A be any event defined over S. Then P(A) $\geq$0.
    - Axiom 2 P(S)=1.
    - Axiom 3 Let A and B be any two mutually exclusive events defined over S. Then
      $$P(A \cup B) = P(A) + P(B).$$
  - When S has an infinite number of members, a fourth axiom is needed

- Axiom 4 Let A_1, A_2, …, be events defined over S. If $A_i \cap A_j = \varnothing$for each i $\neq$j, then

$$P\left(\bigcup_{i=1} \right) = \sum_{i=1}^{\infty} P(A_i).$$

- From these simple statements, all other properties of the probability function can be derived.

- **Subjective probability**
    - What is a person’s measure of belief that an event will occur?

Humorous definitions 1) “probability” = long-run fraction having this characteristic. 2) “probability” = degree of believability. 3) A frequentist is a person whose lifetime ambition is to be wrong 5% of the time. 4) A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule. **http://www.statisticalengineering.com/frequentists_and_bayesians.htm**

Probability function There are two fundamentally different types of probability functions **(Larsen & Marx 2001)**.

A **discrete probability function** is a function defined for a process with a finite or **countably infinite** number of outcomes. Suppose that the sample space for a given experiment is either finite or countably infinite, Then any P such that **a)** 0 $\leq$P(S) for all s $\in$ S and **b)** $\sum_{\text{all}\, s \in S} P(s) = 1.$ The probability of an event A is the sum of the probabilities associated with the outcomes in A:

$$P(A) = \sum_{all\, s \in A} P(s).$$

A **continuous probability function** is defined for a process with an uncountably infinite number of outcomes. If S is a sample space with an uncountable number of outcomes and if f is a real valued function defined on S, then f is said to be a continuous probability function if **a)** 0 $\leq f(y)$ for all y $\in$ S, and **b)** $\int_s f(y)\,dy = 1$. Furthermore, if A is any event defined on S, it must be true that $P(A) = \int_A f(y)\,dy = 1.$

Probability density function **Larsen & Marx (2001, p. 121, 126)** **Describing the variation of a discrete random variable** Associated with each discrete random variable X is a **probability density function** (or **pdf**), $p_x(k)$. By definition $p_x(k)$ is the sum of all the probabilities associated with outcomes in sample space S that get mapped into *k* by the random variable X. That is

$$p_x(k) = P(s \in S \,|\, X(s) = k).$$

Conceptually, $p_x(k)$ describes the probability structure induced on the real line by the random variable *X*.

Describing the variation of a continuous random variable

Associated with each continuous random variable *Y* is also a **pdf**, $f_y(y)$, but $f_y(y)$ in this case is not the probability that the random variable *Y* takes on the value y. Rather, $f_y(y)$ is a function having the property that for all *a* and *b*,

$$P(a \le Y \le b) = P(s \in S \mid a \le Y(s) \le b) = \int_a^b f_y(y)\,dy.$$

Profile Analysis or, 'the multivariate approach to repeated measures' which does not require sphericity as an assumption. **Tabachnick & Fidell (2001, p 422)** describe profile analysis as an alternative to traditional repeated measures designs, 'a special application of multivariate analysis of variance (MANOVA) to a situation when there are several [response variables], all measured on the same scale.' Profile analysis requires more cases than dependent variables in the smallest group. Morrison (1976, p 153) describes profile analysis involving T^2 tests of parallel profiles followed by tests of different levels among groups. Profile analysis is available in SPSS MANOVA and SPSS GLM/Repeated measures as described by **Tabachnick & Fidell (2001, p 391)**.

Propagation of error Variables estimated from data usually have an associated error which should be incorporated in calculations involving those parameter estimates. For example, **Larsen & Marx (2001, p. 222-223)** present the propagation of error formula for the variance of **linear combinations**:

Calculating the variance of a linear combination. Theorem 3.13.1 Let *W* be any random variable, discrete or continuous, and let *a* and *b* be any two constants, Then

$$\mathrm{Var}(aW + b) = a^2\mathrm{Var}(W).$$

Calculating the variance of a sum of random variables. Theorem 3.13.2 Let W_1, W_2, ..., W_n be a set of independent random variables for which $E(W_i^2)$ is finite for all *i*. Then

$$\mathrm{Var}(W_1 + W_2 + \cdots + W_n) = \mathrm{Var}(W_1) + \mathrm{Var}(W_2) + \cdots + \mathrm{Var}(W_n).$$

The formulae and Monte Carlo approaches used to propagate error are covered well in **Bevington & Robinson (1992)** and **Taylor (1997)**.

Pseudoreplication A term coined by **Hurlbert (1984)** for a concept called model misspecification by **Underwood (1997)**. It specifically refers to the use of an inappropriate statistical model, especially one with inflated degrees of freedom used to estimate the error variance.

Q-mode, R-mode **Legendre & Legendre (1983,p. 172)**. The measurement of dependence between two descriptors (variables) is achieved my means of coefficients like Pearson's product-moment correlation, *r*. The study of the correlation or variance-covariance matrices is therefore called an **R analysis**. In contrast, a study of an ecological data matrix based upon the relationship between objects is called **Q analysis**. Cattell (1966) also defined O-,P-,S-, and T-modes. **n.b.,** many authors (**Pielou 1984**) reverse this conventional usage. Occasionally, the terms normal mode and inverse mode are used instead of Q and R mode, but these terms should be avoided due to the overlap with the corresponding statistical terms.

Quadratic equation Any equation of the form: $ax^2 + bx + c = 0$. "In mathematics, a **quadratic function** is a polynomial function of the form $f(x) = ax^2 + bx + c$,where a is nonzero. It takes its name from the Latin *quadratus* for square, because quadratic functions arise in the calculation of areas of squares. In the case where the domain and codomain are R (the real numbers), the graph of such a function is a parabola. If the quadratic function is set to be equal to zero, then the result is a quadratic equation." **http://en.wikipedia.org/wiki/Quadratic**

Quadratic term Any term raised to a power of 2.

Quantile http://mathworld.wolfram.com/Quantile.html

Quartile http://mathworld.wolfram.com/Quartile.html See also **Tukey hinges**

http://upload.wikimedia.org/wikipedia/commons/thumb/6/68/Adolphe_Qu%C3%A9telet_by_Joseph-Arnold_Demannez.jpg/225px-Adolphe_Qu%C3%A9telet_by_Joseph-Arnold_Demannez.jpg

Quetelet, Adolphe (1796-1874) From the **Columbia Encyclopedia**: Belgian statistician and astronomer. He was the first director (1828) of the Royal Observatory at Brussels. As supervisor of statistics for Belgium (from 1830), he developed many of the rules governing modern census taking and stimulated statistical activity in other countries. Applying statistics to social phenomena, he developed the concept of the "average man" and established the theoretical foundations for the use of statistics in social physics or, as it is now known, sociology. Thus, he is considered by many to be the founder of modern quantitative social science. A Treatise on Man (1835; tr., 1842) is his best-known work.

Quota sampling Gave rise to "Dewey beats Truman" *cf.*, **census**, probabilistic sampling

R squared [**coefficient of determination**, R^2] Percentage of the total response variation explained by the regression with the explanatory variables (**Ramsey & Schafer 1997, p. 213**; **Larsen & Marx 2006, p 309-310**)

$$R^2 = 100\left(\frac{Total\ sum\ of\ squares\ -\ Residual\ sum\ of\ squares}{Total\ sum\ of\ squares}\right)\%.$$

Adjusted R squared R^2 adjusted for the number of terms used to fit the model, *cf.*, **PRESS**

$$R^2_{adj} = 100\left(1 - \left(\frac{Error\ sum\ of\ squares\ /\ Error\ df}{Total\ sum\ of\ squares\ /\ Total\ df}\right)\right)\%.$$

Random effects In ANOVA, effects are modeled as fixed or random. The appropriate denominator for **F tests** in factorial ANOVAs differ depending on whether the main effects are fixed or random. A random effect model is one in which the levels are chosen as if they were random samples from a probability distribution. **McCulloch & Searle (2001, p. 17)** discuss whether an effect is fixed or random: "In endeavoriing to decide whether a set of effects if fixed or random, the context of the data, the manner in which they were gathered and the environment from which they came are the determining factors. In considering these points the important question is: are the levels of the factor going to be considered a random sample from a population of values which have a distribution? if "yes" then the effects are to be considered as random effects; if "no" then, in contrast to randomness, we think of the effects as fixed constants and so the effects are considered as **fixed effects**. Thus when inferences will be made about a distribution of effects from which those in the data are considered to be a random sample, the effects are considered as random; and when inferences are going to be confined to the effects in the model, the effects are considered fixed. Another way of putting it is to ask the questions: 'do the levels of a factor come from a probability distribution?' and 'Is there enough information about a factor to decide that the levels of it in the data are like a random sample?' Negative answers to these questions mean that one treats the factor as a fixed effects factor and estimates the effects of the levels; and treating the factor as fixed indicates a more limited scope of inference. On the other hand, affirmative answers mean treating the factor as a random effects factor and estimating the variance component due to that factor. In that case, when there is also interest in the realized values of those random effects that occur in the data, then one can use a prediction procedure for those values." *cf.*, **fixed effects**

Random variable

Rank sum test see **Wilcoxon rank sum test**

Rao-Blackwell theorem (**Hogg & Tanis 1977**, p. 404) *Let V and Y be two random variables such that V has mean E(V)=θ and positive finite variance. Let E(V|Y=y)=w(y). Then the random variable W=w(Y) is such that E(W)=θ and Var(W) ≤ Var(V).*
This theorem means that if a **sufficient statistic** for θ exists, say Y, we may limit our search for a minimum variance unbiased estimator to functions of Y.

Recurrent groups analysis A method to graphically display species associations, introduced by **Fager (1957)**

Reference level To analyze discretely distributed variables in a regression model, they are usually coded as 0,1 **dummy variables**. One of the levels must be left out, and the one level that is left out is called the reference level (see **Ramsey & Schafer 1997, p. 237**)

Regression A term coined by **Francis Galton** in **1879** and **1886** to explain the bivariate association between filial and parental heights. **Yule (1897)** was the first to use least squares to fit a regression of Y on X by minimizing the squared residuals between the regression line and Y. The term regression was later applied to the entire field of fitting linear models with least-squares methods. **Ramsey & Schafer (1997)** state "**regression** refers to the mean of a response variable as a function of an explanatory variable. A **regression model** is a function used to describe the regression. The **simple linear regression** model is a particular regression in which the regression is a straight-line function of a single explanatory variable."
These least squares methods used in regression can be traced back at least to **Legendre (1805)** and the **normal equations** to **Gauss (1822)**.

The **regression phenomenon** – also called **regression to mediocrity**, **regression to the mean**, **the regression artifact** – is expressed mathematically as (**Stigler 1999, p. 176**):

$$E(Y|X{=}x) \;=\; \rho x,\;\; E(X|Y{=}y) \;=\; \rho y,\; and \qquad\qquad (83)$$
$$Var(Y|X{=}x) \;=\; Var(X|Y{=}y) \;=\; (1-\rho^2)\;\sigma^2.$$

Galton was the discoverer of **regression to the mean**, which **Stigler (1999, p. 6)** regards as one of the most original in the last two centuries:

> *" ... **regression to the mean**, one of the trickiest concepts in all of statistics. Galton's completion of his discovery of this phenomenon in the 1880's should rank with the greatest individual events in the history of science — at a level with William Harvey's discovery of the circulation of blood and with Isaac Newton's of the separation of light. In all three cases the discovery is apparently of such an elementary character that it could have been made at least a thousand years earlier, but the fact that it wasn't and the problems the discoverer had in communicating it convincingly to the world hint at the profound difficulty involved. In all three cases the consequences were immense and far-reaching."*

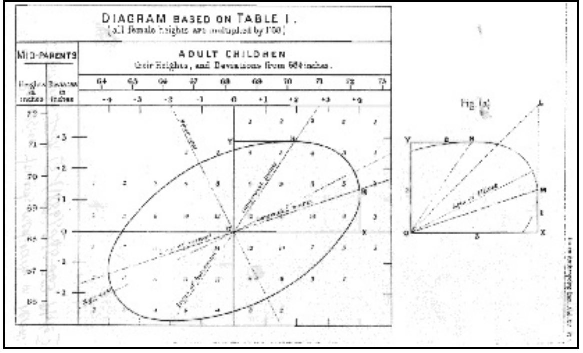


Figure 9. The regression ellipse from p. 248 in Galton (1886), posted at the UCLA statistics history site:

http://www.stat.ucla.edu/history/regression_eclipse.gif

Regression to the mean is described in William Trochim's database: **http://trochim.human.cornell.edu/kb/regrmean.htm**

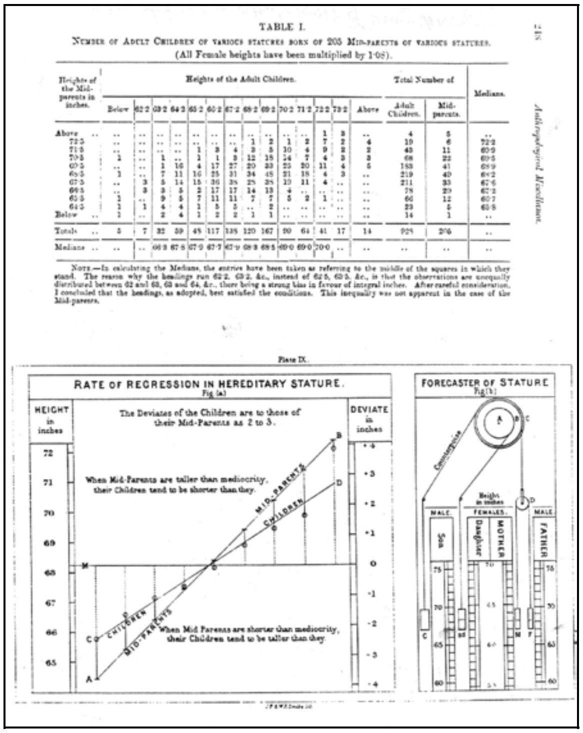


TABLE I.

NUMBER OF ADULT CHILDREN OF VARIOUS STATURES BORN OF 205 MID-PARENTS OF VARIOUS STATURES.
(All Female heights have been multiplied by 1·08).

Heights of the Mid-parents in inches.	Heights of the Adult Children.														Total Number of		Medians.
	Below	62·2	63·2	64·2	65·2	66·2	67·2	68·2	69·2	70·2	71·2	72·2	73·2	Above	Adult Children.	Mid-parents.	
Above ..	..	..	..	..	..	..	..	..	..	..	..	1	3	..	4	5	..
72·5	..	..	..	..	..	..	..	1	2	1	2	7	2	4	19	6	72·2
71·5	..	..	..	..	1	3	4	3	5	10	4	9	2	2	43	11	69·9
70·5	1	..	1	..	1	1	3	12	18	14	7	4	3	3	68	22	69·5
69·5	..	..	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68·9
68·5	1	..	7	11	16	25	31	34	48	21	18	4	3	..	219	49	68·2
67·5	..	3	5	14	15	36	38	28	38	19	11	4	..	..	211	33	67·6
66·5	..	3	3	5	2	17	17	14	13	4	..	..	..	..	78	20	67·2
65·5	1	..	9	5	7	11	11	7	7	5	2	1	..	..	66	12	66·7
64·5	1	1	4	4	1	5	5	..	2	..	..	..	..	..	23	5	65·8
Below ..	1	..	2	4	1	2	2	1	1	..	..	..	..	..	14	1	..
Totals ..	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	..
Medians ..	..	..	66·3	67·8	67·9	67·7	67·9	68·3	68·5	69·0	69·0	70·0	..	..	..	..	..

NOTE.—In calculating the Medians, the entries have been taken as referring to the middle of the squares in which they stand. The reason why the headings run 62·2, 63·2, &c., instead of 62·5, 63·5, &c., is that the observations are unequally distributed between 62 and 63, 63 and 64, &c., there being a strong bias in favour of integral inches. After careful consideration, I concluded that the headings, as adopted, best satisfied the conditions. This inequality was not apparent in the case of the Mid-parents.

248
Anthropological Miscellanea.
Plate IX.
RATE OF REGRESSION IN HEREDITARY STATURE.
Fig (a)
HEIGHT in inches
The Deviates of the Children are to those of their Mid-Parents as 2 to 3.
DEVIATE in inches
When Mid-Parents are taller than mediocrity, their Children tend to be shorter than they.
MID-PARENTS
CHILDREN
When Mid Parents are shorter than mediocrity, their Children tend to be taller than they.
FORECASTER OF STATURE
Fig (b)
MALE
FEMALES
Son
Daughter
MOTHER
FATHER

Figure 10. Plate X from Galton (1886), posted at http://www.stat.ucla.edu/history/regression.gif

OLS Regression Fitting data using **o**rdinary **l**east **s**quares. The assumptions that matter are that there are no outliers which significantly affect the regression fits or statistics, detected by **Cook's D** for example. You don't want to see pattern in the plot of residuals vs. predicted values, nor should there be pattern between the residuals and the order in which samples were taken (in space or time). The former problem could indicate the need for transformation or for a higher order regression, and the latter could indicate lack of independence among the errors.

Regression: Model II Model II regression is called for when both the X and Y variables are measured with considerable error. **Legendre & Legendre (1998)** provide a thorough discussion of methods for Model II regression, including principal components regression.

Regression diagnostics See Jerry Dallal's page: **http://www.tufts.edu/~gdallal/diagnose.htm** for descriptions of

Cook's distance differences between the predicted responses from the model constructed from all of the data and the predicted responses from the model constructed by setting the i-th observatio aside.

DFITS$_i$ scaled difference between the predicted responses from the model constructed from all of the data and the predicted responses from the model constructed by setting the i-th observation aside

DFBeta$_i$ when the i-th observation is included or exlcuded, DFBETAS looks at the change in each regression coefficient.

See also: **studentized residuals**

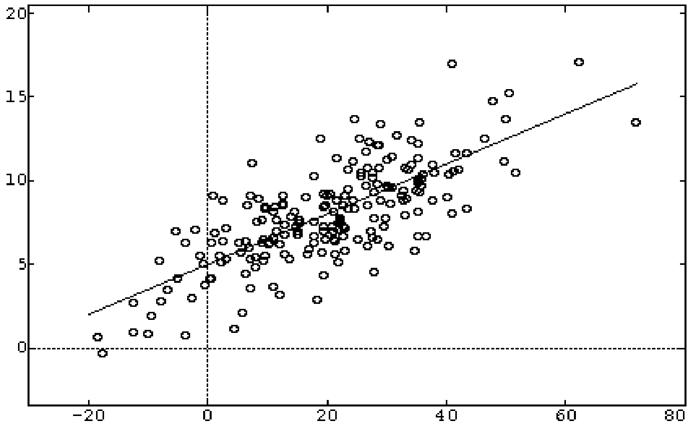


Figure 11.

http://en.wikipedia.org/wiki/File:Linear_regression.png

Relative power efficiency

Relative risk see **risk ratio**

Repeated measures design When 2 or more variables are measured from the same experimental units (often subjects or patients in drug trials). A **paired t test** is a repeated measures design (actually a repeated measures ANOVA with two levels of 1 'within subjects' factor and no 'between subjects' factors).

Residuals Observed minus predicted value

PRESS residuals Prediction residual error sum of squares. Residuals obtained from regression coefficients derived after the effect of each case is removed.

$$e_{(i)} = \frac{e_i}{(1-h_{ii})}$$

$$where,\ h_{ii}\ is\ the\ leverage.$$

Studentized residual, raw residual for a case standardized or ‘studentized’ by scaling with variable standard error after that case is deleted. This can be done by adjusting the mean square error for a regression by that case’s **leverage**

$$r_i = \frac{e_i}{\sqrt{MS_E(1-h_{ii})}}$$
$$where,\ h_{ii}\ is\ the\ leverage.$$

Response variable The variable that is being modeled in a **regression model**. Sometimes called the **dependent variable**. *cf.*, **explanatory variable**

Retrospective studies. **Ramsey & Schafer 1997 p. 529**

Ridge regression Hoerl & Kennard (1970), quoted in **Kendall & Stuart 1979, p. 92** Ridge regression is one method for coping with colinearity among explanatory variables. If X is the matrix of explanatory variables, then the normal equations are solved by adding a small constant to the sum of squares and croxx products matrix before inversion with inv(X’X + lambda*I)) instead of inv(X’X). The effect of adding small amounts to the main diagonal is assessed with a ridge-trace plot. There is an SPSS macro available from Raynald Lavesque to carry out ridge regression.

Risk ratio or **Relative Risk** “The risk ratio is a ratio of probabilities, which are themselves ratios. The numerator of a probability is the number of cases with the outcome, and the denominator is the total number of cases. The risk ratio lends itself to direct intuitive interpretation. For example, if the risk ratio equals X, then the outcome is X-fold more likely to occur in the group with the factor compared with group lacking the factor.” **Holcomb et al. 2001**. **Zhang & Yu (1978)** show the relation between risk ratio and **odds ratio**:

$$Risk\ Ratio = \frac{OR}{[(1 - P_o) + (P_o \times OR)]},\ where$$
$$OR = Odds\ Ratio$$
$$P_o = proportion\ nonexposed\ individuals\ who\ experience\ outcome.$$

As shown by **Zhang & Yu (1978, Fig 1)** the odds ratio overestimates the risk ratio if the event is common. See also **http://www.childrens-mercy.org/stats/journal/oddsratio.asp**

Image removed due to copyright restriction

Robust [**robustness**] P values remain accurate despite slight violations of assumptions.

ROC curve Receiver Operating Characteristic curve. A curve from signal detection theory which describes the classification of a signal in the presence of noise (**Hosmer & Lemeshow 2000, p. 160 & Figure 5.2**) It is used extensively in evaluating diagnostic tests, such as screening tests for cancer. See also **http://www.anaesthetist.com/mnm/stats/roc/** , **specificity**, **sensitivity**

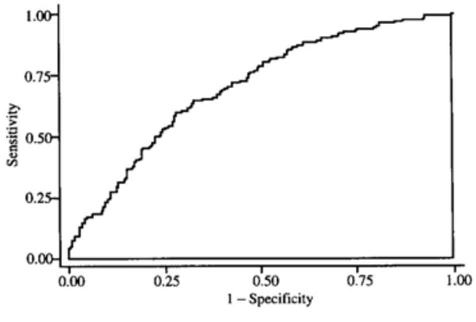


Figure 5.2 Plot of sensitivity versus 1–specificity for all possible cutpoints in the UIS. The resulting curve is called the ROC Curve.

Figure 13. ROC curve from Hosmer & Lemeshow

Rotation **Legendre & Legendre (1983, p. 309)**. Transformations of the axes used to portray data. Both orthogonal (rigid rotations preserving Euclidean distances among data [see VARIMAX]) and oblique rotations have been used in FA. The purpose of rotations is not to improve the degree of fit between the observed data and the factors...the purpose is to achieve **simple structure**.

Runs A consecutive series of events. +++00 represents two runs and ++00+ represents three. There are a variety of runs tests available, several are described in **Larsen & Marx (2001)**

Sample

SAR Simultaneous autoregressive model *cf.*, **CAR**

SARIMA seasonal autoregressive integrated moving-average, *cf.*, **ARIMA**

Sample outcome "Each of the potential eventualities of an **experiment** is referred to as a **sample outcome**, *s*, and their totality is referred to as a **sample space**, *S*. To signify the **membership** of *s* in *S*, we write s ∊ S. Any designated collection of **sample outcomes**, including individual outcomes, the entire **sample space**, and the null set, constitutes and event. The latter is said to occur if the outcome of the **experiment** is one of the members of the event." (**Larsen & Marx 2001, p. 21-22**)

Satterthwaite approximation Satterthwaite's (1946) approximation of the df for the **Welch's t test** *cf.*, **Behrens-Fisher problem**

$$df = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)}.$$

Scheffe's test A very conservative **multiple comparison test** designed to produce an alpha level appropriate for testing linear combinations of the data (*e.g.*, group A+B *vs.* Group C+D+E).

Scree test (described by **Kim & Mueller, 1978**, p. 44) **Cattell (1966)** described the elbow on the log (**eigenvalue**) vs. dimension plot is the point beyond which we are in the "factorial litter" or scree (where scree is the geological term for the debris which collects on the lower part of a rocky slope). The **scree test** retains only those factors at or to the left of the elbow for interpretation or further rotation. **Jackson (1993)** reviewed various stopping rules for PCA, concluding that the scree test overestimated the "true" number of factors by 1.

SEM Structural equation modeling. From a 4/5/04 post from H. Rubin on sci.stat.edu "The formal idea of structural equations modeling, as far as I know, originated in biology in 1919 by **Sewall Wright** [This would be Wright's path analysis]. The idea is that one does not simply have regressions with independent and dependent variables, but a structure is used to describe the probability model for all dependent variables. It was used in psychometrics without formal realization even earlier in multiple factor analysis, and it was heavily used in econometrics as soon as it was realized that regression led to bias. I do not know when the name "Structural Equation Modeling" was formally introduced, but it was clearly understood in the mathematical economics of the 1940s."

Sensitivity In a diagnostic test and **ROC curve**, the **sensitivity** is defined as the proportion of cases (e.g., patients with prostate cancer) with a test value (*e.g.*, PSA antigen level) exceeding a given cutoff value. The **specificity** is the proportion of noncases (cancer-free patients) with a test value (PSA antigen level) equal to or below the cutoff value. The **false-positive rate** is (1-specificity) (see **Thompson *et al.* 2005**) see **ROC curve**

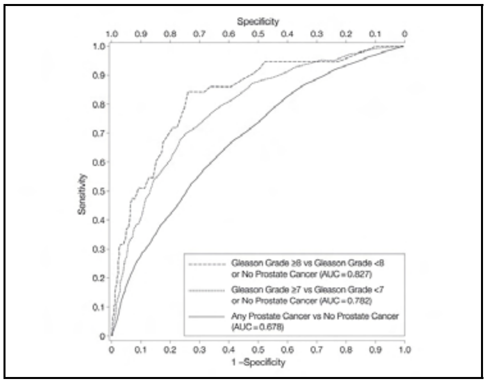


Figure 14. ROC curve for PSA antigen test from Thompson et al. (2005). 1-specificity is the false positive rate.

Shrinkage From **Harrell (2002)**: There are two related meanings. First in regression, when one data set is used for calibration & prediction, the slope will be 1 (by definition). When however parameter estimates are derived from one dataset and applied to another, **overfitting** will cause the calibration plot to have a slope less than 1, a result of **regression to the mean**. Typically low predictions will be too low & high predictions too high. Second, **shrinkage** can refer to pre-shrinking regression plots so that the calibration plot will be more accurate with future data.

Simple structure **Thurstone's (1947)** term for a factor solution with certain properties: Each variable should have factor loadings on as few common factors as possible, and each common factor should have significant loadings on some variables and no loadings on others. (**Kim & Mueller, 1978, p. 86**)

"The principle of ***simple structure****: Once a set of k factors has been found that account for intercorrelations of the variables, these may be transformed to any other set of k factors that account equally well for the correlations....* ***Thurstone (1947)*** *put forward the idea that only those factors for which the variables have a very simple representation are meaningful, which is to say that the matrix of factor loadings should have as many zero elements as possible...a variable should not depend on all common factors but only on a small part of them. Moreover, the same factor should only be connected with a small portion of the variables. Such a matrix is considered to yield* ***simple structure.*** *"*

Reyment & Jöreskog (1993, p. 87)

SIMPLS Algorithm "An alternative estimation method for partial least squares regression components is the SIMPLS algorithm (de Jong, 1993)" **http://www.statsoft.com/textbook/stathome.html**

Serial correlation Regression analysis typically assumes that observational errors are pairwise uncorrelated. In serial correlation, there is a correlation between errors a fixed number of steps apart (**Draper & Smith 1998, p. 179**). Certain types of serial correlation are tested with the **Durbin-Watson test**.

Simple random sampling (SRS): A simple random sample of size n from a population is a subset of the population consisting of n members selected in such a way that every subset of size n is afforded the same chance of being selected. **Ramsey & Schafer (1997)**

Simpson's diversity

Identical to -1 + Hurlbert's **E(S_n)** at n=2 (**Smith & Grassle 1977**)

Simpson (1949) from **Pielou (1969)**:

"Suppose two individuals are drawn at random and without replacement from an S-species collection containing N individuals of which N_j belong to the jth species (j=1:s); Σ_j N_j=N). If the probability is great that both individuals belong to the same species, we can say that the diversity of the collection is low. This probability is $\sum_{j=1}^{S} \frac{N_j(N_j-1)}{N(N-1)}$, and so we may use:

$$Simpson's\ D = 1 - \sum_{j=1}^{S} \frac{N_j\ (N_j - 1)}{N\ (N - 1)}.$$

This assumes a random sample of a population. The biased form of Simpson's diversity is:

$$Simpson's\ biased\ D = 1 - \sum_{j=1}^{S} \left(\frac{N_j}{N} \right)^2.$$

Advantages and properties:

- For m=2, **E(S_n)** = Simpson's unbiased diversity index (**Smith & Grassle 1977**) Simpson's index is an **unbiased estimator**.

Problems:

- ignores species occurring only once. Pays little attention to rare species.
- cannot be decomposed into hierarchical diversity.

Simpson's paradox P{A|C} < P{B|C} and P{A|~C}< P{B|~C} but P(A) > P(B). **Agresti (1996)** presents the example of capital punishment in Florida in which the percentage of black capital defendants being given the death penalty (A) is lower than the percentage of white capital case defendants being given the death penalty (B). But, when the cases are partitioned into those with a white victim and those with a black victim, the percentage of blacks given the death penalty is higher than whites. *cf.* **ecological fallacy**, **http://plato.stanford.edu/entries/paradox-simpson/** and **http://www.cawtech.freeserve.co.uk/simpsons.2.html**

Singular value decomposition (**SVD**) All matrices can be decomposed as the product of three component matrices: **U*S*V'**. **S** is a diagonal matrix of singular values (*i.e.*, the off-diagonal elements are 0s). The columns of U and the rows of V are orthogonal. The ratio of the largest to the smallest singular value is the **condition number** of the matrix. Ill conditioned matrices have large condition numbers (usually in the thousands) and are said to be not of full rank. SVD is the method of choice for creating the powers of matrices. If P is a matrix and Q*D*R' is the singular value decomposition of P, then $Q*D^{10}*R' = P^{10}$. The k'th power of the diagonal matrix D is computed by raising each diagonal element to the kth power. The best low dimensional display of a matrix in a least squares sense is created using the SVD. This is the basis of **Eckart & Young's (1936)** theorem. See: **http://mathworld.wolfram.com/SingularValueDecomposition.html**

Sink species Where individuals of a species use a habitat where their carrying capacity is less than zero, that species is a sink species (**Rosenzweig 1995, p. 260**, **Pulliam 1988**)

Skewness the skewness of a pdf is the 3rd moment about the mean. A symmetric pdf has a skewness of 0. Lognormal pdfs are skewed to the right **http://mathworld.wolfram.com/Skewness.html** *cf.,* **kurtosis**

Snedecor, George W 1882-1974. Described the F distribution and named it for Fisher. Author of a famous statistics textbook (with Cochran).

Somers' D(C |R) and D (R |C) Somers' D(C|R) and Somers' D(R|C) are asymmetric modifications of **Kendall's tau-b**. C|R denotes that the row variable X is regarded as an independent variable, while the column variable Y is regarded as dependent. **http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap28/sect20.htm**

Spearman's ρ A distribution-free, rank-based correlation coefficient. Equivalent to **Pearson's r** after data converted to ranks. *Cf.*, **Kendall's τ**

Specificity See also **ROC curve**

Sphericity A form of variance-covariance matrix assumed by repeated measures ANOVA Departures from sphericity are assessed using $\hat{\varepsilon}$, and F test statistics are assessed with numerator and denominator **df** adjusted in direct proportion to $\hat{\varepsilon}$. Huyhn-Feldt and Greenhouse-Geisser are two adjustments of df for departures from sphericity, with the latter being judged conservative. Prof. William Ware provided this post on the **sphericity assumption** to sci.stat.edu (8/26/96) "Said assumption is relevant in "within subject" designs, either randomized block or repeated measures. Most statistical procedures assume that the errors are independent. In "independent groups" designs, this reduces to no association between the observations in the groups. But of course, in "dependent" samples designs, it is the correlations among the observations that we are employing to reduce the error terms... However, if the correlations are assumed to arise from the "subject" effects, then it implies that all of the pair-wise **covariances** between treatments should be equal to one common value. Thus, the assumption of **sphericity** is met is the variance/covariance matrix is consistent with the data having been drawn from a population in which all of the variances are equal one another, and all of the covariances are equal to one another. If you have multiple groups, then the "group" variance/covariance matrices are tested for equality prior to pooling them. The pooled matrix is then tested for sphericity." See Sphericity and Compound Symmetry in the

ANOVA/MANOVA chapter at **http://name.math.univ-rennes1.fr/bernard.delyon/textbook/stathome.html**

Spearman's ρ The Pearson product moment correlation coefficient after the data have been converted to ranks *cf.*, **Kendall's τ**.

Split-plot design see **ANOVA, split plot**

Spurious correlation A term introduced by **Pearson (1897)** as noted by **Schlager *et al.* (1998, p. 548)**:

> In a classical paper, **Pearson (1897)** pointed to a particular property of compound varialbes, such as ratios, in correlation. He showed that two variables that have no correlation between themselves become correlated when divided by a third uncorrelated variable. **Pearson (1897)** introduced the term '**spurious correlation**' for the 'amount of correlation which would still exist between the indices, were the variables on which they depend distributed at random'

Another definition from the web: "A situation in which measures of two or more variables are statistically related (they cover) but are not in fact causally linked—usually because the statistical relation is caused by a third variable. When the effects of the third variable are taken into account, the relationship between the first and second variable disappears." [**http://www.autobox.com/spur2.html**] [*c.f.* **nonsense correlation**]

SSCP the sum-of-squares-and-cross-products matrix. The SSCP matrix for sites is formed by premultiplying a site x variable matrix times by its transpose. The (i,i)th element of the symmetric SSCP matrix is the sum of squares for the ith variable across sites. The (h,i)th element is the sum of cross-products of the h^th and i^th variables.

Standard deviation The typical distance between a single number and the set's average (**Ramsey & Schafer 2002**); the square root of the **variance**

$$\sqrt{\left(\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{(n-1)}\right)}. \qquad \textbf{(93)}$$

for a proportion:

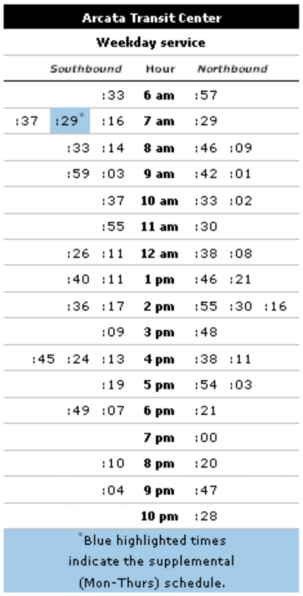
Standard error and **coefficient of variation** $\hat{\sigma}_{\mu} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}}$. The standard error of the sample mean is the sample standard deviation divided by the square root of sample size. With the finite population correction, (1 - n/N), with n being the sample size and N the population size, the **standard error of the sample mean** and the **coefficient of variation of the sample mean** are:

$$S\hat{E}(\bar{Y}) = \sqrt{s^2/n * (1 - \frac{n}{N})}$$

$$C\hat{V}(\bar{Y}) = \frac{S\hat{E}(\bar{Y})}{\bar{Y}}.$$

Statistics **Statistics** is defined by **Sokal & Rohlf (1995)** as the scientific study of numerical data based on variation in nature. **Statistics** can be used in another valid sense as the plural of the noun **statistic**, any quantity that can be calculated from observed data (*e.g.*, the sample standard deviation s *or* $\hat{\sigma}$). Observations are used in the calculation of **sample statistics** (*e.g.*, the sample mean $\bar{x}$ and standard deviation s *or* $\hat{\sigma}$, which are estimates of **population statistics** or **parameters** (μ, σ)). Statistics are usually represented by Roman letters, whereas parameters are represented by Greek letters (**Ramsey & Schafer 1997, p. 20**) *cf.*, **parameter**

Stem-and-leaf plot A quick graphical display method invented by John Tukey. See **http://mathworld.wolfram.com/Stem-and-LeafDiagram.html**



Arcata Transit Center						
Weekday service						
Southbound			Hour	Northbound		
		:33	**6 am**	:57		
:37	:29*	:16	**7 am**	:29		
	:33	:14	**8 am**	:46	:09	
	:59	:03	**9 am**	:42	:01	
		:37	**10 am**	:33	:02	
		:55	**11 am**	:30		
	:26	:11	**12 am**	:38	:08	
	:40	:11	**1 pm**	:46	:21	
	:36	:17	**2 pm**	:55	:30	:16
		:09	**3 pm**	:48		
:45	:24	:13	**4 pm**	:38	:11	
		:19	**5 pm**	:54	:03	
	:49	:07	**6 pm**	:21		
			7 pm	:00		
		:10	**8 pm**	:20		
		:04	**9 pm**	:47		
			10 pm	:28		

*Blue highlighted times indicate the supplemental (Mon-Thurs) schedule.

Figure 15. **Bus Schedule Stem Leaf Diagram (http://wikieducator.org/File:BusSchedStemLeaf.png)**

Stigler's law of eponymy "No scientific discovery is ever named after its original discoverer." **Stigler (1999, p. 277)**

Stochastic variable

Stopping rules There are two meanings for stopping rules. **Jackson (1993)** reviews tests used to decide how many dimensions to retain in a factor analysis or a PCA before rotation. Stopping rules also play a role in sequential medical trials (**Armitage 1975**), in which an investigator performs statistical tests on a small number of subjects, and then sequentially adds subjects if the initial test was deemed inadequate to distinguish between null and alternate hypotheses. In frequentist statistics, an adjustment to experiment-wise error rates must be made to take into account the nature of the stopping rule employed. **Mayo (1996)** strongly criticizes Bayesian statistical inference for its inability to account for stopping rules.

Structural equation modeling (SEM) A technique to create models to explain patterns of covariation among variables. The parameters of an SEM are usually fit by the method of **maximum likelihood** *cf.*, **factor analysis**, **path analysis**, **regression**

Student's t distribution A distribution that is similar to the normal distribution but accounts for the increased dispersion caused by having to estimate the standard deviation from the sample. Developed by **William Gosset**, who published under the *nom de plum* of Student

Student's t test A test for the difference between two means when the variances are unknown and must be estimated from the data. These variances are assumed to be equal in estimating the pooled variance. There are two forms of test: the independent samples t t test and the **paired t test**, for paired data. The probability that the observed results are compatible with a null hypothesis of no difference is assessed using **Student's t distribution**.

The problem of performing a test of mean difference with unequal variance is called the **Fisher-Behrens problem** and the **Welch's t test** was developed as a replacement for the independent samples t test for that purpose.

Sum of Squares Type I, II, III & IV SS as used in SPSS GLM are defined at (login as guest with password guest):
http://www.spss.com/tech/stat/algorithms/7.5/ap11smsq.pdf

Suppressor variable "In the two-predictor situation … traditional and negative suppressors increase the predictive value of a standard predictor beyond that suggested by the predictor's zero order validity." **Conger (1974)**

Survey design [**Sample survey design**] A survey is an observational study of a finite statistical population, not an **experiment**. **Hurlbert (1984)** referred to one type of survey design, measuring a response variable at different levels of a covariate, as a **mensurative experiment**. As A survey design describes the goals of the survey, usually to estimate population parameters, determines the number and manner of sampling the population or populations of interest. Survey design involves the allocation of sampling units, such as quadrat samples, with locations determined by systematic, or random sampling. Transect sampling is one form of systematic sampling, often including a random component, such as random directions or starting points for the transect or random positions along the transect. Cluster sampling involves sampling groups of individuals, sometimes by choice but usually by necessity. For example a quadrat sample of area provides a cluster sample of individuals within that area. Often the statistical population is divided into strata to allow more precise estimates of population parameters for a given sampling effort. As noted by **Hayek & Buzas (1996)**, data from survey designs involving cluster sampling or systematic sampling can't be pooled to estimate means and variances as if the observations were simple random samples; often the variance of cluster or transect samples differ from those calculated assuming simple random sampling. SAS has new procedures that will incorporate survey designs in the estimation of population parameters: **http://support.sas.com/rnd/app/papers/survey.pdf** *Cf.*, **Kendall & Stuart distinction between experiment & survey**

t-test **Student's t test**

test statistic **Hogg & Tanis (1977, p. 255)** A statistic used to define the **critical region** is called a **test statistic**. The **critical region** C is often defined as a set of values of the test statistic that leads to the rejection of the **null hypothesis H_o.**

Time-series analysis In modeling time series through regression, the independence assumption of ordinary least squares regression is often violated. There is positive serial correlation in regression residuals, with nearby points in time being more similar than expected by chance. This lack of independence due to positive temporal serial correlation, also called positive temporal autocorrelation, is that the standard errors of the estimates are too small. Tests based on these standard errors will have inflated Type I errors relative to nominal errors. There are two major solutions: adjusting the standard error to account for the serial correlation or to use filtering to adjust both the response and explanatory variables in regression. Most time-series analysis packages will have routines to fit time series, adjusting for autocorrelation, using maximum likelihood extimation.

Tobit Analysis Tobit analysis is a form of **generalized linear modeling**, appropriate for censored data, e.g., data containing a large number of zeros. Tobit modeling will fit the non-zero data.

Two-sided test also called **two-tailed test** *cf.*, **one-sided test**

Tukey-Kramer test An **a posteriori test** based on the studentized range statistic. It is an extension of **Tukey's HSD**, or honestly significant difference, for unequal sample sizes.

Type I error The error made when rejecting a true **null hypothesis**. The probability of Type I error, called the **alpha level** or **significance level** of the test, is set in advance in the **Neyman-Pearson school** of statistical inference.

Type II error The error made when accepting a false **null hypothesis**. The probability of Type II error is called β and 1-β is the **power** of the test.

Union See **intersection**

Uniqueness the extent to which the common factors fail to account for the total variance of a variable.

Variable From **Mathworld**: "A variable is a symbol on whose value a function, polynomial, etc., depends. For example, the variables in the function f(x,y) are x and y. A function having a single variable is said to be univariate, one having two variables is said to be bivariate, and one having two or more variables is said to be multivariate. In a polynomial, the variables correspond to the base symbols themselves stripped of coefficients and any powers or products."

Variance a measure of the dispersion of a variable; defined as the sum of squared deviations from the mean divided by the number of cases or entities. *cf.*, **standard deviation**

$$s^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{(n-1)}. \tag{99}$$

Variance inflation factor A diagnostic test for **multicollinearity** in multiple regression

Venn diagrams A graphical display, usually consisting of circles on a square background which are used to display the intersection, union and complement of events.

Vertex (vertices) The points or nodes in a **graph**. Vertices may be connected with **edges**.

Wald statistic **Agresti (1996, p. 88)**: any statistic that divides a parameter by its standard error and squares it is called a Wald statistic. In generalized linear models, parameter estimates z= $\hat{\beta}$/Standard error are evaluated with the standard normal distribution or equivalently z^2 has a chi-square distribution with df=1; the p value is the right-tail distribution of the chi-square distribution.

WA-PLS Weighted Average Partial Least Squares see ter Braak & Juggins

Weibull distribution

Welch's t test An modification of **Student's t test** to test for differences in means with samples drawn from populations with different variances. The degrees of freedom for the test are adjusted by the **Sattertherwaite approximation** *cf.*, **Behrens-Fisher problem**.**Fligner-Policello test**

http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap67/sect16.htm

$$t' = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}. \qquad \textbf{(101)}$$

Wilcoxon rank sum test A test for difference in location or central tendency between two samples. The nonparametric equivalent of the independent samples **Student's t test**. The test assumes that the samples are drawn from distributions with equal spread, equivalent to the equal variance assumption of **Student's t test**. The **p values** are identical to those calculated from the **Mann-Whitney U test**, and the Mann-Whitney U and Wilcoxon's rank sum test statistics can be converted exactly. The **Fligner-Policello test** is a rank-based equivalent for samples with unequal spread, but it probably is only appropriate for large sample sizes. **Salsburg (2001)** profiles Wilcoxon, a chemical engineer.

Wilcoxon signed rank test The nonparametric equivalent of the paired-samples Student's t test.

WLS A Weighted Least Squares is a **generalized least squares regression** in which the equal variance assumption is relaxed. **Draper & Smith (1998)** is an excellent reference on WLS regression, which can be peformed readily with SPSS.

Wright, Sewall Founder of quantitative population genetics with **Fisher** and J. B. S. Haldane. Invented **path analysis** early in his career and used this method to analyze patterns of inheritance. His major contribution was describing genetic drift and his shifting balance model of evolution. *Cf.*, **SEM** See **http://books.nap.edu/books/0309049784/html/438.html**

Yule, George Udny (1871-1951) Student of Pearson. **Yule (1897)** adapted Gauss's normal equation approach to estimate the slope of a regression line **(Stigler, 1986, p. 350)**. Our modern approach of estimating the slope and y-intercept of a least-squares regression can be traced to **Yule (1897)**. Yule coined the term **nonsense correlation**.

z transform The standard normal distribution is often called the z distribution. The z transform — subtract the mean and divide by the standard deviation — produces a transformed variate with zero mean and unit standard deviation. The **z-score** is the cut point of the standard normal distribution (e.g., a z-score of -1.96 corresponds to p =0.025 on the cumulative normal probability distribution, z-score =0.5 corresponds to p=0.5 on the cumulative normal probability distribution and z-score =1.96 corresponds to p=0.975 on the cumulative normal probability distribution.

References

Abramowitz, M. and I. A. Stegun, Eds. 1965. Handbook of mathematical functions with formulas, graphs, and mathematical tables. Dover Publications, New York. 1045 pp. [25]

Agresti, A. 1996. An introduction to categorical data analysis. Wiley, New York. [**15**, **31**, **41**, **51**, **56**]

Armitage, 1975. Sequential medical trials, 2nd edition. John Wiley & Sons, New York. [**54**]

Bell, E. T. 1937. Men of mathematics. Simon & Schuster, New York.[**22**, **39**]

Boesch, D. F. 1977. Application of numerical classification in ecological investigation of water pollution. Environmental Protection Agency, Ecological Research Series EPA-600/3-77-033. Corvallis, Oregon. 115 pp.[?]

Box, G. E. P and D. R. Cox. 1964. An analysis of transformations. J. Roy. Statist. Soc. B-26, 211-243, discussion 244-252. As cited in **Draper & Smith (1998)** [**8**]

Campbell, D. T. and D. A. Kenny. 1999. A primer on regression artifacts. The Guilford Press, New York. [**10**]

Cochran, W. G. and G. M. Cox. 1957. Experimental designs. John Wiley & Sons, New York. 611 p and tables. [**5**, **14**]

Cohen, J. et al. 2003. Applied multiple regression/correlation analysis for the behavioral sciences, third edition. Lawrence Erlbaum Associates, Mahwah, NJ. [**12**]

Conger, A. J. 1974. A revised definition for suppressor variables: a guide to their identification and interpretation. Educational and Psychological Measurement *34*: 35-46. [?]

Draper, N. R. and H. Smith. 1998. Applied Regression Analysis, 3rd Edition. John Wiley & Sons, New York. 706 p, with data diskette. [**8**, **19**, **28**, **34**, **51**, **57**, **58**]

Eckart, C. and G. Young. 1936. The approximation of one matrix by another of lower rank. Psychometrika 1: 211-218. [**52**]

Fager, E. W. 1957. Determination and analysis of recurrent groups. Ecology *38*: 586-595. [**45**]

Freedman, D, R. Pisani and R. Purves. 1998. Statistics, 3rd edition. Norton, New York. *[This is a wonderful introduction to probability and statistics. It is very elementary though, and the authors' avoidance of any equations really limits the book's usefulness]*[**47**]

Galton, F. 1877. Typical laws of heredity. Nature 15: 492-495. [?]

Galton, F. 1886. Family likeness in stature. Proc. Roy. Soc. London 40: 42-73. [**45**]

Galton, F. 1888. Co-relations and their measurement, chiefly from anthropological data. Proc. Roy. Soc. London. 45: 133-145. [**12**]

Gauss, C. F. 1822. Awendung der Warhsceinlichkeitsrecnung auf eine Aufgabe der practischen Geometrie. Astronomische Nacricten , vol. 1/6, cols 81-86. [Full citation in **Stigler 1999, p. 445**] [**36**, **45**]

Golumbic, M. C. 1980. Algorithmic graph theory and perfect graphs. Academic Press, New York. [**36**, **45**]

Gondran, M. and M. Minoux. 1984. Graphs and algorithms. John Wiley and Sons, New York. [**18**]

Gower, J. C. 1966. Some distance properties of latent root vector methods used in multivariate analysis. Biometrika *53*: 325-338. [?]

Greenacre, M. 1984. Theory and Application of correspondence analysis. Academic Press, Orlando.[**12**]

Harman, H. H. 1967. Modern Factor Analysis. Univ. Chicago Press, Chicago & London. 474 pp. [**19**]

Hogg, R. V. and E. A. Tanis. 1977. Probability and statistical inference. MacMillan publishing, New York. 450 pp. [**7**, **10**, **14**, **18**, **19**, **26**, **29**, **40**, **45**, **55**]

Holcomb, W. L., T. Chaiworapongsa, D. A. Luke and K. D. Burgdorf. 2001. An Odd Measure of Risk: Use and Misuse of the Odds Ratio. Obstetrics & Gynecology 2001;98:685-688. [**48**]

Hosmer, D. W. and S. Lemeshow. 2000. Applied logistic regression, 2nd Edition. John Wiley & Sons, New York. 373 pp. [**49**]

Hurlbert, S. M. 1971. The non-concept of species diversity: a critique and alternative parameters. Ecology *52*: 577-586. [**17**, **25**]

Jackson, D. A. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology *74*: 2204-2214. [**49**]

Jardine, N. and R. Sibson. 1968. The construction of hierarchic and nonhierarchic classifications. Computer J. *11*: 177-184. [?]

Kemeny, J. G. and J. L. Snell. 1976. Finite Markov chains. Springer-Verlag, New York, New York, U.S.A. [**18**, **23**]

Kendall, D. G. 1969. Some problems and methods in statistical archaeology. World Archaeology *1*: 68-76. [?]

Kendall, M. G. and A. Stuart. 1979. The Advanced Theory of Statistics, Vol. 2. Hafner, New York. [**11**, **20**, **33**, **48**]

King, G. 1997. A solution to the ecological inference problem: reconstructing individual behavior from aggregate data. Princeton University Press, Princeton NJ. 342 pp. [**16**]

Larsen, R. J. and M. L. Marx. 2001. An introduction to mathematical statistics and its applications, 3rd edition. Prentice Hall, Upper Saddle River, NJ . [**12**, **20**, **41**, **43**]

Larsen, R. J. and M. L. Marx. 2006. An introduction to mathematical statistics and its applications, 4th edition. Prentice Hall, Upper Saddle River, NJ . 920 p. [**44**]

Leathwick, J. R. and M. P. Austin. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. Ecology 2560-2573. [**23**]

Legendre, A. M. 1805. Nouvelles méthodes pour la détermination des orbites des comètes. Paris: Courcier [See full citation in **Stigler 1986**, p. 388] [**28**, **36**, **45**]

Legendre, P. and L. Legendre. 1998. Numerical Ecology, 2nd English Edition, Elsevier, Amsterdam. 853 pp. [**21**, **47**]

Legendre, P. and E. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia: *129*: 271-280. [**12**]

Lehmann, E. L. 2006. Nonparametrics. Statistical methods based on ranks, Revised First Edition. Springer, New York. 463 pp. [**6**]

Mayo, D. G. 1996. Error and the growth of experimental knowledge. University of Chicago Press, Chicago & London. 493 pp. [**54**]

McCulloch, C. E. And S. R. Searle. 2001. Generalized, linear, and mixed models. John Wiley & Sons, New York. 325 pp. [**24**, **44**]

Mead, R. 1988. The design of experiments. Cambridge University Press, Cambridge. 620 p.[**14**, **15**]

Nahin, P. J. 2002. Duelling idiots and other probability puzzlers. Princeton University Press, Princeton N.J. [**25**]

Neter, J, M. H, Kutner, C. J. Nachtsheim and W. Wasserman. 1996. Applied linear statistical models. Irwin, Chicago. 1408 pp. with data diskette. [**14**]

Pearson, K. 1897. On a form of spurious correlations which may arise when indices are used in the measurement of organs. Proc. Roy. Soc. London 60: 489-502. {Cited by **Schlager et al. (1998)**}[**53**]

Pielou, E. C. 1969. An introduction to mathematical ecology. Wiley-Interscience, New York. [**51**]

Pielou, E. C. 1984. The interpretation of ecological data: a primer on classification and ordination. John Wiley & Sons, New York. Read pp. 13-81 [**44**]

Pielou, E. C. 1984. The interpretation of ecological data: a primer on classification and ordination. John Wiley & Sons, New York. [**44**]

Popper, K. R. 1959. The Logic of Scientific Discovery. Hutchinson & Co., London. [**40**]

Pulliam, H. R. 1988. Sources, sinks, and population regulation. Amer. Natur. *132*: 652-661. *[p. 52]*

Ramsey, F. L. and D. W. Schafer. 1997. The statistical sleuth: a course in methods of data analysis. Duxbury Press, Belmont CA. 742 pp. [**4**, **11**, **18**, **24**, **26**, **27**, **38**, **44**, **45**, **48**, **51**, **54**]

Ramsey, F. L. and D. W. Schafer. 2002. The statistical sleuth: a course in methods of data analysis, 2nd Edition. Duxbury Press, Pacific Grove CA. 742 pp. [**3**, **4**, **10**, **11**, **18**, **24**, **26**, **27**, **30**, **38**, **44**, **45**, **48**, **51**, **53**, **54**]

Robert, C. P. and G. Casella. 1999. Monte Carlo statistical methods. Springer-Verlag, New York. 507 pp. [**6**]

Roberts, F. S. 1976. Discrete mathematical models with applications to social, biological, and environmental problems. Prentice-Hall, Englewood Cliffs, New Jersey. [**2**, **31**]

Robinson, W. S. 1950. Ecological correlation and the behavior of individuals. American Sociological Review 15: 351-357. [**16**]

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge University Press, Cambridge. [p. **52**]

Salsburg, D. 2001. The lady tasting tea: how statistics revolutionized science in the twentieth century. W. H. Freeman & Co., New York. 340 pp. [**57**]

Schlager, W., D. Marsal, P. A. G. van der Geest, and A. Sprenger. 1998. Sedimentation rates, observation span, and the problem of spurious correlation. Mathematical Geology *30*: 547-556. [p.**53**, **60**]

Shmida, A. and S. Ellner. 1984. Coexistence of plant species with similar niches. Vegetatio 58: 29-55. [p. ?]

Shmida, A. and M. V. Whittaker. 1981. Pattern and biological microsite effects in two shrub communities, southern California. Ecology *62*: 234-251. *[p. ?]*

Shmida, A. and M. V. Wilson. 1985. Biological determinants of species diversity. J. Biogeography *12*: 1-20. *[p. ?]*

Smith, W. and J. F. Grassle. 1977. Sampling properties of a family of diversity measures. Biometrics *33*: *[**51**]*

Sokal, R. R. and F. J. Rohlf. 1995. Biometry, 3rd Edition. W. H. Freeman & Co., New York. 887 pp. *[A top-notch guide to statistics with many biological examples. This text does a particularly good job with one-way ANOVA. and multiple-comparison tests]*[**34**, **54**]

Stevens, S. S. 1951. Mathematics, measurement, and psychophysics. Pp. 21-30 *in* S. S. Stevens, ed. Handbook of Experimental Psychology. Wiley, New York. [**31**]

Stigler, S. M. 1986. The history of statistics: the measurement of uncertainty before 1900. Belknap Press, Cambridge. [**10**, **12**, **15**, **28**, **35**, **36**, **57**]

Stigler, S. M. 1999. Statistics on the Table. Belknap Press, Cambridge.[**36**, **45**, **46**, **54**]

Tabachnick, B. G. & L. S. Fidell. 2001. Using multivariate statistics, 4th Ed. Allyn & Bacon, Boston. 966 pp. [**16**, **43**]

Thompson, I. M., D. P. Ankerst, C. Chi, M. S. Lucia, P. J. Goodman, J. J. Crowley, H. L. Parnes, C. A. Coltiman. 2005. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. J. Amer. Med. Assoc. 294: 66-70.[**50**]

Toothaker, L. E. 1993. Multiple comparison procedures. Sage Publications, Newbury Park, CA. 96 pp. [**16**]

Torgerson, W. S. 1952. Multidimensional scaling: I. Theory and method. Psychometrika *17*: 401-419.[?]

Vellman, P. F. and Wilkinson, L. 1993. Nominal, ordinal, interval and ratio typologies are misleading. The American Statistician, 47(1), 65-72. [?]

Yule, G. U. 1897. On the theory of correlation. J. Roy. Stat. Soc. 60: 812-854. [**45**, **57**]

Van Kampen, N. G. 1981. Stochastic process in physics and chemistry. North Holland, Amsterdam.[**31**]

Index