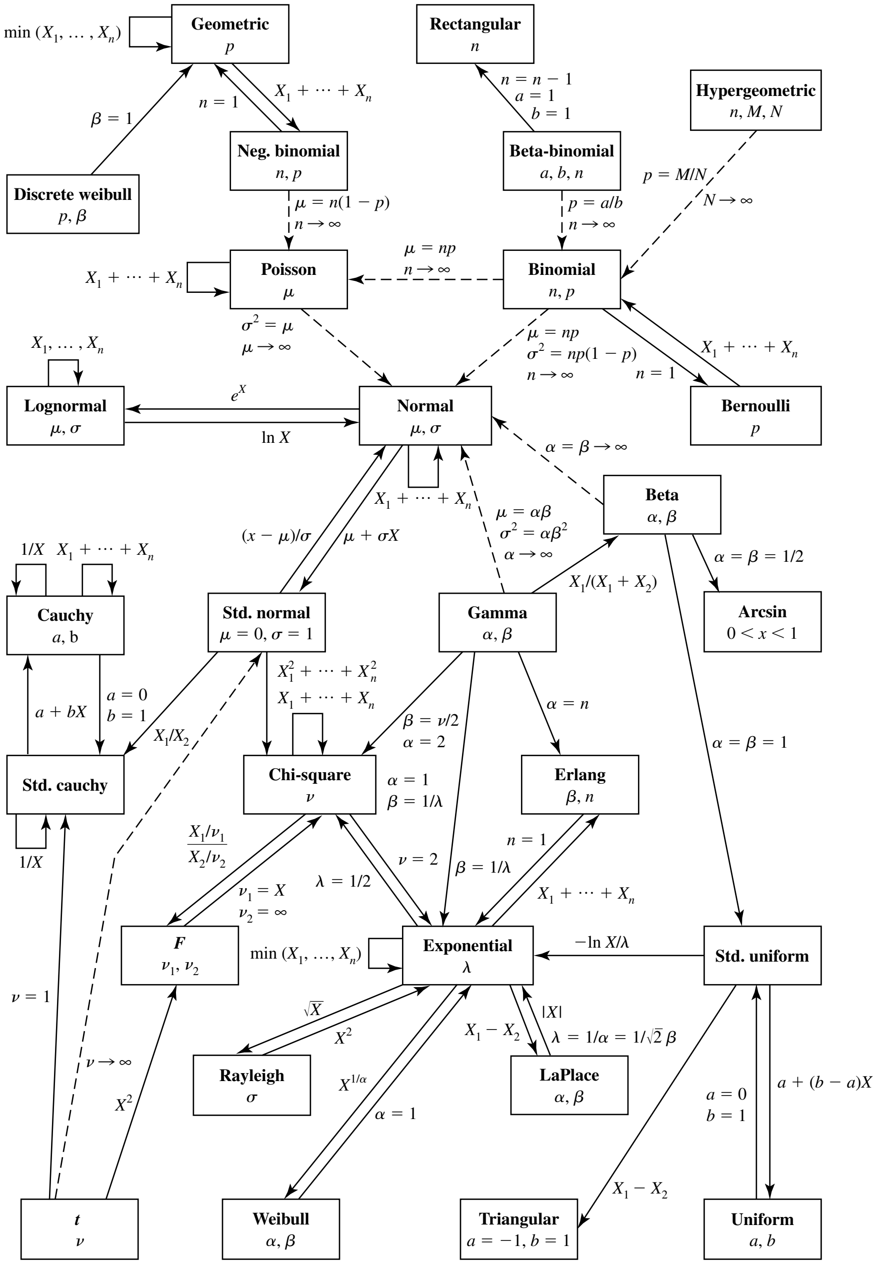
Appendix G

Relationships among distributions

This appendix exhibits relationships among some of the common univariate (discrete and continuous) distributions. The first line in each box gives the name of the distribution and the second line lists the parameters of the distribution. The flowchart, as shown on the next page, represents the three types of relationships: transformations (independent random variables are assumed) and special cases (both indicated with a solid arrow), and limiting distributions (indicated with a dashed arrow).



Adapted from Leemis, L. M. (1986), "Relationships among common univariate distributions," *The American Statistician,* Vol. 40, No. 2, pp. 143–146. With permission.

Appendix H

Maximum likelihood (ML) and unbiased estimators for parameters of some common probability distributions

Distribution	Parameter	ML estimator	Unbiased estimator
Uniform	α	$X_{(1)}$	$X_{(1)} - \left(\dfrac{X_{(n)} - X_{(1)}}{n-1}\right)$
	β	$X_{(n)}$	$X_{(n)} + \left(\dfrac{X_{(n)} - X_{(1)}}{n-1}\right)$
Normal	μ	$\overline{X}$	$\overline{X}$
	σ^2	$\sum_{i=1}^{n} \dfrac{(X_i - \overline{X})^2}{n}$	$\sum_{i=1}^{n} \dfrac{(X_i - \overline{X})^2}{n-1}$
Exponential	θ	$\dfrac{1}{\overline{X}}$	$\dfrac{n-1}{n\overline{X}}$
Poisson	λ	$\overline{X}$	$\overline{X}$
Binomial	p	$\dfrac{Y}{n}$	$\dfrac{Y}{n}$
Negative binomial	p	$\dfrac{k}{\overline{X}}$	$\dfrac{k(n-1)}{n\overline{X}}$
Geometric	p	$\dfrac{1}{\overline{X}}$	$\dfrac{n-1}{n\overline{X}}$

Note: $X_{(n)} = \max(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$
$X_{(1)} = \min(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$
$\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$
Y = total number of successes
k = number of items in a population labeled "successes"
n = sample size

Appendix I

Some important statistical formulas

Classification of Data

Relative frequency of a class	$\dfrac{\text{Frequency of the class}}{\text{Total frequency}}$
Approximate number of classes	$\dfrac{\text{Largest data value} - \text{smallest data value}}{\text{Class width}}$
Real lower limit (lower class boundary)	(Apparent lower limit) $- \dfrac{1}{2}$(unit difference)
Real upper limit (upper class boundary)	(Apparent upper limit)$-\dfrac{1}{2}$(unit difference)
Midpoint of a class	$\dfrac{\text{Lower class limit} + \text{upper class limit}}{2}$
Frequency density of a class	$\dfrac{\text{Frequency of the class}}{\text{Class size}}$

Descriptive Statistics

Finite population mean

Ungrouped data

$$\mu=\frac{\sum_{i=1}^{N} X_i}{N}$$

Grouped data

$$\mu = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i}$$

Sample mean

Ungrouped data

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Grouped data

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$

Weighted arithmetic mean

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Geometric mean

Ungrouped data

$$\mathrm{GM} = \sqrt[n]{x_1 x_2 \ldots x_n}$$

grouped data

$$\mathrm{GM} = \left(\prod_{i=1}^{k} x_i^{f_i}\right)^{1 \Big/ \sum_{i=1}^{k} f_i}$$

Harmonic mean

Ungrouped data

$$\mathrm{HM} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Grouped data

$$\mathrm{HM} = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \left(\frac{f_i}{x_i}\right)}$$

Median

Ungrouped data

odd number of observations — the $(n+1)/2$ th observation in an ordered set

even number of observations — mean of $(n/2)$th and $(n/2+1)$th observations in an ordered set

Grouped data

$$\text{Median} = l_1 + \frac{h}{f}\left(\frac{N}{2} - CF\right)$$

where ℓ_1 = lower class boundary of the median class
h = size of the class interval of the median class
f = frequency of the median class
N = total frequency
CF = cumulative frequency preceding the median class
Median class is the $(N/2)$th frequency class

Mode

Ungrouped data — Most frequent data value

Grouped data

$$\text{Mode} = \ell_1 + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right) \times h$$

where ℓ_1 = lower class boundary of the modal class
h = size of the class interval of the modal class
f_m = frequency of the modal class
f_1 = frequency preceding the modal class
f_2 = frequency following the modal class
Modal class is the most frequent data class

Mean deviation about mean

Ungrouped data $$\text{MD} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Grouped data $$\text{MD} = \frac{\sum_{i=1}^{k} f_i\, |x_i - \bar{x}|}{\sum_{i=1}^{k} f_i}$$

Mean deviation about median

Ungrouped data $$\text{MD} = \frac{\sum_{i=1}^{n} |x_i - \tilde{x}|}{n}$$

Grouped data $$\text{MD} = \frac{\sum_{i=1}^{k} f_i\, |x_i - \tilde{x}|}{\sum_{i=1}^{k} f_i}$$

Range Largest data value − Smallest data value

Finite population variance

Ungrouped data $$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Ungrouped data (computing formula) $$\sigma^2 = \frac{\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2 \Big/ N}{N}$$

Grouped data $$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (X_i - \mu)^2}{\sum_{i=1}^{k} f_i}$$

Grouped data (computing formula) $$\sigma^2 = \frac{\sum_{i=1}^{k} f_i X_i^2 - \left(\sum_{i=1}^{k} f_i X_i\right)^2 \Big/ \sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} f_i}$$

Sample variance

Ungrouped data $$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

Ungrouped data (computing formula) $$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \Big/ n}{n - 1}$$

Grouped data $$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} f_i - 1}$$

Grouped data (computing formula)

$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2 - \left(\sum_{i=1}^{k} f_i x_i\right)^2 \Big/ \sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} f_i - 1}$$

Standard deviation

Population

$$\sigma = \sqrt{\sigma^2}$$

Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

Coefficient of variation

$$\left(\frac{\text{Standard deviation}}{\text{Mean}}\right) \times 100\%$$

Covariance

Population

$$\sigma_{xy} = \frac{\sum_{i=1}^{N} (X_i - \mu_x)(Y_i - \mu_y)}{N}$$

Sample

$$s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Probability

Probability of an event

$$P(A) = \frac{\text{number of outcomes in event } A}{\text{number of equally likely and mutually exclusive outcomes in the sample space}}$$

Complement rule $P(A) = 1 - P(\bar{A})$ or $P(\bar{A}) = 1 - P(A)$

Addition rule

For two events

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B) = P(A) + P(B)$ if A and B are mutually exclusive

$P(A \cup B) = P(A) + P(B) - P(A)P(B)$ if A and B are independent

For three events

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

For *n* events

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Conditional probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \text{provided } P(B) \neq 0$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \qquad \text{provided } P(A) \neq 0$$

Multiplication rule

For two events

$P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$

$P(A \cap B) = P(A)P(B)$ if A and B are independent

For three events

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B)$

$P(A \cap B \cap C) = P(A)P(B)P(C)$ if A, B, C are independent

Theorem of total probability

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n)$$

Bayes' theorem $$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n)}$$

Mathematical Expectation

Expected value (mean) of a random variable

Discrete $$\mu = E(X) = \sum_x x\,p(x)$$

Continuous $$\mu = E(X) = \int_{-\infty}^{\infty} x\,f(x)\,dx$$

Variance of a random variable

Discrete $$\sigma^2 = \text{Var}(X) = \sum_x (x-\mu)^2 p(x)$$

Continuous $$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)$$

Variance of a random variable (computational formula)

Discrete $$\sigma^2 = \text{Var}(X) = \sum_x x^2 p(x) - \mu^2$$

Continuous $$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

Properties of the mean and variance of a random variable
If X is a random variable and a and b are any arbitrary constants, then

Mean of $(a + bX)$ $\mu_{(a+bX)} = a + b\mu_X$

Variance of $(a + bX)$ $\sigma^2_{(a+bX)} = b^2\sigma^2_X$

Properties of a discrete probability distribution $p(x)$

1. $p(x) \geq 0$ for any value of x

2. $\sum_x p(x) = 1$

Properties of a continuous probability distribution $f(x)$

1. $f(x) \geq 0$ for any value of x

2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Moment generating function (about origin)

Discrete $$M_X(t) = E(e^{tX}) = \sum_X e^{tx}\,p(x)$$

Continuous $$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

Moment generating function (about mean)

Discrete $$M_{(X-\mu)}(t) = E\left(e^{t(X-\mu)}\right) = \mathrm{e}^{-\mu t} \sum_x e^{tx}\,p(x)$$

Continuous $$M_{(X-\mu)}(t) = E\left(e^{t(X-\mu)}\right) = e^{-\mu t} \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

Characteristic function (about origin)

$$\left.\begin{array}{ll}\textbf{Discrete} & \phi_X(t) = E(e^{itX}) = \sum_x e^{itx}\,p(x) \\ \textbf{Continuous} & \phi_X(t) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f(x)dx\end{array}\right\} i = \sqrt{-1}$$

Characteristic function (about mean)

$$\left.\begin{array}{ll}\textbf{Discrete} & \phi_{X-\mu}(t) = E\left(e^{it(X-\mu)}\right) = e^{-i\mu t}\sum_x e^{itx}\,p(x) \\ \textbf{Continuous} & \phi_{X-\mu}(t) = E\left(e^{it(X-\mu)}\right) = e^{-i\mu t}\int_{-\infty}^{\infty} e^{itx} f(x)\,dx\end{array}\right\} i = \sqrt{-1}$$

Combinatorics

Number of permutations

For n distinct objects taken all together $\quad n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$

For n distinct objects taken r at a time $\quad P_r^n = \dfrac{n!}{(n-r)!}$

For n objects in which n_1 are of first kind, n_2 are of second kind, . . . , n_k are of kth kind, such that $n_1 + n_2 + \cdots + n_k = n$ $\quad \dfrac{n!}{n_1!n_2!\cdots n_k!}$

Circular arrangement for n distinct objects $\quad (n-1)!$

Number of combinations

For r objects selected from n distinct objects $\quad \dbinom{n}{r} = {}^nC_r = \dfrac{n!}{r!(n-r)!}$

Number of experimental outcomes with r successes in n Bernoulli trials $\quad \dfrac{n!}{r!(n-r)!}$

Binomial expansion $\quad (a+b)^n = \sum_{k=0}^{n} \dbinom{n}{k} a^k b^{n-k}$

Multiplicative rule

Total number of possible outcomes for a sequence of k events in which the first one has n_1 possibilities, the second one has n_2 possibilities, the third one has n_3 possibilities, . . . , and the kth event has n_k possibilities $\quad n_1 \cdot n_2 \cdot n_3 \cdots n_k$

Total number of possible outcomes for a sequence of k events in which each event has n possibilities $\quad n^k$

Sampling Distributions

Inferences about a population mean μ and variance σ^2

Sample mean $\quad \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

Sample variance $\quad S^2 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$

Sample standard deviation $S = \sqrt{S^2}$

Expected value of $\bar{X}$ $E(\bar{X}) = \mu$

Standard deviation of $\bar{X}$
(with correction for finite population)

$$\sigma_{\bar{X}} = \sqrt{\frac{N-n}{N-1}}\frac{\sigma}{\sqrt{n}}$$

Standard deviation of $\bar{X}$
(without correction for finite population, if the sample size is less than or equal to 5% of the population size)

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Sampling distribution of $\bar{X}$
(when parent population is normal and σ^2 is known)

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

Sampling distribution of $\bar{X}$
(when parent population is not necessarily normal, σ^2 is unknown, and sample is large)

$$Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

Sampling distribution of $\bar{X}$
(when parent population is normal, σ^2 is unknown, and sample is small)

$$t_{n-1} = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

Sampling distribution of S^2
(when parent population is normal and the mean μ is unknown)

$$\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma^2}$$

Inferences about a population proportion *p*

Sample proportion $\bar{p}$

$$\bar{p} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad x_i = 0, 1$$

Sample variance

$$S^2 = \frac{\bar{p}(1-\bar{p})}{n}$$

Sample standard deviation

$$S = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Expected value of $\bar{p}$ $E(\bar{p}) = p$

Standard deviation of $\bar{p}$
(with correction for finite population)

$$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}\frac{p(1-p)}{n}}$$

Standard deviation of $\bar{p}$
(without correction for finite population, if the sample size is less than or equal to 5% of the population size)

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Sampling distribution of $\bar{p}$
[when p is known and sample is large $(np(1-p) \geq 5)$]

$$Z = \frac{\bar{p}-p}{\sqrt{p(1-p)/n}}$$

Sampling distribution of $\bar{p}$
[when p is unknown and sample is large $(n\bar{p}(1-\bar{p}) \geq 5)$]

$$Z = \frac{\bar{p}-p}{\sqrt{\bar{p}(1-\bar{p})/n}}$$

Inferences about the difference between two population means $(\mu_1 - \mu_2)$

Sample means $\bar{X}_1 = \dfrac{\sum_{i=1}^{n_1} X_{1i}}{n_1}$ $\quad \bar{X}_2 = \dfrac{\sum_{i=1}^{n_2} X_{2i}}{n_2}$

Sample variances $S_1^2 = \dfrac{\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2}{n_1 - 1}$ $\quad S_2^2 = \dfrac{\sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2}{n_2 - 1}$

Sample standard deviations $S_1 = \sqrt{S_1^2} \quad S_2 = \sqrt{S_2^2}$

Expected value of $\bar{X}_1 - \bar{X}_2$ $E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$

Standard deviation of $\bar{X}_1 - \bar{X}_2$ $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Sampling distribution of $\bar{X}_1 - \bar{X}_2$
(when parent populations are normal and σ_1^2 and σ_2^2 are known)

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

Sampling distribution of $\bar{X}_1 - \bar{X}_2$
(when parent populations are not necessarily normal, σ_1^2 and σ_2^2 are unknown, and samples are large)

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}$$

Sampling distribution of $\bar{X}_1 - \bar{X}_2$
(when parent populations are normal, σ_1^2 and σ_2^2 are unknown, $\sigma_1^2 = \sigma_2^2$, and samples are small)

$$t_v = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{(1/n_1) + (1/n_2)}}$$

$$\text{where } S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$v = n_1 + n_2 - 2$$

Sampling distribution of $\bar{X}_1 - \bar{X}_2$
(when parent populations are normal, σ_1^2 and σ_2^2 are unknown, $\sigma_1^2 \neq \sigma_2^2$, and samples are small)

$$t_v = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}$$

$$\text{where } v' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Inferences about the difference between two population proportions $(p_1 - p_2)$

Sample proportions $\bar{p}_1 = \dfrac{\sum_{i=1}^{n_1} x_{1i}}{n_1} \quad \bar{p}_2 = \dfrac{\sum_{i=1}^{n_2} x_{2i}}{n_2}$

$x_{1i}, x_{2i} = 0, 1$

Sample variances $S_1^2 = \dfrac{\bar{p}_1(1 - \bar{p}_1)}{n_1} \quad S_2^2 = \dfrac{\bar{p}_2(1 - \bar{p}_2)}{n_2}$

Sample standard deviations $S_1 = \sqrt{\dfrac{\bar{p}_1(1 - \bar{p}_1)}{n_1}} \quad S_2 = \sqrt{\dfrac{\bar{p}_2(1 - \bar{p}_2)}{n_2}}$

Expected value of $(\bar{p}_1 - \bar{p}_2)$ $E(\bar{p}_1 - \bar{p}_2) = p_1 - p_2$

Standard deviation of $(\bar{p}_1 - \bar{p}_2)$ $$\sigma_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Sampling distribution of $(\bar{p}_1 - \bar{p}_2)$
[when p_1 and p_2 are known and samples are large $(n_1 p_1(1-p_1) \geq 5, n_2 p_2(1-p_2) \geq 5)$]

$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Sampling distribution of $\bar{p}_1 - \bar{p}_2$
[when p_1 and p_2 are unknown and samples are large $(n_1 \bar{p}_1(1-\bar{p}_1) \geq 5, n_2 \bar{p}_2(1-\bar{p}_2) \geq 5)$]

$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}}$$

Inferences about the ratio of two population variances

Sampling distribution of S_1^2/S_2^2
[when parent populations are normal and their means (μ_1 and μ_2) are unknown

$$F_{\nu_1,\nu_2} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

where $\nu_1 = \mathrm{n}_1 - 1,\ \nu_2 = \mathrm{n}_2 - 1$

Tests for Goodness of Fit and Independence

Chi-square statistic for goodness of fit

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Expected frequencies for contingency table under the assumption of independence or homogeneity

$$E_{ij} = \frac{(i\text{th row total})\ (j\text{th column total})}{\text{grand total}}$$

Chi-square statistic for contingency table

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Analysis of Variance

Completely randomized design

Total sum of squares

$$\mathrm{SS}_T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{..}^2}{N}$$

Sum of squares due to treatments

$$\mathrm{SS}_{Tr} = \sum_{i=1}^{k} n_i (\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^{k} \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{N}$$

Sum of squares due to error

$$\mathrm{SS}_E = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{k} \frac{x_{i.}^2}{n_i}$$

Randomized block design

Total sum of squares

$$\mathrm{SS}_T = \sum_{i=1}^{t} \sum_{j=1}^{b} (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{t} \sum_{j=1}^{b} x_{ij}^2 - \frac{x_{..}^2}{tb}$$

Sum of squares due to treatments

$$\mathrm{SS}_{Tr} = b \sum_{i=1}^{t} (\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^{t} \frac{x_{i.}^2}{b} - \frac{x_{..}^2}{tb}$$

Sum of squares due to blocks

$$\mathrm{SS}_B = t \sum_{j=1}^{b} (\bar{x}_{.j} - \bar{x}_{..})^2 = \sum_{j=1}^{b} \frac{x_{.j}^2}{t} - \frac{x_{..}^2}{tb}$$

Sum of squares due to error

$$\mathrm{SS}_E = \sum_{i=1}^{t}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$$
$$= \sum_{i=1}^{t}\sum_{j=1}^{b}x_{ij}^2 - \sum_{i=1}^{t}\frac{x_{i.}^2}{b} - \sum_{j=1}^{b}\frac{x_{.j}^2}{t} + \frac{x_{..}^2}{tb}$$

Randomized block design with replication

Total sum of squares

$$\mathrm{SS}_T = \sum_{i=1}^{t}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^{t}\sum_{j=1}^{b}\sum_{k=1}^{n}x_{ijk}^2 - \frac{x_{...}^2}{tbn}$$

Sum of squares due to treatments

$$\mathrm{SS}_{Tr} = bn\sum_{i=1}^{t}(\bar{x}_{i..} - \bar{x}_{...})^2 = \sum_{i=1}^{t}\frac{x_{i..}^2}{bn} - \frac{x_{...}^2}{tbn}$$

Sum of squares due to blocks

$$\mathrm{SS}_B = tn\sum_{j=1}^{b}(\bar{x}_{.j.} - \bar{x}_{...})^2 = \sum_{j=1}^{b}\frac{x_{.j.}^2}{tn} - \frac{x_{...}^2}{tbn}$$

Sum of squares due to interaction

$$\mathrm{SS}_{TB} = n\sum_{i=1}^{t}\sum_{j=1}^{b}(\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2$$
$$= \sum_{i=1}^{t}\sum_{j=1}^{b}\frac{x_{ij.}^2}{n} - \sum_{i=1}^{t}\frac{x_{i..}^2}{bn} - \sum_{j=1}^{b}\frac{x_{.j.}^2}{tn} + \frac{x_{...}^2}{tbn}$$

Sum of squares due to error

$$\mathrm{SS}_E = \sum_{i=1}^{t}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij.})^2$$
$$= \sum_{i=1}^{t}\sum_{j=1}^{b}\sum_{k=1}^{n}x_{ijk}^2 - \sum_{i=1}^{t}\sum_{j=1}^{b}\frac{x_{ij.}^2}{n}$$

Latin square design

Total sum of squares

$$\mathrm{SS}_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{h=1}^{p}(x_{ij(h)} - \bar{x}_{...})^2 = \sum_{i=1}^{p}\sum_{j=1}^{p}x_{ij(h)}^2 - \frac{x_{...}^2}{p^2}$$

Sum of squares due to rows

$$\mathrm{SS}_R = p\sum_{i=1}^{p}(\bar{x}_{i..} - \bar{x}_{...})^2 = \sum_{i=1}^{p}\frac{x_{i..}^2}{p} - \frac{x_{...}^2}{p^2}$$

Sum of squares due to columns

$$\mathrm{SS}_C = p\sum_{j=1}^{p}(\bar{x}_{.j.} - \bar{x}_{...})^2 = \sum_{j=1}^{p}\frac{x_{.j.}^2}{p} - \frac{x_{...}^2}{p^2}$$

Sum of squares due to treatments

$$\mathrm{SS}_{Tr} = p\sum_{h=1}^{p}(\bar{x}_{..(h)} - \bar{x}_{...})^2 = \sum_{h=1}^{p}\frac{x_{..(h)}^2}{p} - \frac{x_{...}^2}{p^2}$$

Sum of squares due to error

$$\mathrm{SS}_E = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{h=1}^{p}(x_{ij(h)} - \bar{x}_{i..} - \bar{x}_{.j.} - \bar{x}_{..(h)} + 2\bar{x}_{...})^2$$
$$= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{h=1}^{p}x_{ij(h)}^2 - \sum_{i=1}^{p}\frac{x_{i..}^2}{p} - \sum_{j=1}^{p}\frac{x_{.j..}^2}{p} - \sum_{h=1}^{p}\frac{x_{..(h)}^2}{p} + \frac{2x_{x...}^2}{p}$$

Regression and correlation

Pearson product moment correlation coefficient

Population
$$\rho = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{N}(y_i - \mu_y)^2}}$$

Sample
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

• Computational formula
$$r = \frac{\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)\Big/ n}{\left\{\left(\sqrt{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\Big/ n}\right)\left(\sqrt{\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\Big/ n}\right)\right\}}$$

Slope and y intercept of a regression line

Slope
$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

• Computational formula
$$b = \frac{\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)\Big/ n}{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\Big/ n}$$

Intercept $a = \bar{y} - b\bar{x}$

Estimated regression line $\hat{y} = a + bx$

Total sum of squares
$$\mathrm{SS}_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

• Computational formula
$$\mathrm{SS}_T = \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\Big/ n$$

Sum of squares due to regression
$$\mathrm{SS}_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

• Computational formula
$$\mathrm{SS}_R = \frac{\left[\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)\Big/ n\right]^2}{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\Big/ n}$$

Sum of squares due to error
$$\mathrm{SS}_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

• Computational formula $\mathrm{SS}_E = \mathrm{SS}_T - \mathrm{SS}_R$

Coefficient of determination
$$r^2 = \frac{\mathrm{SS}_R}{\mathrm{SS}_T} = 1 - \frac{\mathrm{SS}_E}{\mathrm{SS}_T}$$

Estimate of σ^2
$$\mathrm{MS}_E = \frac{\mathrm{SS}_E}{n - 2}$$

Estimated variance of *b*
$$S_b^2 = \frac{\text{MS}_E}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Mean square due to regression
$$\text{MS}_R = \frac{\text{SS}_R}{\text{regression degrees of freedom}}$$

F statistic
$$F = \frac{\text{MS}_R}{\text{MS}_E}$$

Confidence interval for $E(y_p)$
$$\bar{y}_p \pm t_{(n-2),\,\alpha/2}\sqrt{\text{MS}_E\left(\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)}$$

Prediction interval for y_p
$$\hat{y}_p \pm t_{(n-2),\,\alpha/2}\sqrt{\text{MS}_E\left(1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)}$$

Nonparametric Methods

Sign test

Test statistic: X = number of positive or negative signs $\quad E(X) = \frac{1}{2}n$

Null distribution: binomial $b(x; n, 0.5)$ $\quad \text{Var}(X) = \frac{1}{4}n$

Wilcoxon signed-rank test

Test statistic: T = sum of positive or negative ranks $\quad E(T) = \frac{n(n+1)}{4}$

Null distribution (large sample): Normal $\quad \text{Var}(T) = \frac{n(n+1)(2n+1)}{24}$

Run test for randomness

Test statistic: R = number of runs $\quad E(R) = \frac{2n_1n_2}{n_1+n_2} + 1$

Null distribution (large sample): Normal $\quad \text{Var}(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$

where n_1 and n_2 are the number of symbols of type I and type II, respectively

Wilcoxon rank-sum test

Test statistic: W = sum of the ranks of the first sample $\quad E(W) = \frac{1}{2}n_1(n_1+n_2+1)$

Null distribution (large sample): Normal $\quad \text{Var}(W) = \frac{1}{12}n_1n_2(n_1+n_2+1)$

where n_1 and n_2 are sample sizes of the first and second samples, respectively

Mann–Whitney test

Test statistic: U = number of pairs (x_i, y_i) such that $x_i < y_i$ $\quad E(U) = \frac{1}{2}n_1n_2$

Null distribution (large sample): Normal $\quad \text{Var}(U) = \frac{1}{12}n_1n_2(n_1+n_2+1)$

where n_1 and n_2 are sample sizes of the first and second samples, respectively

Kruskal–Wallis test

Test statistic:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

where R_i = sum of the ranks of the observations in the ith treatment group
n_i = number of observations in the ith group
k = number of groups
$N = n_1 + n_2 + \cdots + n_k$

Null distribution (large sample): χ^2_{k-1}

$$E(R_i) = \frac{n_i(N+1)}{2}$$

$$\text{Var}\,(R_i) = \frac{n_i(N+1)(N-n_i)}{12}$$

Friedman's test

Test statistic:

$$F_r = \frac{12}{bk(k+1)} \sum_{i=1}^{k} \left[R_i - \frac{b(k+1)}{2}\right]^2$$

where R_i = sum of the ranks of the observations in the ith treatment group
b = number of blocks
k = number of treatment groups

Null distribution (large sample): χ^2_{k-1}

$$E(R_i) = \frac{b(k+1)}{2}$$

$$\text{Var}\,(R_i) = \frac{b(k+1)(k-1)}{12}$$

Spearman's rho test

Test statistic:

$$R_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where d_i is the difference between the ranks of the ith pair

Null distribution (large sample): Normal

$$E(R_s) = 0$$

$$\text{Var}\,(R_s) = \frac{1}{n-1}$$

Control Charts

Control lines for the $\bar{x}$-chart

$$\text{LCL}_{\bar{X}} = \bar{\bar{x}} - A_2\bar{R}$$

$$\text{UCL}_{\bar{X}} = \bar{\bar{x}} + A_2\bar{R}$$

where $\bar{\bar{x}}$ is the average of all the subgroup means $\bar{x}_i$, $\bar{R}$ is the average of all the subgroup ranges (taken from a pilot set of about 20 rational subgroups), and A_2 is a multiplier obtained from the following table:

n	**2**	**3**	**4**	**5**	**6**	**7**
A_2	1.880	1.023	0.729	0.577	0.483	0.419

Control lines for the *R* chart

$\text{LCL}_R = D_3\bar{R}$
$\text{UCL}_R = D_4\bar{R}$

where $\bar{R}$ is the average of ranges taken from a pilot set (about 20 rational subgroups), and D_3 and D_4 are multipliers obtained from the following table:

n	**2**	**3**	**4**	**5**	**6**	**7**
D_3	0	0	0	0	0	0.076
D_4	3.267	2.574	2.282	2.114	2.004	1.924

Control lines for the *s* chart

$\text{LCL}_s = B_3\bar{s}$
$\text{UCL}_s = B_4\bar{s}$

where $\bar{s}$ is the average of all the subgroup standard deviations taken from a pilot set (about 20 rational subgroups) and B_3 and B_4 are determined from the following table:

n	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
B_3	0	0	0	0	0.030	0.118	0.185	0.239	0.284
B_4	3.267	2.568	2.666	2.089	1.970	1.882	1.815	1.761	1.716

Control lines for the *p* chart

$$\text{LCL}_p = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$\text{UCL}_p = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

where $\bar{p}$ is the average proportion defective taken from a pilot set (about 20 rational subgroups) and n is the sample size.

Control lines for the *c* chart

$\text{LCL}_c = \bar{c} - 3\sqrt{\bar{c}}$
$\text{UCL}_c = \bar{c} + 3\sqrt{\bar{c}}$

where $\bar{c}$ is the average number of defects taken from a pilot set (about 20 rational subgroups)