

## Chapter 1

Statistics (统计学): 研究数据资料的收集、整理、分析和解释 (interpretation) 的科学。

Biostatistics (生物统计学): 统计学应用于生物科学

Variable (变量): 指某种特征, 它的表现在不同个体间或不同组间存在变异性。

Observation (观测值): 指对变量的表现进行观察或测量所获得的数值, 有时也被称为变数 (variate)

Population (总体): 又叫“统计总体”, 是指一个统计问题研究对象的全体, 它是具有某种 (或某些) 共同特征的元素的集合。

Individual (个体): 总体中每一个研究对象称作个体。

Sample (样本): 从总体中按一定方法抽取部分具有代表性的个体, 这部分个体称为样本。

Parameter (参数): 描述总体特征的数, 如总体平均数、总体方差等。

Statistic (统计量): 描述样本特征的量, 如样本平均数、样本方差、样本相关系数等。

Accuracy (准确性): 指观测值或估计值与真值的接近程度。

Precision (精确性): 对同一物体的重复观察值或估计值彼此之间的接近程度。

## Chapter 2

Raw data (直接数据): 数据调查与实验未经处理的数据;

Continuous data (连续性数据): 指在一定范围内可取任何实数值的数据。

Discrete data (离散性数据): 在一定范围内只能取有限种可能值的数据。

Count data (计数数据): 用计数的方式得到的数据资料, 必须用整数来表示。

Classification data (分类资料): 可自然的或人为的分为 2 个或多个不同类别的资料。例如: 男生记做 1 女生记做 2

频数 (率) 分布 (frequency distribution); 下四分位数 (lower quartile); 中位数 (median); 上四分位数 (upper quartile); 条形图 (bar chart); 直方图 (histogram); 饼图 (pie chart); 散点图 (scatter plot), 组间距 (interval)

Percentile (百分位数): 一组  $n$  个观测值按数值大小排列, 小于某数值的数据个数占全体个数的  $x\%$ , 则为  $x\%$  分位数。25% 分位数为  $0.25n$  个数后面一个数。

Central tendency (集中趋势): 变量分布的中心位置。

Arithmetic mean (算术平均数): 观测值总和被观测值个数除所得的商数。

Geometric mean (几何平均数):  $n$  个观察值相乘积开  $n$  次方所得的根。

Mode (众数): 资料中出现次数最多的数。

Dispersion tendency (离散趋势/变异程度): 反映集中趋势对数据的代表程度。

方差 (variance); 自由度 (degree of freedom, df) (分母  $n-1$ ); 离均差平方和 (sum of squares,

SS) ( $\sum (X_i - \bar{X})^2$ ); 标准差 (standard deviation); 校正项 (correction factor/correction term)。

Range (范围/全距/极差): 样本中最大值与最小值之差。

Mean absolute deviation (平均绝对离差 MD): 各观测值离均差绝对值的平均数。

Coefficient of variation (变异系数 CV): 数据相对变异程度大小的度量。

## Chapter 3

Random variable (随机变量): 在一定范围内随机取值的变量。

Discrete Random variables (离散型随机变量): Continuous Random variables (连续型随机变量):

Probability function (概率函数): 描述离散型随机变量取各个可能值的概率的函数。

Probability density function (概率密度函数): 描述连续性随机变量取某值的概率密度的函数。

Probability distribution function (概率分布函数): 描述随机变量取值小于等于某值的概率的函数。

Expectation (数学期望): 随机变量的期望。

Normal distribution (正态分布); Binomial distribution (二项分布); Poisson (泊松分布)。

Skewness (偏度): 衡量随机变量分布对称性 (比较  $M_0, M_d, \bar{X}$ )

Kurtosis (峰度): 衡量随机变量分布平坦程度

## Chapter 4

Statistical inference (统计推断): 对样本数据的分析来对总体进行推断。

Parameter estimation (参数估计): 用样本统计量来对总体参数进行估计。

Hypothesis testing (假设检验): 用样本统计量对总体分布特征进行检验。

Sampling distribution (抽样分布): 从总体中随机抽取含量为  $n$  的样本, 并由样本计算各种统计量的概率分布。

Standard error (标准误差 SE)。

Central limit theorem (中心极限定理): 无论原总体是什么分布, 只要样本足够大, 样本平均数就近似服从正态分布。

Point estimation (点估计): 以某个样本统计量作为该参数的一个估计值。

Moment estimation (矩估计法): 用于总体参数相应的样本统计量作为估计量, 必要时可对用计量做适当调整。

Maximum likelihood estimation (最大似然法): 用使样本观测值的似然函数达到最大的统计量作为估计量。

Unbiased estimation (无偏估计量):  $\bar{\theta}$  为  $\theta$  的一个估计量,  $E(\bar{\theta}) = \theta$ , 称  $\bar{\theta}$  为  $\theta$  的无偏估计量。

Sampling variance (抽样方差): 估计量的方差。  $D(S^2) = \frac{2\sigma^4}{(n-1)}$

Standard error of Mean 标准误:  $SEM = \sigma / \sqrt{n}$

Mean squared error (均方误):  $E(\bar{\theta} - \theta)^2$   $MSE$

consistency 一致性: 一致收敛于真值;

efficacy 有效性: 抽样方差达到最小的无偏估计;

sufficiency 充分性: 估计函数包含了关于被估参数的全部信息

Interval estimation (区间估计): 以一定的置信度对参数真值的可能取值范围进行估计。

Confidence interval (置信区间): 区间  $(\bar{\theta}_1, \bar{\theta}_2)$

confidence level (置信水平):  $1 - \alpha$

significant level (显著性水平):  $\alpha$

Power of the Test (检验功效): 表示一个错误的原假设能够被否定的概率, 由于 II 型错误的概率是一个错误的原假设被接受的概率, 所以检验功效为  $1 - \beta$ 。

Test of significance (显著性检验): 假设检验。

Two-tailed test (双侧检验): 假设检验的否定域分别位于检验统计量抽样分布的两个尾部。

One-tailed test (单侧检验): 否定域在检验统计量抽样分布的一侧。

Type I error (第一类错误):  $H_0$  为真时, 样本落入拒绝域, 否定原假设。

Type II error (第二类错误):  $H_1$  为真时, 样本落入接受域, 接受原假设。

Power of the test (检验功效): 一个错误的原假设能够被否定的概率。

test statistic (检测统计量): 用于检验假设量是否正确的量, 有  $Z$  和  $t$

## Chapter 5/6

Analysis of variance (方差分析): 用于两个及两个以上样本均数差别的显著性检验。

One-way classification (单项分类): 以一个标志来分类 (或称分组) 的。

均方 (mean squares); 差异不显著 (not significant); 多重比较 (multiple comparison); 最小显著差数 (least significant difference,  $LSD_{\alpha}$ )

## Chapter 7

interact effect: 交互, 每个因子并不是独立地对观测值起作用, 两因子不同水平的组合也会起作用, 从而使得一个因子的某个水平在另一个因子的不同水平中有不同的效应;

main effect: 主效应, 每个因子简单效应的平均。 $a_1 - a_0$ : 称为  $a_1$  与  $a_0$  比较的简单效应 (除了交互效应)

Crossed classification (交叉分组): 每一因子的每一水平都与所有其他因子的各个水平发生交叉组合, 每个水平组合就是一个组。

## Chapter 8

Cross classification: 一个因子的所有水平可以与其他因子的任何水平相结合。

Nested classification (嵌套分组): 一个因子的不同水平分别与另一因子的不同水平发生组合,  $B$  的水平嵌套在  $A$  内

Fixed factor: 一个因子的各个水平是我们有目的地挑选出来的, 而我们的研究目的是要比较这些因子之间有无差异或估计这些水平的效应, 则称之为固定因子, 该因子各个水平的效应称为固定效应。

Random factor: 一个因子的各个水平是从该因子的所有可能水平中随机抽取的, 我们的研究目的不仅是要比较这些水平之间有无差异, 而是了解该因子不同水平的总体变异情况, 即对该总体的方差进行检验或估计, 则称之为随机因子, 该因子各个水平的效应称为随机效应。

fixed model: 因子的不同水平之间有无差异

mixed model: 检验  $A$  因子不同水平之间有无差异,  $B$  因子不同水平效应的方差是否为 0

random model: 检验每个因子不同水平效应的方差是否为 0

## Chapter 9

Relationship (统计相关): 统计学将非确定关系称为统计相关关系。

Correlation (相关关系): 研究变量间关系的强弱程度。

Regression (回归关系): 研究变量间的因果关系。

$SP_{xy}$  Sum of products (离均差乘积和):  $(X - \bar{X})(Y - \bar{Y})$

Covariance (样本协方差):  $Cov(X, Y)$

## Chapter 10

Multiple correlation (复相关): 一个变量与其他多个变量之间的相关

Partial correlation (偏相关) 其他变量保持不变时,  $X_i$  与  $X_j$  之间的相关

## Chapter 11

线性化 (linearization); 幂函数 (power function); 对数函数 (logarithmic function); 双曲线函数 (hyperbolic function); 多项式回归 (polynomial regression)。

## Chapter 12

辅助变量 (instrumental variable); 协变量 (co-variable)。

Analysis of covariance (协方差分析): 用回归分析对初始值的影响进行校正, 再进行方差分析。

## Chapter 13

Rate(率): 指其中的某个类别的出现概率。

Interval estimation of Rate: 率的区间估计

The comparison of two rate: 率的比较

Chi square fitness test (卡方适合性检验): 检验某一分类资料所在总体的分布是否符合某个假设的或理论的分布。例如群体中的性别比例是否符合 1: 1 的理论比例。

Chi square independence test (独立性检验): 也就是说不同类别的概率分布是否与处理相关联, 或者说类别与处理是否独立无关。

## Chapter 14

Test(试验): 在人为控制条件下进行的一种有目的的实践活动与观察不同是获取数据资料的过程

Test Unit (试验单元): 根据试验目的确定的试验材料的基本单元, 如: 细胞, 组织, 个体

Test index (试验指标): 对每个试验单元度量的实验结果标志 (日增重, 产奶量)

Test Factor (试验因子): 试验中要研究的影响试验指标的因素

Factor levels (因子水平): 对试验因子划分等级状态

Treatment (处理): 根据试验因子的不同水平对试验单元所采取的不同措施 (单因子试验: 每个水平是一个处理; 多因子试验: 每个水平组合是一个处理)

Test error (试验误差): 由于除试验处理外的其他因素所引起的观测值的变异

systematical error (系统误差): 由于干扰因子造成的不同处理内的试验单元之间存在整体性的差异

random error (随机误差): 由于随机和偶然因素所造成的误差, 造成试验单元之间的随机变异

Probability sample (概率抽样): 以一定的概率来对总体中的个体进行抽样

Non-probability sampling (非概率抽样): 从总体中非随机地选择特定的个体

Simple random sampling (简单随机抽样): 用完全随机的方法对总体中的个体进行抽样, 总体中的每一个个体被抽中的概率相等

Stratified random sampling (分层随机抽样): 将总体按所要研究的指标分为若干不重叠的层次或组, 再从每个层次中按该层次在总体中所占比例进行简单随机抽样