1. 取样误差：样本统计量与相应的总体参数之间的差距
2. 偏态分布：分数堆积在分布的一端，而另一端成为比较尖细的尾端，其与对称分布对应
3. 离散型变量：由分离的、不可分割的范畴组成，临近范畴之间没有值存在
4. 连续型变量：在任何两个观测值之间都存在无限多个可能值，它可被分割成无限多个组成部分
5. 集中量数：集中量数又叫集中趋势，是体现一组数据一般水平的统计量。它能反映频数分布中大量数据向某一点集中的情况。

1．算数平均数 $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{N}$

特点 ①在一组数据中每个变量与平均数之差的总和等于零 ②在一组数据中，每一个数都加上一个常数 C，所得的平均数为原来的平均数加常数C ③在一组数据中，每一个数都乘以一个常数 C，所得的平均数为原来的平均数乘以常数 C

优点：反应灵敏、计算严密、计算简单、简明易解、适合于进一步用代数方法盐酸、较少受抽样变动的影响

缺点：易受极端数据的影响、不能在出现模糊数据时计算

差异量数：差异量数就是对一组数据的变异性，即离中趋势特点进行度量和描述的统计量，也称为离散量数。

1. 离差与平均差$x = X - \mu$

平均差：次数分布中所有原始数据与平均数绝对离差的平均值

$$A.D. = \frac{\Sigma \left| X_i - \bar{X} \right|}{n}$$

2. 方差与标准差

$$SS = \sum (X - \mu)^2 = \sum X^2 - \frac{(\sum X)^2}{N}$$

（1）总体的方差和标准差

方差：每个数据与该组数据平均数之差乘方后的均值，即离均差平房后的均数

作为样本统计量用符号 s_2 表示，作为总体参数用符号 σ_2表示，也叫均方。

$$\sigma^2 = \frac{SS}{N}$$

标准差：方差的平方根

作为样本统计量用符号 s 表示，作为总体参数用符号 σ表示。$\sigma = \sqrt{\sigma^2}$

（2）样本的方差和标准差

样本的变异性往往比它来自的总体的变异性要小。为了校正样本数据带来的偏差，在计算样本方差时，我们用自由度来矫正样本误差，从而有利于对总体参数更好

的无偏差估计：$S^2 = \frac{SS}{n-1}$ $S = \sqrt{S^2}$

①每一个观测值都加一个相同的常数 C 之后，计算得到的标准差等于原来的标准差 ②每一个观测值都乘以一个相同的常数 C，所得到的标准差等于原标准差乘以这个常数

3. 变异系数

差异系数：一种最常用的相对差异量，为标准差对平均数的百分比

$$CV = \frac{s}{\bar{X}} \times 100\%$$

（四）相对量数

1. 百分位数：在整个分布中，在某一值之下或等于该值的分数的百分比，所对应的分数。百分位数和百分等级是同一操作定义的两端。当我们求累计次数占总体的百分比是，所对应的分数和百分比的值分别为百分位数和百分等级。

2. 百分等级：常模团体中低于该分数的人所占总体的百分比。百分等级一定要对应分数区间的精确上限。百分等级和百分位数都可以由已知数据用差值法求解。

3. 标准分数：以标准差为单位表示一个原始分数在团体中所处位置的相对位置量数，也叫 Z 分数。离平均数有多远，即表示原始分数在平均数以上或以下几个标准差的位置。

$$Z = \frac{X - \bar{X}}{S}$$

①Z 分数无实际单位，是以平均数为参照点，以标准差为单位的一个相对量

②一组原始分数转换得到的 Z 分数可正可负，所有原始分数的 Z 分数之和为零

③原始数据的 Z 分数的标准差为 1

④若原始分数呈正态分布，则转换得到的所有 Z 分数均值为 0，标准差为 1 的标准正态分布

①可比性——不同性质的成绩，一经转换为标准分数，就可在同一背景下比较

②可加性——不同性质的原始数据具有相同的参照点，因此可相加

③明确性——知道了标准分数，利用分布寒暑表就能知道其百分等级

④稳定性——转换成标准分数之后，规定了标准差为 1，保证了不同性质分数在总分数中权重一样

（五）相关量数

相关系数：两列变量间相关程度的数字表现形式。作为样本的统计量用 r 表示，作为总体参数一般用 ρ 表示。

正相关：两列变量变动方向相同

负相关：两列变量中有一列变量变动时，另一列变量呈现出与前一列变量方向相反的变动

零相关：两列变量之间没有关系，各自按照自己的规律或无规律变化

1. 积差相关

①数据要成对出现，即若干个体中每个个体都有两种不同的观测值，并且每队数据与其它对子相互独立

②两列变量各自总体的分布都是正态的，至少接近正态

③两个相关的变量是连续变量，也即两列数据都是测量数据

④两列变量之间的关系应是直线性的

（2）公式

$$r=\frac{SP}{\sqrt{SS_X\ SS_Y}}=\frac{\sum xy}{\sqrt{\sum x^2\cdot\sum y^2}}=\frac{\sum XY-\frac{\sum X\sum Y}{N}}{\sqrt{\sum X^2-\frac{(\sum X)^2}{N}}\cdot\sqrt{\sum Y^2-\frac{(\sum Y)^2}{N}}}$$

r 也就等于 X 和 Y 共同变化的程度除以 X 和 Y 各自变化的程度。

2. 等级相关

①当研究考察的变量为顺序型数据时，若原始数据为等比货等距，则先转化为顺序型数据②当研究考察的变量为非线性数据时

（2）公式

将原始数据转化为顺序型数据，仍然用 Pearson 相关公式计算即可。

3．肯德尔等级相关

（1）肯德尔 W 系数

也叫肯德尔和谐系数，原始数据资料的获得一般采用等级评定法，即让 K 个被试对 N 件实物进行等级评定。其原理是评价者评价的一致性除以最大变异可能性。

$$W = \frac{\sum_i R_2 - \frac{(\sum R)_2}{N}}{\frac{1}{12} K_2 (N_3 - N)}$$

R_i 代表评价对象获得的 K 个等级之和

N 代表等级评定的对象的树木

K 代表等级评定者的数目

4. 点二列相关与二列相关

（1）点二列相关

适用于一列数据为等距正态变量，另一列为离散型二分变量。

$$r_{pb} = \frac{\bar{X}_p - \bar{X}}{s_t} \cdot \sqrt{pq}$$

$\bar{X}_p$ 是与二分称名变量的一个值对应的连续变量的平均数

$\bar{X}_q$ 是与二分称名变量的另一个值对应的连续变量的平均数

p 与 q 是二分称名变量两个值各自所占的比率

s_t 是连续变量的标准差

（2）二列相关

适用于两列变量都是正态等距变量，但其中一列变量被人为地分成两类。

$$r_b = \frac{\bar{X}_p - \bar{X}}{s_t} \cdot \frac{pq}{y}$$

y 为标准正态曲线中 p 值对应的高度，查正态分布表能得到

5. Φ相关

适用于两个变量都是只有两个点值或只表示某些质的属性。

$$r_{\Phi} = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

其中 a、b、c、d 分别为四格表中左上、右上、左下、右下的数据

1. 概率：表明随即时间出现可能性大小的客观指标

概率的定义包含以下两种，当观测次数够多时他们是相等的。

后验概率：对随机事件进行 n 次观察，某一事件 A 出现的次数 m 与观测次数 n 的比值在 n 趋近无穷时所稳定在的常数 p

先验概率：在满足试验可能结果数有限且每一种结果出现的可能性相等的条件下，随机事件包含的结果数除以结果总数

2. 正态分布：当样本量足够大时，我们会发现生活中许多变量的分布都近似于正态曲线。

特点

①正态曲线的形状就像一口挂钟，呈对称分布，其均值、中数、众数实际上对应于同一个数值

②大部分的原始分数都集中分布在均值附近，极端值相对而言比较少

③曲线两端向靠近横轴处不断延伸，但始终不会与横轴向交

④正态分布曲线转化为 z 分数后人以 z 分数与零点对应曲线下面积固定

用法①依据 Z 分数求概率，即已知标准分数求面积 ②从概率求 Z 分数，即从面积求标准分数值 ③已知概率或 Z 值，求概率密度，即正态曲线的高

3. 二项分布：对于一个事件有两种可能 A 和 B，但我们对这一事件观察 n 次，事件 A 发生的总次数的概率分布就是二项分布

二项分布的均值为$\mu = pn$

方差公式为$\sigma 2 = npq$

标准差的公式为$\sigma = \sqrt{npq}$

4. 抽样原理与抽样方法

抽样原理：抽样的基本原则是随机性原则，所谓随机性原则，是指在进行抽样时，总体中每一个个体是否被抽选的概率完全均等。由于随机抽样使每个个体有同等机会被抽取，因而有相当大的可能使样本保持和总体有相同的结构，或者说，具有最大的可能使总体的某些特征在样本中得以发现，从而保证由样本推论总体。

抽样方法：①简单随机取样法　②系统随机取样法　③分层随机取样法

④多段随机取样法

5. 抽样分布　样本分布：样本统计量的分布，是统计推论的重要依据

（1）正态分布及渐近正态分布

样本统计量为正态分布或者接近正态分布的情况都可根据正态分布的概率进行统计推论。

总体分为正态或接近正态，方差已知，样本平均数和方差的分布为正态分布

①样本平均数分布的平均数和方差与母体的平均数和方差有如下关系：

$$\mu_{\bar{X}} = \mu$$

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

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

②样本的方差及标准差的分布也渐趋于正态分布，其分布的平均数与标准差和总体有如下关系：

$$\bar{X}_s = \sigma \qquad \sigma_s = \frac{\sigma}{\sqrt{2n}}$$

$$\bar{X}_{s^2} = \sigma^2 \qquad \sigma^2_s = \frac{\sigma^2}{2n}$$

（2）t 分布

t 分布是一种与方差无关而与自由度有关的分布，很类似正态分布，我们可以将正态分布看作 t 分布当自由度为正无穷时的特例。

总体分布为正态，方差未知时，样本平均数的分布为 t 分布：

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n-1}} \quad \text{其中} \quad s_{n-1} = \sqrt{\frac{SS}{n-1}}$$

（3）χ^2 分布

χ^2 分布的构造是从一个服从正态分布的总体中每次抽去 n 个随机变量，计算其平方和之后标准化的一个分布。分布曲线下的面积都是1，但伴随着n 取值的不同，自由度改变，曲线分布形状不同，而当自由度趋近于正无穷时 χ^2 分布即为正态分布，因此其于 t 分布一样都是一族分布，而正态分布都是其中的特例。

$$\chi^2 = \frac{\Sigma (X-\mu)^2}{\sigma^2}$$

（4）F 分布

如果有两个正态分布的总体，我们从其中各自取出两个样本，各自计算出 χ^2，则：

$$F = \frac{\chi_1^2 / df_1}{\chi_2^2 / df_2}$$

更多情况下，我们所计算的 F 两样本取自相同总体，此时可将上式化简为：

$$F = \frac{S^2_{\frac{n_1-1}{n_2-1}}}{S^2}$$

（二）参数估计

当在研究中从样本获得一组数据后，如何通过这组信息，对总体特征进行估计，也就是如何从局部结果推论总体的情况，称为总体参数估计。总体参数估计问题可以分为点估计与区间估计。

良好估计量的标准

①无偏性——用多个样本的统计量估计总体参数的估计值，其偏差的平均数为零

②有效性——当总体参数的无偏估计不止一个统计量时，无偏估计变异小者有效性高，变异大者有效性低，即方差越小越好

③一致性——当样本容量无限增大时，估计值应能够越来越接近它所估计的总体

参数

④充分性——样本的统计量是否充分地反映了全部 n 个数据所反映总体的信息

点估计：用样本统计量来估计总体参数，因为样本统计量为数轴上某一点值，估计结果也以一个点的数值表示

区间估计：根据估计量以一定可靠程度推断总体参数所在的区间范围，这个区间就叫做置信区间，相应的概率成为置信度，这两个量是共通变化的，置信区间越大，置信度越高；区间估计是用数轴上的一段距离表示未知参数可能落入的范围及落入该范围的概率。

标准误：样本平均数分布的标准差

总体方差未知时用估算的总体方差计算标准误。$\sigma_X = \frac{\sigma}{\sqrt{n}}$

2. 总体平均数的估计 $\bar{x} - Z_{\alpha/2}\sigma_X < \mu < \bar{x} + Z_{\alpha/2}\sigma_X$

当总体方差未知时，则使用 t 分布对应置信度

3. 标准差与方差的区间估计

（1）标准差的区间估计 $s_{n-1} - Z_{\alpha/2}\sigma_s < \sigma < s_{n-1} + Z_{\alpha/2}\sigma_s$

（2）方差的区间估计

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{(1-\alpha)/2}}$$

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
https://d.book118.com/166152054134010134