第四章　差异量数

一、差异量数

（一）全距

1.概述

全距又称两级差，指数据中最大值与最小值之差，用 R 表示。

2.评价

全距计算简单、容易理解；但全距不稳定、不可靠、不灵敏，也易受极端值的影响。

3.计算：$R=X_{max}-X_{min}$

（二）离差和平均差

1.离差

（1）概述

离差是某一数据与平均数的差，表示每一个观测值与平均数距离的大小，正负号说明了偏差的方向，所以，所有观测值离差的总和总是为 0。

（2）计算：$x=X-\mu$

2.平均差

（1）概述

平均差是所有原始数据离差绝对值的平均值。

（2）评价：平均差充分考虑了每个数值的离中情况，完整地反映了全部数值的分散程度，在反映离中趋势方面比较灵敏，计算方法也比较简单；但是它需要对离差取绝对值，不利于进一步做统计分析，较低效。

（3）计算

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

（三）方差和标准差

方差和标准差是表示一组数据离散程度的最好指标，其值越大，说明次数分布的离散程度越大。

1.方差

①含义

方差是每个数据与该组数据平均数之差平方后的均值，即离均差平方后的均数。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。

②特点

具有可加性和可分解性的特点。方差分析就是利用方差的这个特点，进一步说明各种变异对总结果的影响。

2.标准差

①含义

标准差是方差的平方根。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。

②特点

A.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。

B.每一个观测值都乘以一个相同的常数 C，所得标准差等于原标准差乘以这个常数。

C.每一个观测值都乘以一个相同的常数 C，再加上一个常数 d，所得标准差等于原标准差乘以这个常数 C。

③评价

A.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演算，较少受抽样变动的影响。

B.缺点：运算繁琐、易受极值影响、难理解。

④应用

切比雪夫定理，“正负三个标准差法则”

3.方差和标准差的计算

（1）方差

①未分组

A.基本式

$$s^2=\frac{\sum\left(X-\overline{X}\right)^2}{N}$$

B.原始式

$$s^2=\frac{\sum X^2}{N}-\left(\frac{\sum X}{N}\right)^2$$

②已分组

A.基本式

$$s^2=\frac{\sum f\left(X_C-\overline{X}\right)^2}{N}$$

B.估算式

$$s^2=\left[\frac{\sum fd^2}{N}-\left(\frac{\sum fd}{N}\right)^2\right]\times i^2\text{，}\ d=(Xc-AM)/i$$

③总方差合成

只有在应用同一种观测手段、测量的是同一个特质，只是样本不同时，才能应用总方差合成的公式。

$$s_T^2=\frac{\sum N_i\times s_i^2+\sum N_i\times d_i^2}{\sum N_i}$$

$d_i=\overline{X_T}-\overline{X_i}$ 其中，$\overline{X_T}$ 为总平均数，$\overline{X_i}$ 为各小组平均数

（2）标准差

$$s=\sqrt{s^2}\text{，}\quad \sigma=\sqrt{\sigma^2}$$

（四）百分位差和四分位差

1.百分位差

百分位差是指某一百分位数与另一百分位数之间的差值。常用的百分位差除 P_{90}-P_{10} 外，还有 P_{93}-P_{7}。

百分位差容易理解，易计算且较少受两极数值的影响；但不能反映出分布的中间数值的差异情况，稳定性较差（没有考虑全部数据）。

2.四分位差

四分位差是数据中间 50%数据的全距，常用 Q 来表示。它的值等于 P_{25} 到 P_{75} 距离的二分之一。

$$Q=\frac{Q_3-Q_1}{2}=\frac{P_{75}-P_{25}}{2}$$

二、相对量数

（一）差异系数

1.概述

（1）含义

差异系数又称变异系数、相对标准差等，它是一种相对量数，用 CV 表示，是标准差对平均数的百分比。

（2）适用条件

①同一团体不同观测值离散程度的比较。

②对于水平相差较大，但进行的是同一种观测的各种团体的观测值的离散程度的比较。

（3）使用须知

测量数据必须等距；测量工具具备绝对零；由于尚无有效的检验方法，目前不能进行推理统计。

2.计算

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

（二）标准分数

1.概述

（1）含义

以标准差为单位，表示个原始分数在团体中所处位置的相对位置量数，也叫 Z 分数或基分数。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。

（2）特点

若将一个分布中的所有原始分数都转化为 Z 分数，则所得的新分布就被称为 Z 分数分布，也称标准分布，这个过程叫标准化。标准分布有以下特点：

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所有原始分数的 Z 分数之和为 0，Z 分数的平均数也为 0。一组原始分数转换得到的

Z 分数可正可负。

③所有原始分数的 Z 分数的标准差为 1。

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

⑤原始分数转换为 Z 分数后，两者分布形状相同。

（3）评价

①优点

A.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。

B.可加性：标准分数能使不同性质的原始分数具有相同的参照点，因此可相加。

C.明确性：知道了标准分数，利用分布函数表就能知道该分数在全体分数中的位置。

D.稳定性：原始分数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的权重一样。

②缺点

A.计算繁杂。

B.有负值和零、有小数。

C.在进行比较时需满足数据原始形态相同这一条件。

（4）应用

①比较几个分属性质不同的观测值在各自数据分布中相对位置的高低。

②计算不同质的观测值的总和或平均值，以表示在团体中的相对位置。

③若标准分数中有小数、负数等不易被人接受的问题，可通过线性公式将其转化成新的标准分数（如韦氏成人智力量表）。

2.计算

$$Z=\frac{X-\overline{X}}{s}$$

（三）百分位数与百分等级

1.百分位数

（1）概述

百分位数是处于某一百分比的人对应的分数。在此分数以下，包括数据分布中全部数据个数的一定百分比。

第 P 个百分位数是指在其值 P 的数据以下，包括分布中全部数据的百分之 P。例如，第 50 位百分位数 P_{50} 等于 89 分，那么在 89 分以下就包括 50%的数据。

（2）计算

①基本式

$$P_p=L_b+\frac{i}{f}\times(\frac{P}{100}\times N-F_b)$$

P_P 为所求的第 P 个百分位数，L_b 为百分位数所在组的精确下限，f 为百分位数所在组的次数，F_b 为小于 L_b 的各组次数的和，N 为总次数，i 为组距。

②内插法

例如：某个测下限的测验中，小赵获得最高分为 695，其百分等级为 100；小磊获得最低分为 103 分，百分等级为 1，那么百分等级为 80 的百分位数 P_{80} 可如此计算：

$$\frac{100-80}{695-P_{80}}=\frac{80-1}{P_{80}-103}$$

2.百分等级

（1）概述

百分等级是低于某测验分数的人数百分比，它与百分位数互为逆运算。例如，89 分对应的百分等级为 50，则说明有 50%的数据小于 89 分。

（2）计算

①已分组

$$P_R=\frac{100}{N}\times[F_b+\frac{f(X-L_b)}{i}]$$

②未分组

$P_R=100-(100\times R-50)/N$，$R$ 为某人在测验中的排序。