

地理建模实验1 实验报告

42109232 吕文博 地信2101班

2024-05-13

未分组数据常见统计指标计算

地块编号	1	2	3	4	5	6	7	8	9	10	11	12
面积	12	83	50	35	55	50	72	40	85	29	65	75
统计指标	平均值	众数	中位数	极差	离差	离差平方和	方差	标准差	无偏估计 ，样本方	变异系数		
数值												
离差												

```
# 定义计算函数
comstatindic = \(x){
  meanx = mean(x)
  u = unique(x)
  tab = tabulate(match(x, u))
  modex = u[tab == max(tab)]
  medianx = median(x)
  rangex = max(x) - min(x)
  deviationx = x - meanx
  squre_deviationx = sum(deviationx ^ 2)
  variancex = squre_deviationx / length(x)
  sdx = sqrt(variancex)
  unbias_sdx = sqrt(squre_deviationx / (length(x) - 1))
  cvx = unbias_sdx / meanx
  dt1 = c(meanx,modex,medianx,rangex,squre_deviationx,
    variancex,sdx,unbias_sdx,cvx)
```

```
names(dt1) = c("平均值","众数","中位数","极差","离差平方和",
               "方差","标准差","标准差的无偏估计","变异系数")
return(list(dt1,"离差" = deviationx))
}
```

```
ug = c(12,83,50,35,55,50,72,40,85,29,65,75)
comstatindic(ug)
```

```
## [[1]]
##          平均值          众数          中位数          极差
##      54.2500000      50.0000000      52.5000000      73.0000000
##      离差平方和          方差          标准差 标准差的无偏估计
##      5666.2500000      472.1875000      21.7298757      22.6961150
##      变异系数
##      0.4183616
##
## $离差
## [1] -42.25  28.75  -4.25 -19.25   0.75  -4.25  17.75 -14.25  30.75 -25.25
## [11]  10.75  20.75
```

分组数据平均值,中位数,众数的计算

家庭收入分组/元	2000-3000	3000-4000	4000-5000	5000-6000	6000-7000	7000-8000	8000-9000
户数	300	1300	200	150	100	50	30
向上累计频率	300	1600	1800	1950	2050	2100	2130
向下累计频率	2130	1830	530	330	180	80	30
组中值	2500	3500	4500	5500	6500	7500	8500
	平均值	众数	中位数				
	平均值	众数	中位数				

```
# 首先构造输入数据
g = tibble::tibble(
  家庭收入分组 = paste(seq(2000,8000,by = 1000),
                        seq(3000,9000,by = 1000),
                        sep = "-"),
  户数 = c(300,1300,200,150,100,50,30),
  向上累计频率 = c(300,1600,1800,1950,2050,2100,2130),
  向下累计频率 = c(2130,1830,530,330,180,80,30),
  组中值 = c(2500,3500,4500,5500,6500,7500,8500))
```

```
)
g
```

```
## # A tibble: 7 x 5
##   家庭收入分组  户数 向上累计频率 向下累计频率 组中值
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 2000-3000      300          300          2130      2500
## 2 3000-4000     1300          1600          1830      3500
## 3 4000-5000      200          1800           530      4500
## 4 5000-6000      150          1950           330      5500
## 5 6000-7000      100          2050           180      6500
## 6 7000-8000       50          2100           80      7500
## 7 8000-9000       30          2130           30      8500
```

平均值计算

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

$x_i (i = 1, 2, \dots, m)$ 代表第 i 组的组中值, 如果第 i 组的下限值为 a_i , 上限值为 b_i , 则 $x_i = a_i + \frac{(b_i - a_i)}{2}$; f_i 代表第 i 组的频数, m 为分组个数。

```
# 计算分组数据的平均值
meang = sum(g$`组中值` * g$`户数`) / sum(g$`户数`)
cat("分组数据平均值为", meang)
```

```
## 分组数据平均值为 3899.061
```

众数计算 :

法1

$$M_o = L + d \times \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

法2

$$M_o = U - d \times \frac{\Delta_2}{\Delta_1 + \Delta_2}$$

M_o 为代求的分组数据的众数, L 为众数所在组的下限值, U 为众数所在组的上限值, Δ_1 为众数组频数与上一组频数之差, Δ_2 为众数组频数与下一组频数之差, d 为众数所在组的组距

```

calcul_mode = \(popn,floorn,ceiln){
  i = match(max(popn),popn)
  L = floorn[i];U = ceiln[i]
  delta1 = popn[i] - popn[i-1]
  delta2 = popn[i] - popn[i+1]
  d = U - L
  m1 = L + d * delta1 / (delta1 + delta2)
  m2 = U - d * delta2 / (delta1 + delta2)
  m = c(m1,m2)
  setNames(m,c('法1计算的众数','法2计算的众数'))
}

xfloor = seq(2000,8000,by = 1000)
xceil = seq(3000,9000,by = 1000)
xpop = g$`户数`
modeg = calcul_mode(xpop,xfloor,xceil)
modeg

```

```

## 法1计算的众数 法2计算的众数
##          3476.19          3476.19

```

中位数计算：

法1

$$M_e = L + d \times \frac{\frac{1}{2} \sum_{i=1}^n f_i - S_{m-1}}{f_m}$$

法2

$$M_e = U - d \times \frac{\frac{1}{2} \sum_{i=1}^n f_i - S_{m+1}}{f_m}$$

M_e 为代求的分组数据的中位数, L 为中位数所在组的下限值, U 为中位数所在组的上限值, n 为分组个数, f_i 为第*i*组对应的频数, f_m 为中位数所在组的频数, S_{m-1} 为中位数所在组以下的累积频数, S_{m+1} 为中位数所在组以上的累积频数, d 为中位数所在组的组距

```

calcul_median = \(popn,floorn,ceiln){
  halfpop = sum(popn) * 0.5
  i = which(cumsum(popn) - halfpop > 0)[1]
}

```

```

L = floorn[i];U = ceiln[i]
d = U - L
fm = popn[i]
sms1 = sum(popn[1:(i-1)])
sma1 = sum(popn[(i+1):length(popn)])
m1 = L + d * (halfpop - sms1) / fm
m2 = U - d * (halfpop - sma1) / fm
m = c(m1,m2)
setNames(m,c('法1计算的中位数','法2计算的中位数'))
}

xfloor = seq(2000,8000,by = 1000)
xceil = seq(3000,9000,by = 1000)
xpop = g$`户数`
mediang = calcul_median(xpop,xfloor,xceil)
mediang

```

```

## 法1计算的中位数 法2计算的中位数
##          3588.462          3588.462

```

所以分组数据的平均值为3899.0610329,众数为3476.1904762,中位数为3588.4615385.