心理统计

第五讲：z分数、概率、正态分布和二项分布

严超赣

Chao-Gan Yan, Ph.D.

yancg@psych.ac.cn
http://rfmri.org/yan

Institute of Psychology, Chinese Academy of Sciences

1

1

第四章

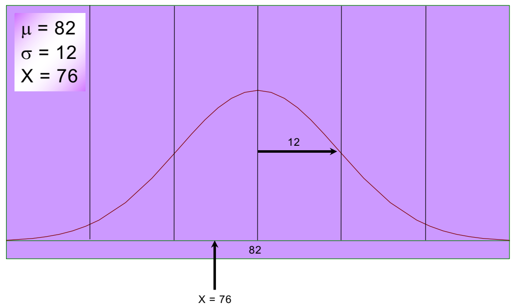
2

2

Z分数

3

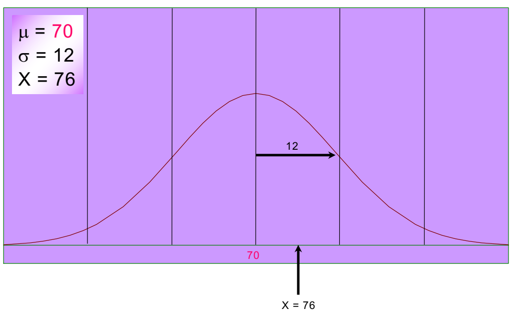
你如何解释76分?



4

4

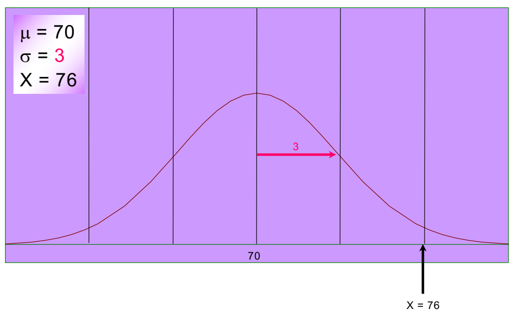
变换了均值的效果



5

5

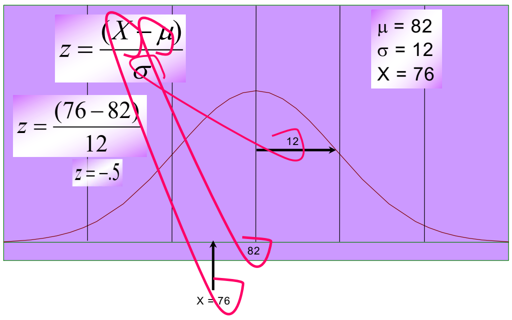
变换了标准差的效果



6

6

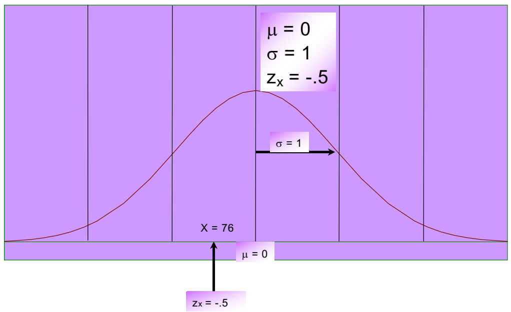
计算z分数(分布1)



7

7

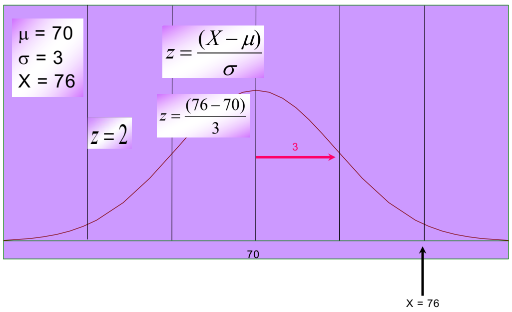
z 分布中的Z 分数(分布1)



8

8

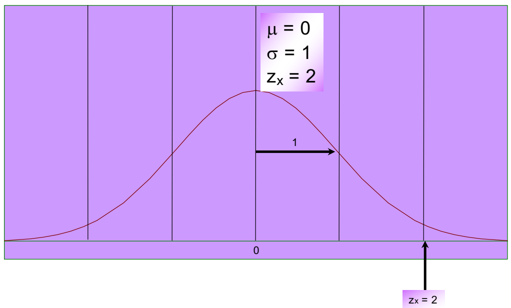
计算z分数(分布3)



9

9

z 分布中的Z 分数(分布3)



10

10

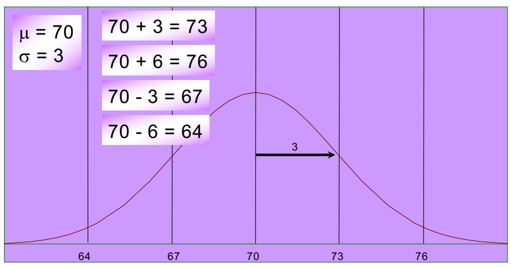
分布间的比较

μ	σ	X	z
82	12	76	-.5
70	12	76	.5
70	3	76	2

11

11

我们可以把分布按标准差分成6份



12

12

由Z分数转换回原始分数

- 如果分布中的总体/样本的均值和标准差已知，Z分数也可转换回原始分数。

$$z = \frac{X-\mu}{\sigma} \rightarrow X = z\sigma + \mu$$

- 例：如果某人说他的SAT分数高于均值 2 SD。他得了多少分？
  - mean = 500, SD = 100,
  - Xbar = (Z)(s)+ m = 2*100 + 500 = 200+500 = 700

13

z分数的另一用途——将整个分布标准化

- 如果我们将一个分布中的所有原始分数转化为z分数，所得的新分布就被称为z分数分布，也称标准分布（standardized distribution）

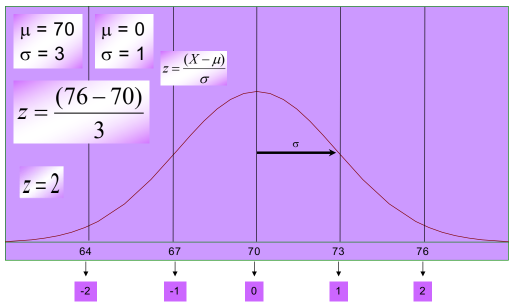
14

标准分布

- 标准分布具有以下一些特征：
  - 形状 - Z分数分布的形状与原始分数分布完全相同。每个分数所在的相对位置亦完全相同。
  - 均值- 当原始分数转换成Z分数, mean=0.
  - 标准差 -当原始分数转换成Z分数, standard deviation = 1.

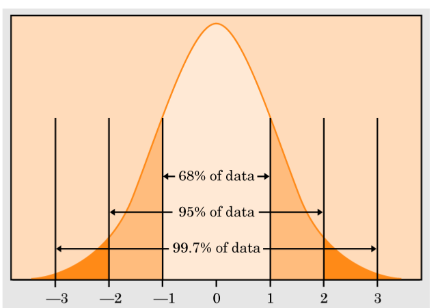
15

. . . 将它转换成z分布



16

即标准正态分布



17

概率 (Probability)

- 概率简介
- 概率与正态分布
- 概率与二项分布

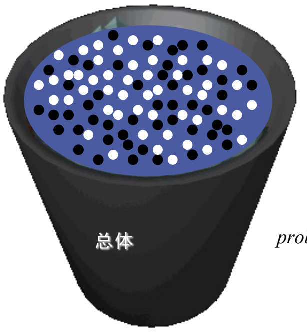
18

概率

- 推论统计所必需的概念，根据样本的信息对总体作出判断。
- 概率在日常生活中的例子
  - 彩票，天气预报，患病风险率

19

19



50 黑棋子

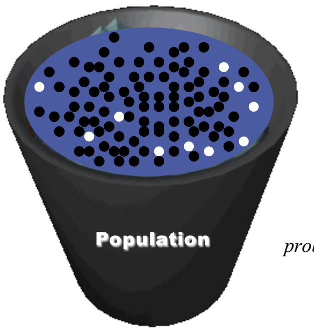
50 白棋子

$$probability(A) = \frac{\# A \ \ outcomes}{total \ \# outcomes}$$

$$p(black) = \frac{50}{100} = \frac{1}{2} = 50\%$$

20

20



90 黑棋子

10 白棋子

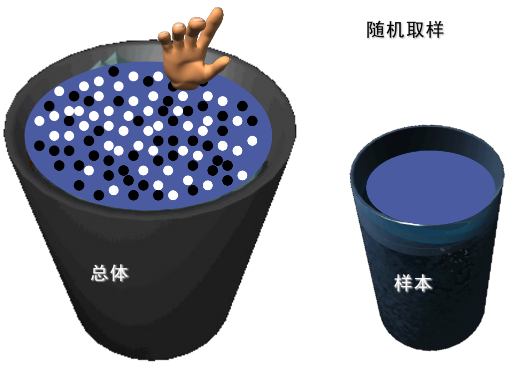
$$probability(A) = \frac{\# A \ \ outcomes}{total \ \# \ \ outcomes}$$

$$p(black) = \frac{90}{100} = \frac{9}{10} = 90\%$$

21

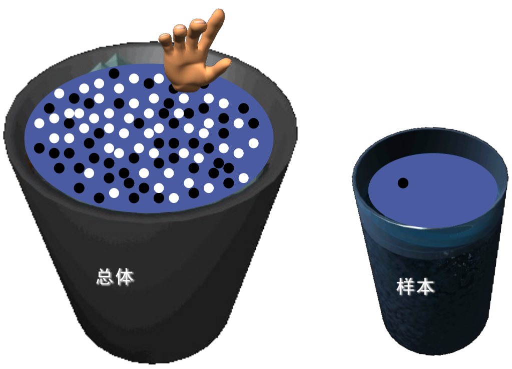
21

随机取样



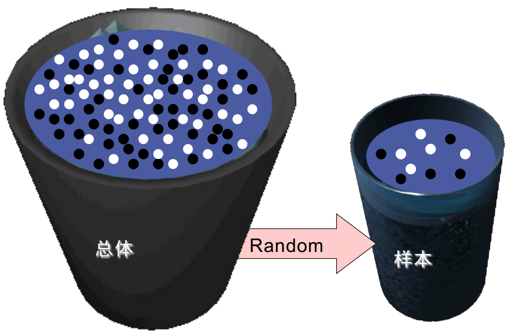
22

22



23

23



24

24

概率与随机取样（random sampling）

- 在次数分布表中. F/N曾用来计算比率. 而概率也常常表达为比率的形式 (或百分比或分数).
- 但是, 为获得正确定义的概率, 个体的选取 (取样) 一定要通过随机取样
- 随机取样应满足以下两个条件:
  - 1. 总体中的每个个体有同样的机会被选择
  - 2. 如果样本中要选择多于一个的个体，每次选择的概率应当恒定

25

25

回置取样

- **回置取样** （Sampling with replacement） - 一种取样方法， 在选择下一个个体（下次取样之前），将每个已选择个体放回总体之中

26

26

在推论统计中，我们从如下样本开始

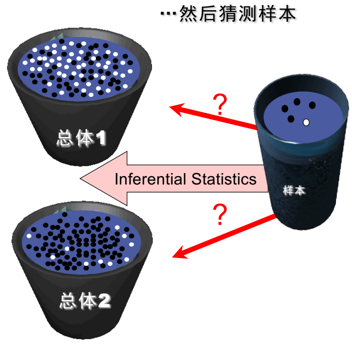
$$probability(A) = \frac{\#A \ \ outcomes}{total\ \# \ \ outcomes}$$

$$p(black) = \frac{4}{5} = 80\%$$

27

27

…然后猜测样本

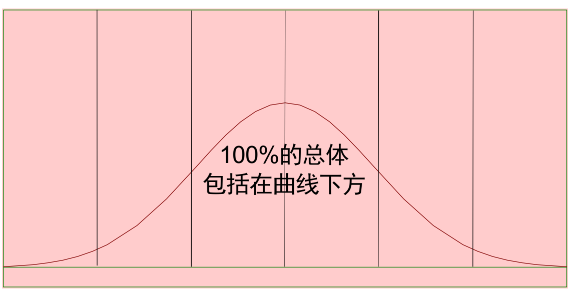


是来自哪个总体?

28

28

正态曲线



29

29

正态曲线

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

1. 正态曲线的形状肖似一口挂钟，呈对称分布。其均值、中数、众数对应于同一个数值。
2. 大部分的原始分数都集中分布在均值附近，极端值相对而言是比较少的。
3. 曲线两端向靠近横轴处不断延伸，但始终不会与横轴相交。正态分布是最常见的分布， 单峰和具对称性.

30

30

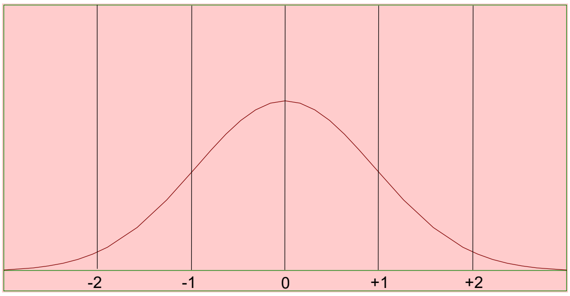
正态分布的注意点

- 并非所有的单峰，对称曲线都是正态分布，但很多是
- 在本课程中，无须担心所研究分布与正态分布有多接近，在本课程所遇到的问题中，多数情况下，分布是正态
- 上述的平滑的曲线是指密度曲线（而并非次数曲线）
- 曲线下方的面积总和必定为1．因为曲线下方的面积相当于概率（或比率）总概率应当等于 1.

31

31

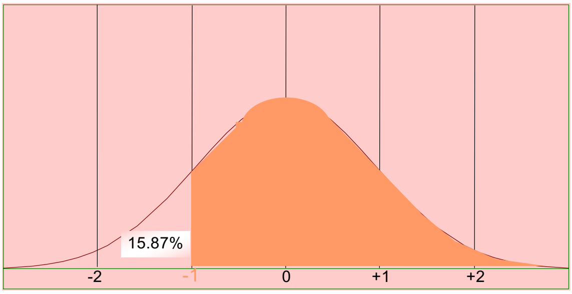
单位正态曲线



32

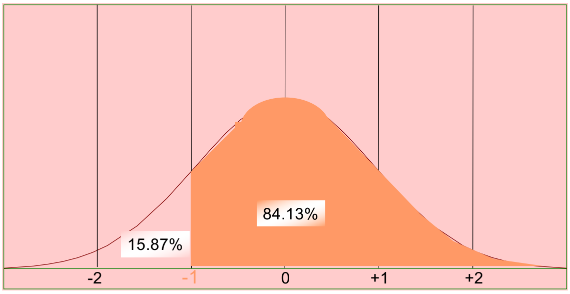
32

可以用单位正态曲线来确定百分位数



33

. . . 以及比这个分数高的分数比率



34

34

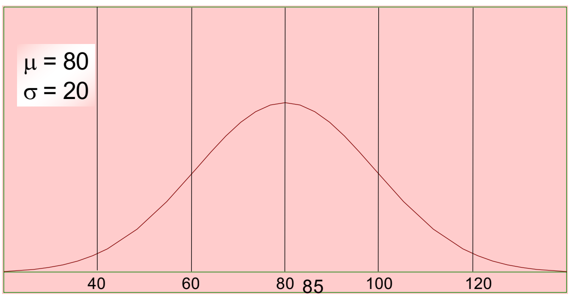
例题

- 在一个μ = 80 ,σ = 20 的正态分布中,找到低于85的百分比

35

35

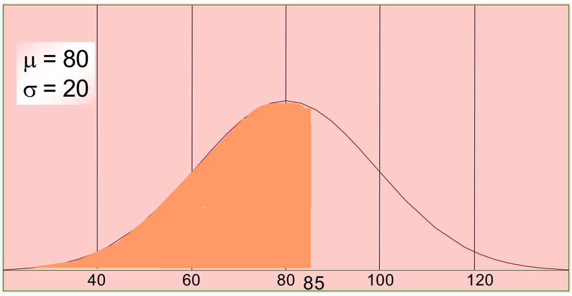
Step 1 – 绘出分布:



36

36

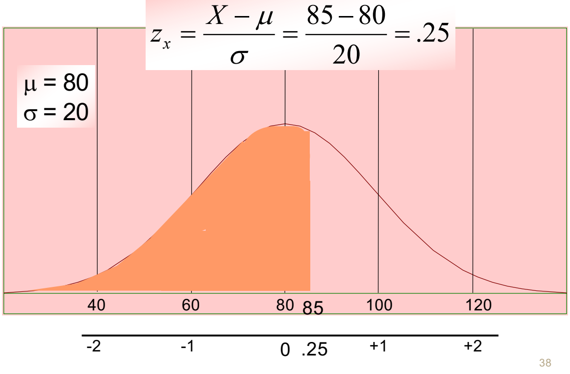
Step 2 – 将分布的要求解部分加重



37

Step 3 – 将X 转换成 z

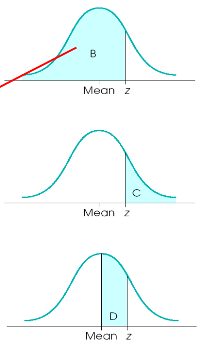
$$z_x = \frac{X-\mu}{\sigma} = \frac{85-80}{20} = .25$$



38

Step 4 – 查正态分布表

(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z
0.00	.5000	.5000	.0000
0.01	.5040	.4960	.0040
0.02	.5080	.4920	.0080
0.03	.5120	.4880	.0120
0.21	.5832	.4168	.0832
0.22	.5871	.4129	.0871
0.23	.5910	.4090	.0910
0.24	.5948	.4052	.0948
0.25	.5987	.4013	.0987
0.26	.6026	.3974	.1026
0.27	.6064	.3936	.1064
0.28	.6103	.3897	.1103
0.29	.6141	.3859	.1141
0.30	.6179	.3821	.1179
0.31	.6217	.3783	.1217
0.32	.6255	.3745	.1255
0.33	.6293	.3707	.1293
0.34	.6331	.3669	.1331



39

Step 5 – 确定相应的概率

59.87% 或 60% = 低于85的百分比.

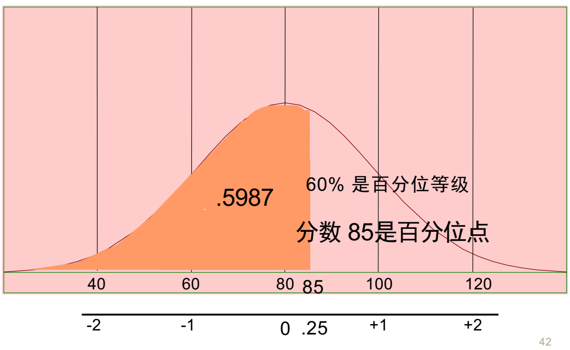
40

用z分数求百分位数的一般步骤

1. 绘出分布,包括均值,标准差和有关的分数. 将分布的要求解部分加重.
2. 将分数转换成 z分数.
3. 查正态分布表
4. 确定相应的概率.

41

确定百分位点和百分位等级



42

插值法示例

- 从概率查到相应的 z-分数. 常常用到插值法

  例：相当于人群顶端 5%的IQ 是多少？

43

43

插值法示例

- 从概率查到相应的 z-分数. 常常用到插值法

  例：相当于人群顶端 5%的IQ 是多少？

  p = 0.05　　z=?

  查表p = 0.05：

  Z = 1.64　p = 0.0505

  Z = ?　　p = 0.05

  Z = 1.65　p = 0.0495

  Z = 1.645 ≈ 1.65

  故 X = (1.65)(15) + 100 = 124.75

44

44

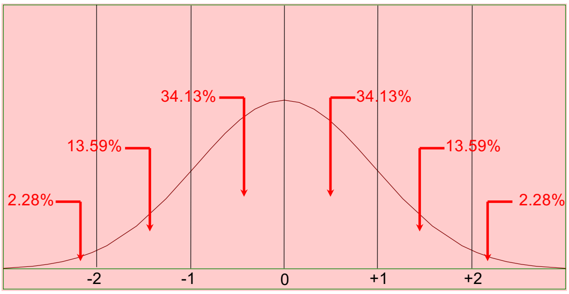
示例

例2：IQ为130的百分位等级是多少？

45

45

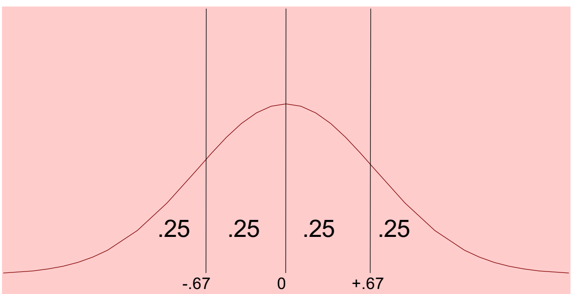
确定正态分布中的固定比率



46

46

确定四分位距



47

47

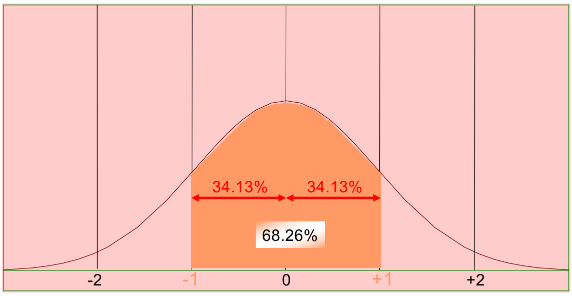
正态分布与概率

- 95% 的分数会落入-1.96与1.96标准差之间.
- 95% 的分数会落入1.65标准差以左.
- 99% 的分数会落入-2.58与2.58标准差之间
- 99.9%的分数会落入-3.30与3.30标准差之间

48

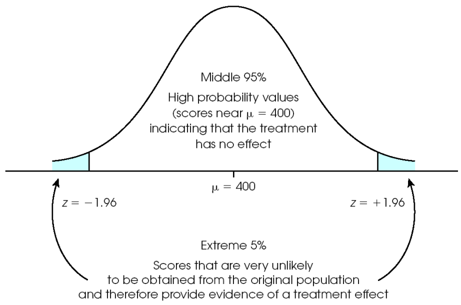
48

以便使我们能够确定在z分数界限里的人群百分比



49

. . . 以及尾端的比率



50

二项分布
（Binomial Distribution）

51

概念

- 如果在某种特定的情境下，只有两种可能的结果．其结果就形成一个二项分布
- 例如，投掷硬币得到正面或反面，对是否题的回答，一个人的生或死等等
- 二项分布表示为：B(n,p)，其方程非常复杂。
- 如果 n 足够大，二项分布可以近似为正态分布。

52

二项分布的概率

- 两个类目：A 和 B
  p = p(A) = A 的概率
  q = p(B) = B 的概率
- p + q = ?? ----> 1.0
- n =样本中所包含个体（或观察）的数目
  X =样本中事件类目 A 发生的数目
- 二项分布表达了与从X = 0 到 X = n 的每一个X值有关的概率。

53

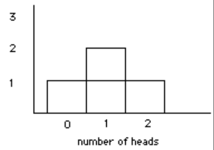
例 1　掷硬币. A = 正面; B = 反面

p = p(A) = ½; q = p(B) = ½

假设 n = 2 (即, 将硬币掷 2 次), 有多少可能的结果

- B(2, 0.5)? -4 ·

第 1次	第 2次	正面次数
正面	正面	2
正面	反面	1
反面	正面	1
反面	反面	0



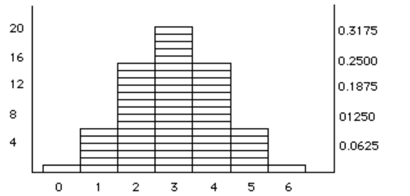
54

- 两次掷到正面的概率是多少?
- 掷不到正面的概率是多少?
- 只一次掷到正面的概率是多少?
- 至少一次掷到正面的概率是多少??
- 假设 n = 6. 有多少可能的结果? 64 种 （公式是: 2^n）

55

55

t1	t2	t3	t4	t5	t6	#正面
正面	正面	正面	正面	正面	正面	6
正面	正面	正面	正面	正面	反面	5
正面	正面	正面	正面	反面	正面	5
正面	正面	正面	正面	反面	反面	4
:	:	:	:	:	:	:
反面	反面	反面	反面	反面	反面	0

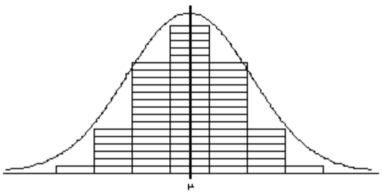


56

56

什么条件下，二项分布可以近似为正态分布？

- 看n = 6 时的情况 (pn = .5*6 = 3).



当 n 足够大 (pn > 10) 和(qn > 10),二项分布可以近似为正态分布.

57

57

二项分布的均值和标准差

Mean: $\mu = pn$

Standard deviation: $\sigma = \sqrt{npq}$

$$z = \frac{X-\mu}{\sigma} = \frac{X-pn}{\sqrt{npq}}$$

58

58

二项分布的均值和标准差

Mean

$$\boxed{\mu = E(x) = np}$$

Variance and Standard Deviation

$$\boxed{\sigma^2 = npq}$$

$$\boxed{\sigma = \sqrt{npq}}$$

n = sample size
p = probability of A
q = (1 – p)

59

59

利用正态分布表求二项分布的概率

- 正态分布中X 的值是一段，而并非一点，所以当二项分布近似为正态分布时，需要考虑精确上下限。因为我们是在用连续型分布（正态）来估计离散型分布的值.

60

60

例子

- 例2：某职业学校，有时学生入学后会中途退学．如果每个人中途退出的概率是0.10，在100人的班上，
  1. 有不少于15个学生退学的概率是多少？
  2. 有多于15个学生退学的概率是多少？

61

61

例子

- 例3：有时学生入学后会中途退学．如果每个人中途退出的概率是0.10，在100人的班上，
  1. 有不少于15个学生退学的概率是多少？
  2. 有多于15个学生退学的概率是多少？

n = 100　p = 0.10　q = 0.90

np = .10*100 = 10　nq = 90

m_x = pn = 10

s_x = sqrt(npq) = sqrt (100*.10*.90)

= sqrt (9) = 3

1. p(X ≥ 15的精确下限) = P(X ≥ 14.5)

= P(Z ≥ 14.5-10) / 3.0= P(z ≥ 1.5) = 0.0668

2. p(X ≥ 15的精确上限) = P(X ≥ 15.5)

= P(Z ≥ 15.5-10) / 3.0= P(z ≥ 1.833) = 0.0335

62

62

例子

例4：假设你参加一个48道题的多项选择题测验，只有4种可能的答案．你全凭猜测作答．猜对14 道题的概率是多少？

63

63

例子

例4：假设你参加一个48道题的多项选择题测验，只有4种可能的答案．你全凭猜测作答．猜对14 道题的概率是多少？

p = P(正确) = 1/4

q = P(错误) = 3/4

pn = (1*48)/4 = 12

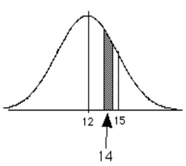
qn = (3*48)/4 = 36

- pn 和qn 都大于10，所以可以假定分布近似正态.
- 分数 14 其实是对应 从 13.5 到 14.5 之间这段距离

64

64

例子



$m = pn = 12$

σ = sqroot (*pqn*) = sqr(48*.25*.75) = sqroot (9) = 3

查表

$\frac{X - \mu}{\sigma} = \frac{13.5 - 12.0}{3} = 0.50$ --> 0.3085

$\frac{X - \mu}{\sigma} = \frac{14.5 - 12.0}{3} = 0.83$ --> 0.2033

两个 z分数间的面积是：0.3085 - 0.2033 = 0.1052

65

65

作业

有一次心理测验的成绩（成绩分布的总体为正态分布）μ= 80，σ= 8. 此测验中，Tom 得分 X=84， Mary得分在第 60个百分点上， John的得分换算成 z分数是 z=0.75。将此三人的分数从高到低排序。

66

66

作业

指出一个正态分布中位于下列z分数区间的概率:

a) z = 0.25 —— z = 0.75

b) z = -1.00 —— z = 1.50

c) z = -0.75 —— z = 2.00

67

67

作业

有一正态分布 μ= 75， σ= 9，指出下列情况发生的概率：

a) 该分布中某一样本值小于80的概率，即X<80

b) 该分布中某一样本值小于94的概率，即X<94

c) 该分布中某一样本值大于63，且小于88的概率，即63<X<88

d) 从中随机取出一个分数，其值处于72-78之间的概率。

68

68

作业

一个正偏态的分布均值为 100， 标准差 12. 从中随机抽取一个分数，其值大于106的概率是多少？

69

69

作业

一个特殊制作的硬币正面向上的概率是0.8, 反面向上的概率是0.2.

a) 如果掷硬币100次，正面向上的平均期望值是多少？

b) 如果掷硬币100次，有95次以上正面向上的概率是多少？

c) 如果掷硬币100次，有95次以下正面向上的概率是多少？

d) 如果掷硬币100次，有正好95次正面向上的概率是多少？

70

70

作业

一个是非判断测验有36道题。如果答对24题或以上算及格，单凭猜测获得及格或以上的概率是多少？

71

71