心理统计

第六讲：样本均值的分布 假设检验初步

严超赣

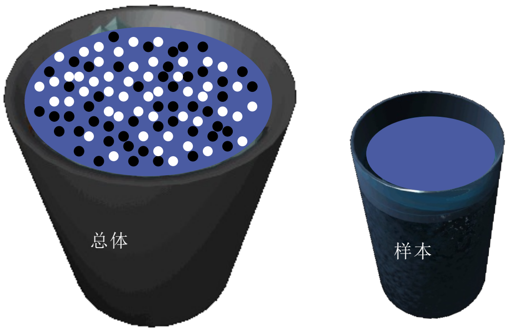
Chao-Gan Yan, Ph.D.

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

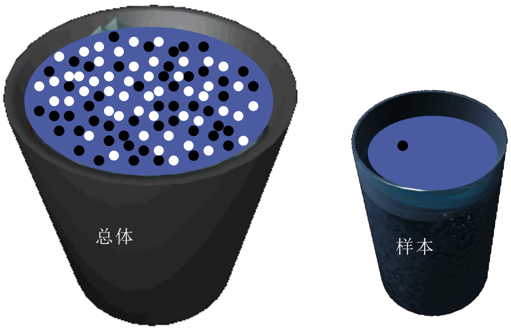
Institute of Psychology, Chinese Academy of Sciences

1

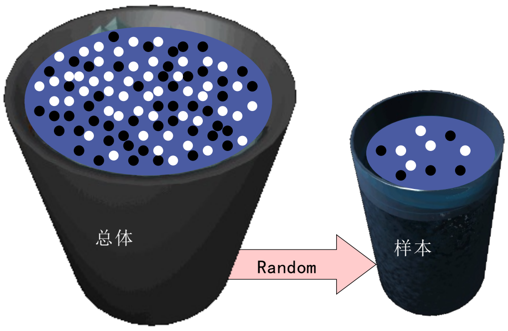
随机取样



2



3



4

在推论统计中:

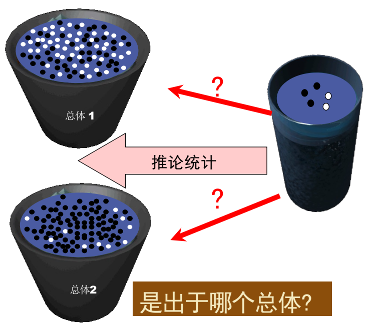
我们从下列样本开始

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

$$p(black) = \frac{3}{5} = 60\%$$

5

…猜一猜样本



6

从样本开始

计算概率

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

$$p(black)=\frac{3}{5}=.60$$

是样本的均值

7

样本均值的含义

1 ---黑子, 0 ---白子.

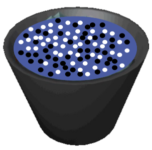
# 1 黑子:	1
# 2 黑子:	1
# 3 白子:	0
# 4 黑子:	1
# 5 白子:	0

S = 3

M = 3/5 = .60

8

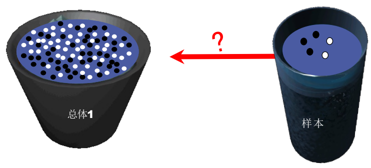
虚无分布是 50%



50% 是黑子

9

我们收集了5个棋子作样本



总体 μ = .50 黑子

样本 M = .60黑子

这个样本是总体的一个随机样本吗?

10

- 这个问题并不简单
- 因为样本会变化, 有时会包含一些极端值.
- 问题的关键在于60%的黑棋子这样的事件发生的概率有多小.

11

样本均值分布

我们需要一个新的分布

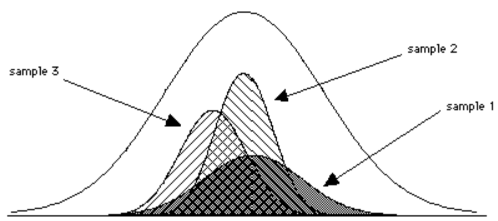
12

取样分布和样本均值的分布

- 样本均值分布（distribution of sample mean）：
  - 总体中可抽取的所有可能的特定容量(n)的随机样本的均值的集合。
- 抽样分布 （sampling distribution ）：
  - 总体中可抽取的所有可能的特定容量分布的统计量所形成的统计分布。
  - 样本均值分布是抽样分布的特例
- 为了对总体均值作出最佳估计，我们所要做的就是考察所有可能的样本（n一定，这点很重要）然后根据其特性作出预测。

13

从同一总体取3次不同样本 （n=100）



- 每一个都不同．不同形状，不同均值，不同方差．如何对总体均值作出最佳估计？
- 可能从该总体取多少个样本？

14

5个样本均值分布

- 样本容量分别是n=1，n=2，n=4，n=6，n=10(色子的只数）
- 对于每个样本，建立100个分布(掷色子的次数）.
- 真正的样本均值分布由无限个分布的均值构成，所以100次只是一个粗略的近似
  - 每个色子代表一个被试.
  - 掷色子的点数(1-6)代表因变量指标，如自尊分数
  - 掷色子代表随机选取被试.
- 计算每个样本的均值.
- 这样我们建立了5个样本均值分布

15

样本均值分布的均值

- 样本均值分布的均值是这100个样本的均值的平均数.

$$M_M = \frac{\sum X}{N} = \frac{2.8 + 3.0 + 2.4 + etc.}{100}$$

- 样本均值分布的期望值应当与总体均值相等（3.5）.

16

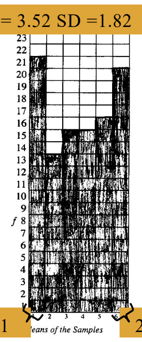
计算标准差

$$\sigma = \sqrt{\frac{\sum\left((2.8-\mu)^2 + (3.0-\mu)^2 + ...\right)}{N}}$$

17

n=1

M = 3.52 SD =1.82



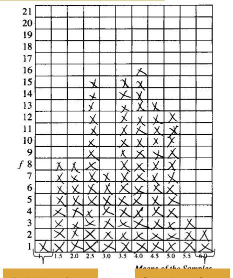
21 M = 1

20 M =6

18

n=2

M = 3.51 SD =1.2



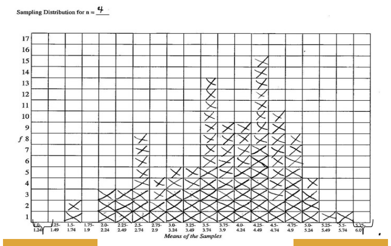
1 $M = 1$

2 $M = 6$

19

n=4

M = 3.7 SD = 0.88



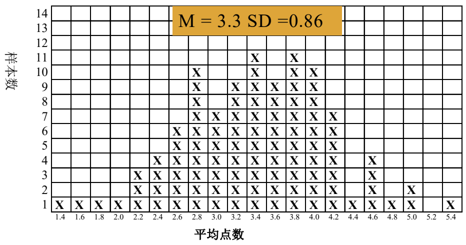
0 $M = 1$

0 $M = 6$

20

n=5

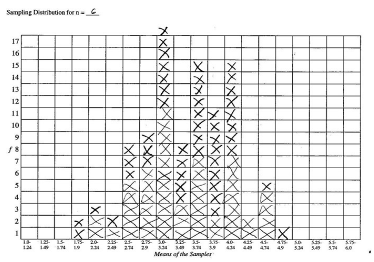
M = 3.3 SD =0.86



21

Sample Size n=6

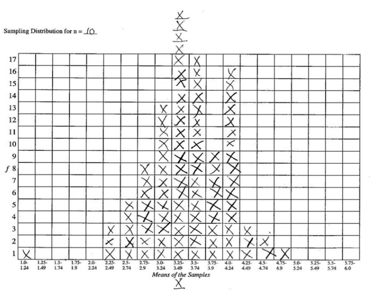
M = 3.4 SD =0.66



22

n=10

M = 3.46 SD =0.58



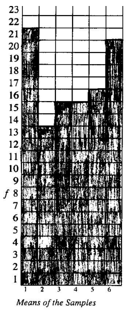
23

色子实验的小结

	Mean	SD	Minimum	Maximun
n = 1	3.52	1.82	1.0	6.0
n = 2	3.51	1.20	1.0	6.0
n = 4	3.71	0.88	1.5	5.5
n = 6	3.38	0.66	1.8	4.8
n = 10	3.46	0.58	1.4	4.9
总体	3.50	1.71	1.00	6.00

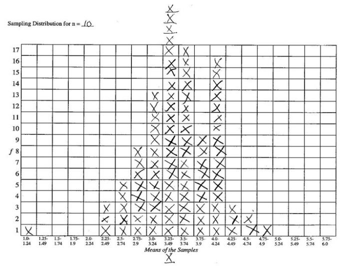
24

n = 1 近似矩形



25

n = 10 就非常正态了



26

例子

- 我们作了一个犯罪青少年自尊的调查发现均值为2.3.
- 我们想知道犯罪青少年自尊低吗?
- 我们知道一般青少年自尊均值为3.5 －这是总体均值.
- 均值为2.3是否低到我们认为它不是来自于一般青少年这一总体?
- 为回答这一问题,我们需要两章的知识,样本均值分布是基础

27

样本均值分布的特性

28

样本均值分布的特性1：形状

- 样本均值的分布形状接近正态分布.
- 当 n 较大时(30 以上)，样本均值的分布几乎是完全的正态分布．而不在乎原始分布的形状

29

样本均值分布的特性 2：均值

- 如果在同一总体中选择一组样本，大部分均值应当堆积在总体均值 μ 附近
- 如果不是这样，取样一定有偏差
- 这些样本均值的平均应该等于总体均值．样本均值的平均叫做 X 的期望值．期望值的意思因为这个值会在总体均值 μ 的附近.

30

样本均值分布的特性 3：变异性

- 样本均值分布的变异性 用标准误来描述
- 定义：样本均值分布的标准差 叫做 X 的标准误 （standard error of X ; SE）

  X 的标准误= $\sigma_{\bar{X}}$ = X 与 μ 的标准距离.

  - 这个统计量描述了与均值的标准(或称典型，平均)距离. 在这里，它也是样本均值 X 和 总体均值 μ 的差值.
  - 这个统计量的主要目的和用途是告诉我们样本均值对总体均值的估计是否准确. 换言之，取样误差是多大.
- 标准误的数值取决于两个特征：总体方差和样本容量

31

标准误

32

标准差和标准误

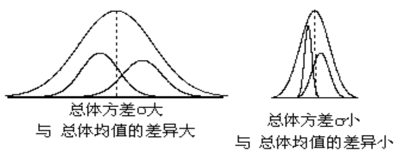
- 标准差:
  - 指总体中一个分数与总体均值的标准距离
  - 是衡量次数分布变异性的一个指标
- 标准误
  - 指一个样本均值和其相应的总体均值之间的标准距离
  - 衡量的是样本均值分布的变异性
  - 当n=1时,

$$\text{标准误} = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{1}} = \sigma = \text{标准差}$$

33

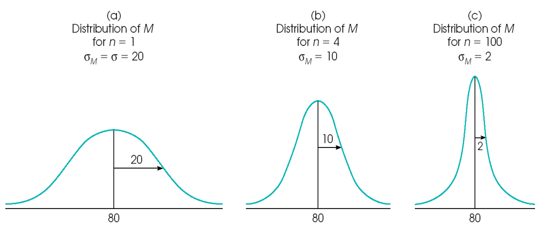
总体方差与标准误

- 总体方差越大， 样本均值的方差越大.



34

样本容量与标准误



35

标准误的定义公式

- 将上述两个特征合并起来，就是标准误的定义公式.：

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

36

中心极限定律 (Central Limit Theorem)

- 综合了样本均值分布的特性：形状，均值，方差
- 定律：对于任何均值为μ，标准差为σ的总体，样本容量为n的样本均值的分布，随着n趋近无穷大时，会趋近均值为μ，标准差为 $\sigma/\sqrt{n}$ 的正态分布
- 因此，当 n 足够大时(30或以上)：

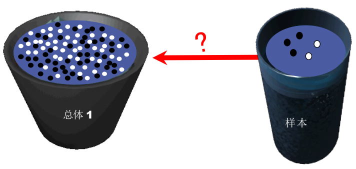
$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

37

样本均值分布与概率

38

我们可以解决这个问题了吗?



总体 μ = .50 黑子
σ = 0.5

样本 M = .60黑子

问题是:这个样本是总体的一个随机样本吗?

39

计算z分数

$$z = \frac{M-\mu}{\sigma_M}$$

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{.5}{2.24} = .22$$

n=2上限：2.5/5=0.5；n=2下限：1.5/5=0.3
Z1=(0.5-0.5)/0.22=0
Z2=(0.3-0.5)/0.22=0.91
Z1对应的概率0.50
Z2对应的概率0.1814
2颗白子的概率0.3186

40

样本均值分布与概率

例 1：一位老师对班上学生的IQ感兴趣. 她班上有9位学生，她认为他们都很聪明. 这班学生IQ 的均值大于等于112的概率是多少?

IQ test: μ = 100, s = 15

1. 我们需要知道样本的分布（注意：即使n 小于 30，我们仍然假定正态分布.）

   X ~ N（μ, σ/(sqrt(n)）= N(100,5)

2. 我们需要知道对应这个样本均值112的z 分数

   Z_X = (112 – 100)/ 5 = 2.4

   P(X ≥ 112) = P(Z ≥ 2.4) = 0.0082

41

例子

- 例2:如果班上有25位学生，如果让其样本均值位于顶端10%的IQ分布，均值应该有多大?

42

例子

- 例2:如果班上有25位学生，如果让其样本均值位于顶端10%的IQ分布，均值应该有多大?
  - 我们需要知道样本的分布 (注意: 即使n 小于 30，我们仍然假定正态分布.)

  $X \sim N(\mu, \sigma/(sqrt(n)) = N(100, 3)$
  - 然后找出对应于这个部分的均值：

  $X = Z_X * \sigma_X + \mu$

  查正态分布表：90% 概率对应的 Z 分数是1.28

  $X = 1.28 * \sigma_X + 100 = (1.28)(3) + 100 = 103.84$
  - 所以, 对于 25 个人的样本, 他们的均值必须在103.84 以上才能位于分布顶端的10%

43

假定上例中样本较小， n = 16 答案会不会改变?

44

假定上例中样本较小， n = 16 答案会不会改变?

- 查正态分布表：90% 概率对应的 Z 分数
- $X = 1.28 * \sigma_X + 100 = (1.28)(3.75) + 100 = 104.80$
- 所以, 对于16个人的样本, 他们的均值必须在104.80以上才能位于分布顶端的10%

45

上例中对于不同的样本容量

n = 9

$X = 1.28 * 15/sqrt(9) + 100 = (1.28)(5) + 100 = 106.40$

- n = 4

$X = 1.28 * 15/sqrt(4) + 100 = (1.28)(7.5) + 100 = 109.60$

- n = 1

$X = 1.28 * 15/sqrt(1)+ 100 = (1.28)(15) + 100 = 119.20$

- 注意: 如果 n = 1, 标准误等于总体标准差
- 所以，样本容量越小, 取样误差 (标准误) 越大.

46

样本稳定性与样本容量

- 在不同容量的样本中添加或去掉一个分数或改变某一分数对总体影响如何呢?
- 总体 $X \sim N(50, 10)$
- 比较这两个样本:
  - 样本 1: $X_1 = 50$, n = 4

  $\sigma_X = \sigma/(sqrt(n)) = 10/2 = 5$
  - 样本 2: $X_2 = 50$, n = 100

  $\sigma_X = \sigma/(sqrt(n)) = 10/10 = 1$

  假定我们在这两个样本中分别添加一个新的分数20

  样本一新的均值是：（50×4＋20）/5＝44；

  样本二新的均值是：（50×100＋20）/101＝49.7

47

假设检验初步(I)

- 假设检验的定义
- 示例
- 假设检验的步骤
- I类错误和II类错误

48

假设检验

- 假设检验 （Hypothesis testing）是一种用样本数据来评价有关总体的某一假设的可置信性的推论程序。
- 以样本为依据，对总体作出判断.

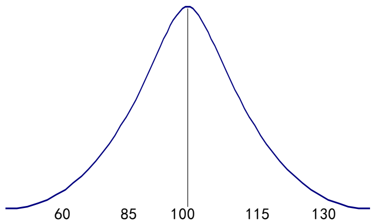
49

假设检验

- 我们从一个假设开始
  少年大学生的空间想象力与一般大学生不一样吗?
- 我们取了样本.
  N = 25， M = 110， SD = 12
- 我们知道对于总体
  μ = 100， σ = 15
- 我们的假设有没有得到验证?

50

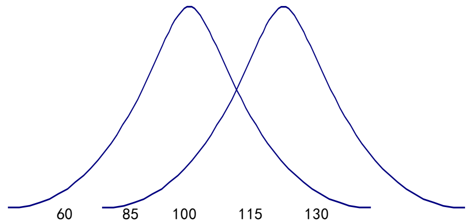
样本是从虚无总体中抽取的吗?



总体分布

51

. . . 或者是从另一总体中抽取?

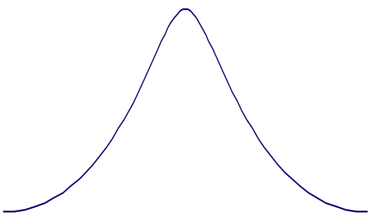


一般大学生

少年大学生

52

我们对另一总体并不了解

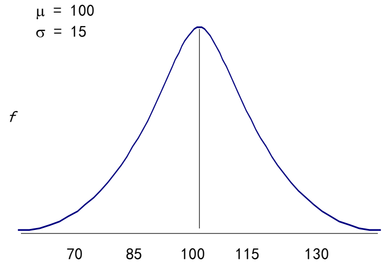


少年大学生

53

只了解虚无总体

总体参数:
μ = 100
σ = 15



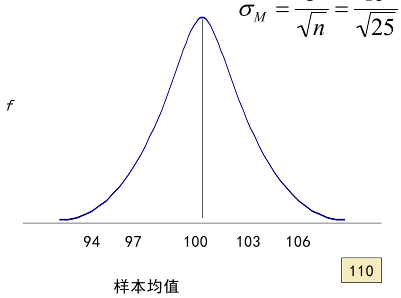
分数

54

样本均值分布

均值 = μ = 100

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = 3$$



55

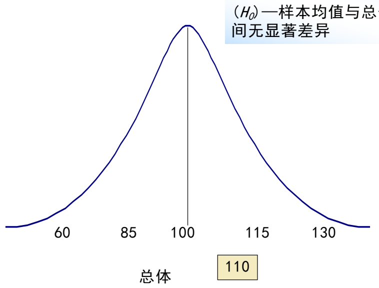
因此我们猜到结论:

- 这个样本不是来自这个总体.
- 换句话说-少年大学生与一般大学生的空间想象能力不同.
- 少年大学生的较高.

56

Step 1: 陈述假设

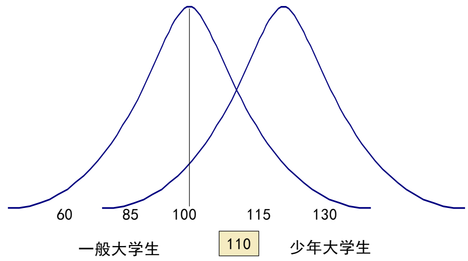
虚无假设(null hypothesis) (H_0)—样本均值与总体均值间无显著差异



57

Step 1: 陈述假设

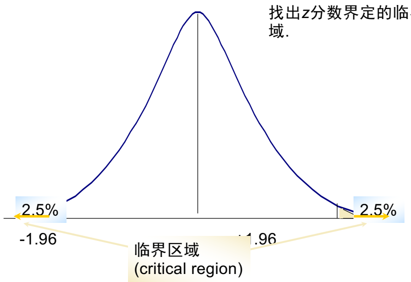
研究假设(Alternate hypothesis) (H_1)--样本均值与总体均值间存在显著差异



58

Step 2: 设定决策标准

找出z分数界定的临界区域.



59

Step 3: 收集数据, 计算样本统计量

$$n = 25$$
$$M = 110$$
$$SD = 15$$

$$Z_M = \frac{M - \mu}{\sigma_M} = \frac{110 - 100}{3} = \frac{10}{3} = 3.33$$

60

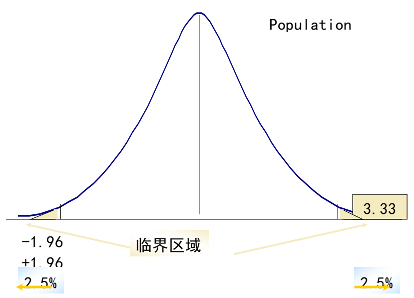
Step 4: 作出决策

Obtained $z_M = 3.33$
Critical $z = \pm 1.96$

因此, z_{ob} 落在拒绝区域中.

61

Step 4: 作出决策



62

步骤总结

1. 陈述虚无假设和研究假设.
2. 设定决策标准找出临界.
   如a (alpha) =.05
3. 收集数据, 计算样本统计量.
4. 作出决策.
   如果得到的z分数落在拒绝区域中, 就拒绝H_0

63

论文的写作

- 我们得到结论样本均值与总体均值有真实的差异.
  - 统计术语: 拒绝 H_0
  - 论文写作: 样本均值与总体均值有显著差异, $z = 2.11$, $p < .05$.
- 我们得到结论样本均值与总体均值没有不同.
  - 统计术语: 接受 H_0
  - 论文写作: 样本均值与总体均值没有显著差异, $z = 1.11$, $p > .05$.

64

事先设定拒绝虚无假设的标准

- 例如, 有一个总体 $\mu = 65$ $\sigma = 10$, 取到一个均值是80的样本 (容量为 100) 的概率是多大?
  - "取到一个均值是80的样本的概率是0.00002. 这已经很小了. 我有把握说这个样本不是从这个总体中来, 它是来自另一个总体。"
- 事先设定一个标准会为我们说"这已经很小了"提供更客观的依据. 这个事先设定的标准会限定概率有多小才足够拒绝虚无假设. 换言之, 发现的差别有多大才可以拒绝虚无假设.
  - 这惯例常常随学科不同而不同. 例如, 一些领域认为 $p \leq 0.05$ 就足够拒绝 H_0. 而另一些领域则把界限设定在 $p \leq 0.01$

65

假设验证的可能结果

- 实际情况是怎样?
  - - H_0 正确 (没有差异, 自变量没有效应)
  - - H_0 错误 (差异存在, 自变量产生效应)
- 研究结论是怎样?
  - - H_0 正确 (未发现差异, 未发现自变量的效应)
  - - H_0 错误 (发现差异, 发现自变量的效应)
- 这就构成了4种可能性 (2 * 2):
  - - 2 种错误方式
    - 2 种正确方式

66

I类错误和II类错误

真相

		没有效应 H_0 正确	真实效应 H_0 错误
统计决策	拒绝 H_0	I类错误	正确
	接受 H_0	正确	II类错误

67

两类错误

- 两类错误反映的情形不同，它们有不同的名称
  - α 错误（type I error） – 拒绝 H_0时所犯的错误.
    即侦察到不存在的差异
  - β 错误（type II error） – 接受H_0 时所犯的错误.
    - 即未能侦察到存在的差异

68

法庭裁决的类比

实际情况

		X 没犯罪	X 犯了罪
法庭的裁决	有罪	I 类错误	正确
	无罪	正确	II 类错误

I 类错误 – 将无辜者送进监狱

II 类错误 – 让罪人逃脱法网

69

设定可接受的alpha 水平

- 在科学研究中，我们常常采取保守的策略，设定标准来尽量减少I类错误 (自变量其实没有效应，但研究得到结论是有效应)。或者说，科学家要设定一个可接受的alpha 水平 (a)，或称显著性水平（level of significance）

70

显著性水平

- 显著性水平规定了当虚无假设正确时，样本结果非常不可能出现的概率值
- 当实验产生非常不可能（以alpha为标准）的数据时，我们就会拒绝虚无假设
- 所以，alpha 水平 也规定了I 类错误的概率 – 即，当H_0事实上正确时，拒绝H_0的概率
- 在心理学中，a 通常定在0.05

71

作业

- 1. 有一总体 μ=60， σ=10。请指出下列样本对应的z分数：

a. 样本1：n=4， *Xbar* =55

b. 样本2：n=25， *Xbar* =64

c. 样本3：n=100， *Xbar* =62

72

72

作业

- 2.　有一正态分布：μ=100， σ=20

a. 从中随机抽取一个样本 n = 25, 求其样本平均值 *Xbar*

b. 使用z分数, 指出位于此分布中间95%的样本的上下限。

c. 现从中取出一样本，均值 *Xbar* =106 ，n=25 ，问此样本是否处于分布的两端 5%的区域?

73

73

作业

- 3.　有一正态总体 μ=50， σ=10.

a. 请问，其中一个样本的值为45-55，即45<X<55的概率为多少?

b. 同一分布形态，但容量 n=25.，则此时其中一个样本的值为45-55，即45<Xbar<55的概率又为多少?

74

74

作业

- 4.　有一成绩的正态分布 μ=75， σ=12.

a. 如果从中抽取n=4的样本，试问，样本均值为70-80的概率为多少?

b. 如果从中抽取n=16的样本，试问，样本均值为70-80的概率为多少?

75

75

作业

- 5.　有一总体 μ=300， σ=80.

a. 如果从中随机抽取n=25 的样本, 则样本均值于总体均值间的误差将为多大?

b. 如果从中随机抽取n=100 的样本, 则样本均值于总体均值间的误差将为多大?

c. 如果从中随机抽取n=400 的样本, 则样本均值于总体均值间的误差将为多大?

d. 比较以上 a, b, c的答案, 指出样本容量与标准误之间的关系

76

76

作业

- 6. 讨论假设检验中可能犯的错误：
  - a) 什么是I类错误？为什么会发生?
  - b) 什么是II类错误？它是如何发生的?

77

77