

《传播统计学》

抽样分布

教师：林志良

邮箱：linzhl@nfu.edu.cn

个人网站：www.zhilianglin.com

目录

- 抽样分布：基本概念
- 总体服从正态分布的情形
- 总体不服从正态分布的情形
- 习题



抽样分布

抽样分布 (Sampling Distributions) 的概念

- A sampling distribution is a distribution of all of the possible values of a sample statistic for a given sample size selected from a population.
- 抽样分布是一组统计数据的分布，这组统计数据是由从一个总体中取出所有可能的固定大小的样本得到的。

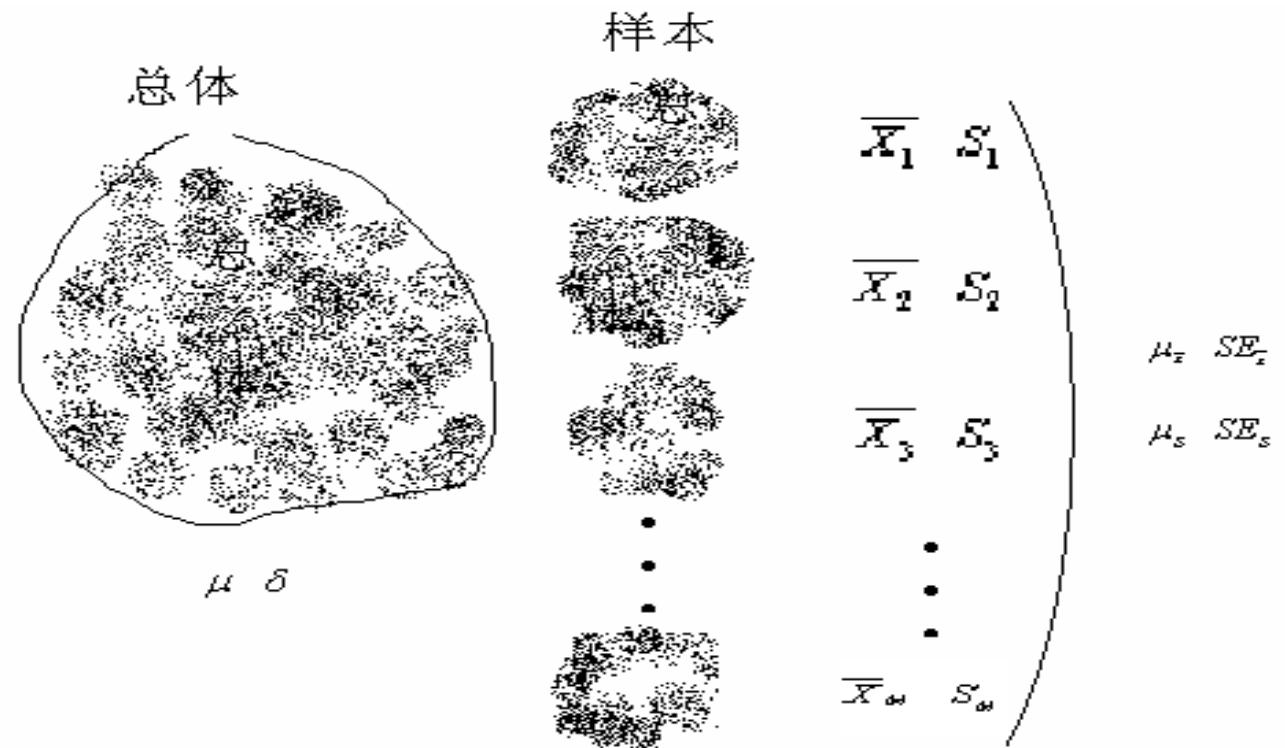


抽样分布

抽样分布 (Sampling Distributions) 的概念

- 抽样分布是指样本统计量的概率分布。如样本均值的分布、样本方差的分布

本节课主要关注
样本均值的分布



抽样分布

例子

- 总体样本量 $N=4$
- 变量 X 为个体的年龄
- X 的取值是: 18, 20, 22, 24

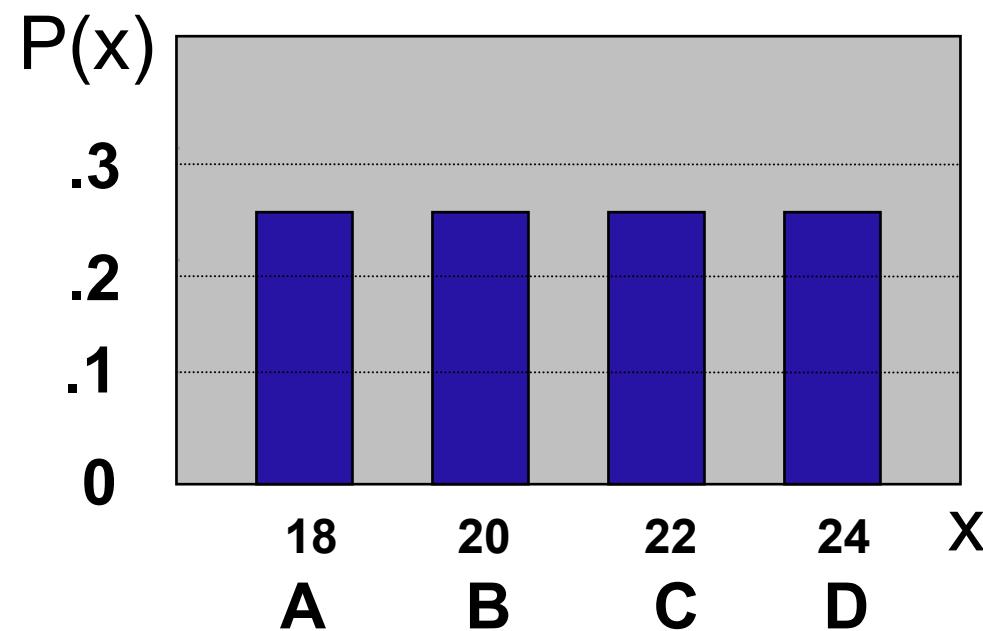


抽样分布

总体的平均值、标准差、分布

$$\mu = \frac{\sum X_i}{N}$$
$$= \frac{18 + 20 + 22 + 24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = 2.236$$



抽样分布

取多个样本量n=2的样本

1 st Obs	2 nd Observation			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24



16个样本 (有放回抽样)

1 st Obs	2 nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

16个样本均值

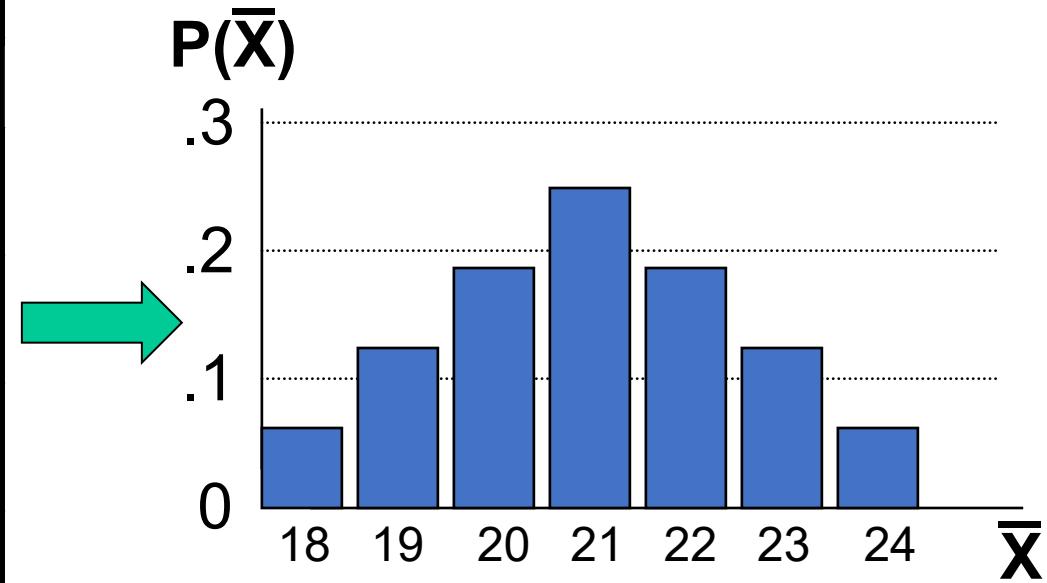
抽样分布

样本均值 (Sample Means) 的分布

16 个样本均值

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

样本均值分布



抽样分布

样本均值的平均值、标准误

$$\mu_{\bar{x}} = \frac{18+19+19+\dots+24}{16} = 21$$

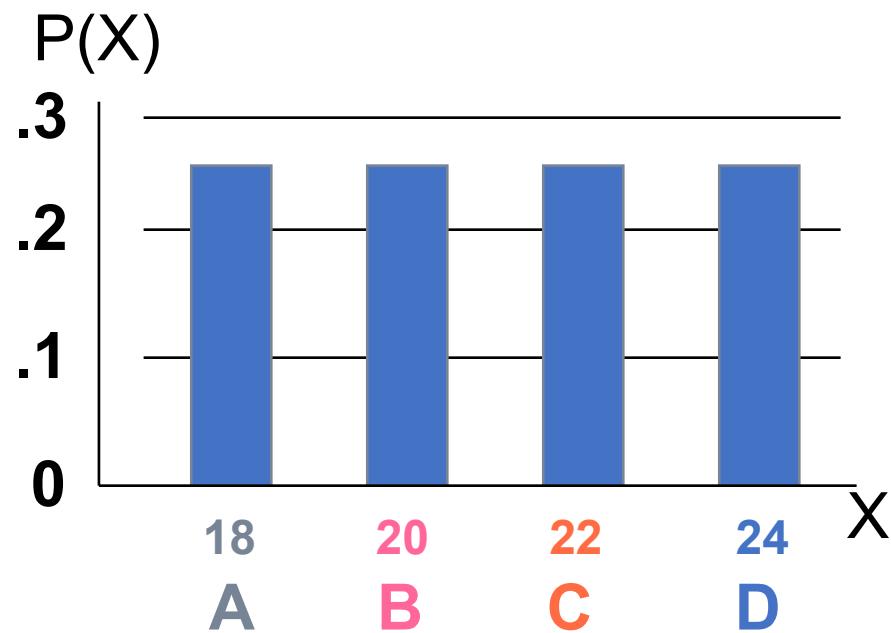
$$\sigma_{\bar{x}} = \sqrt{\frac{(18-21)^2 + (19-21)^2 + \dots + (24-21)^2}{16}} = 1.58$$

抽样分布

比较总体分布与样本均值的分布

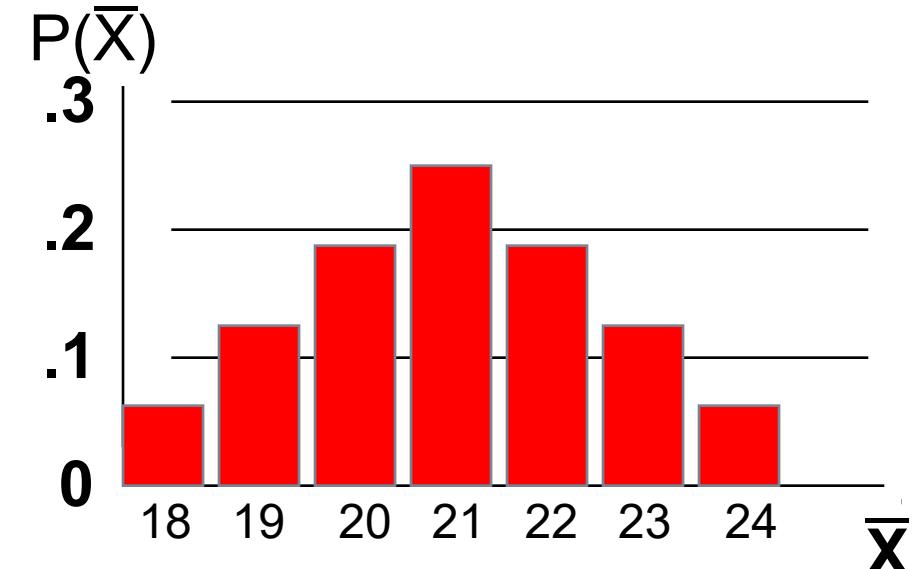
总体
 $N = 4$

$$\mu = 21 \quad \sigma = 2.236$$



样本均值的分布 ($n=2$)

$$\mu_{\bar{X}} = 21 \quad \sigma_{\bar{X}} = 1.58$$





抽样分布

样本均值的标准误 (Standard Error)

样本均值的标准差叫做**标准误 (standard error)**，样本均值的标准误由总体标准差及样本量决定：

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



抽样分布

样本均值的抽样分布：总体是正态分布的情形

如果总体服从正态分布（均值为 μ , 标准差为 σ ），那么 \bar{x} 的抽样分布同样服从正态分布，且：

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

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

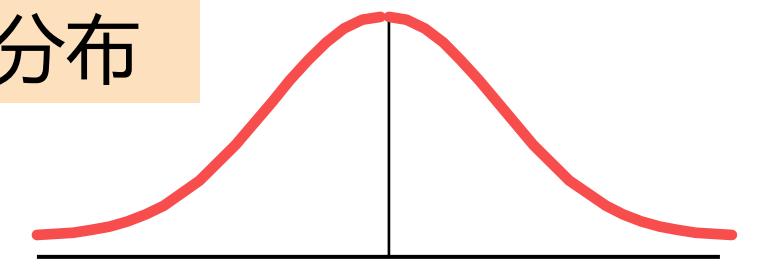


抽样分布

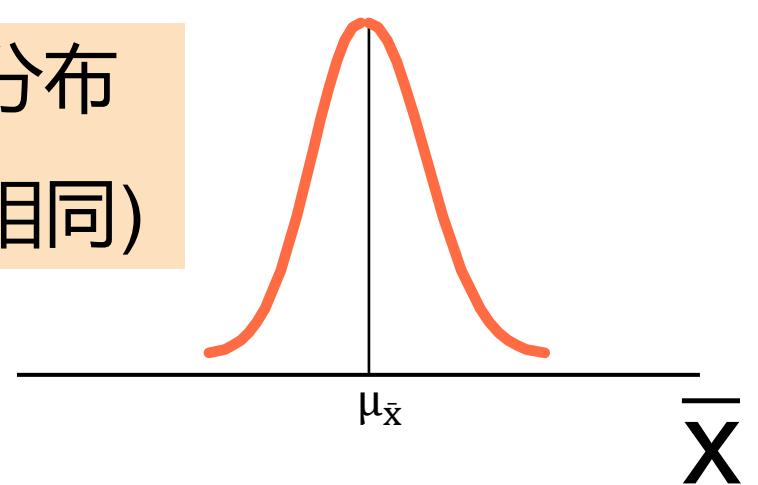
抽样分布的特征

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

正态分布



正态分布
(均值相同)



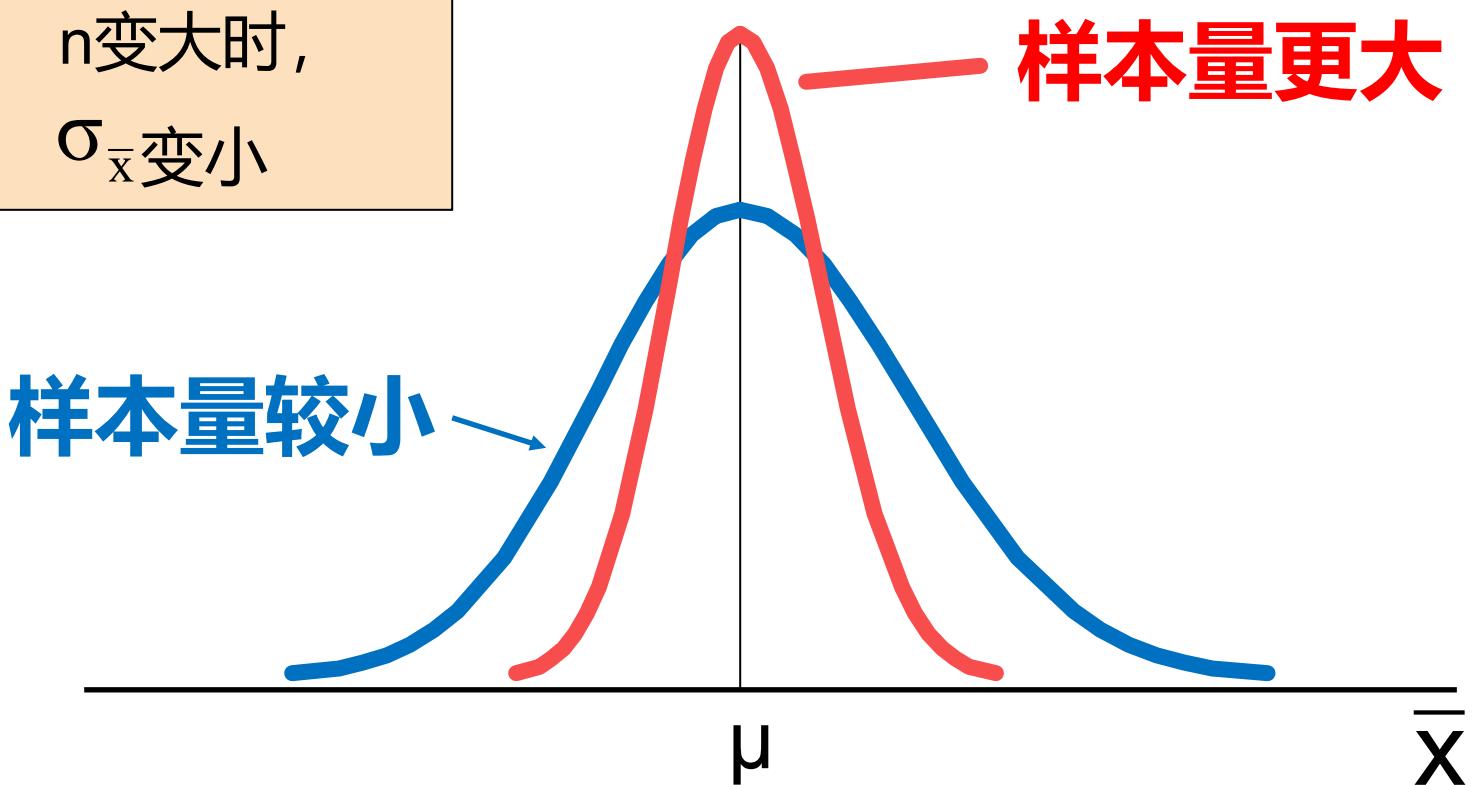


抽样分布

抽样分布的特征

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

n变大时，
 $\sigma_{\bar{X}}$ 变小





抽样分布

样本均值的抽样分布：总体不是正态分布的情形

根据**中心极限定理 (Central Limit Theorem)**，即使总体不是正态分布，当抽取的**样本n足够大时 ($n > 30$)** 时，样本均值的抽样分布近似服从正态分布，且样本均值的抽样分布有以下性质：

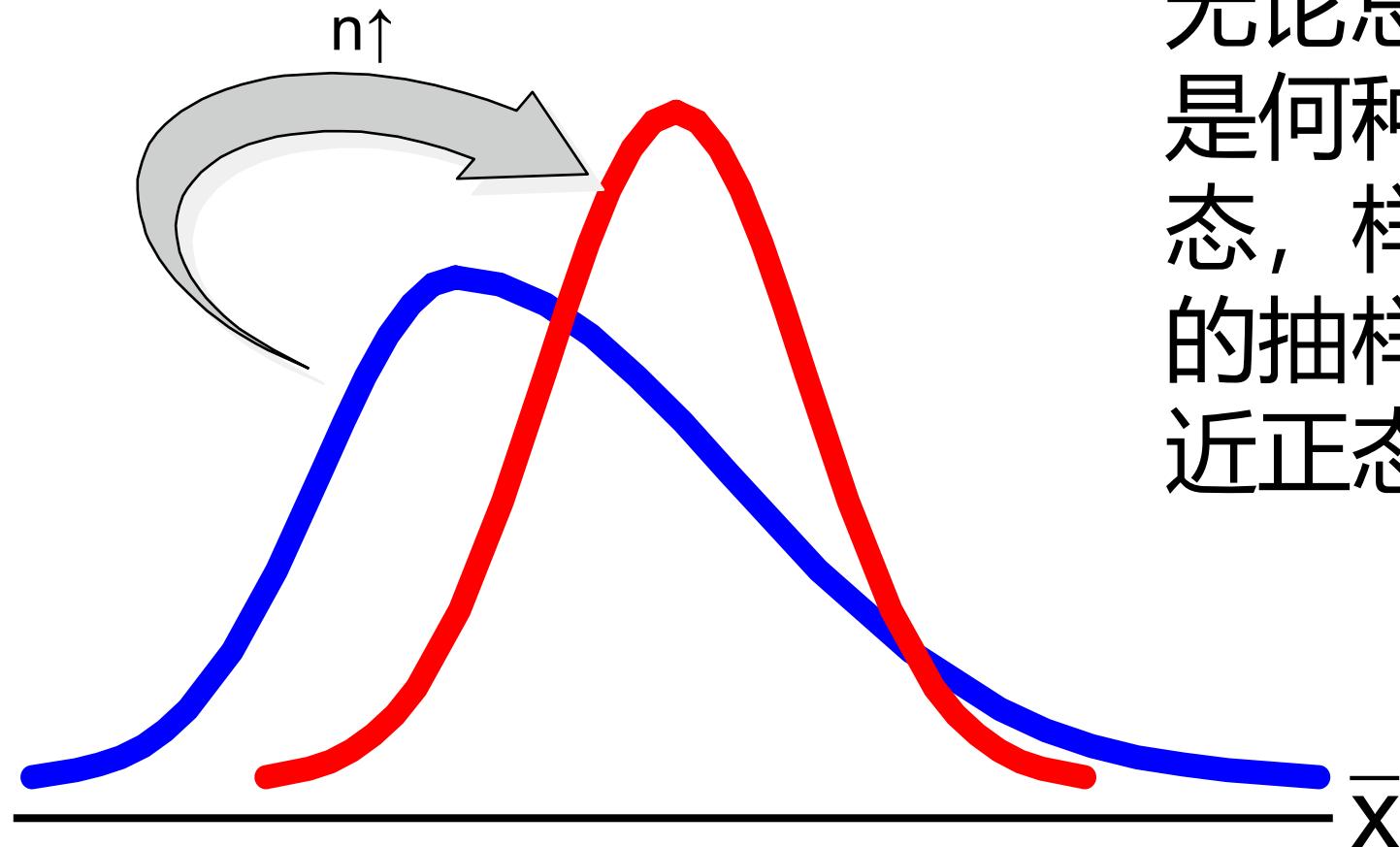
$$\mu_{\bar{x}} = \mu$$

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

抽样分布

中心极限定理

当样本容量足够大时.....



无论总体原来是何种分布形态，样本均值的抽样分布接近正态分布



抽样分布

样本均值的抽样分布的Z分数

\bar{X} 的抽样分布的Z分数:

$$Z = \frac{(\bar{X} - \mu_{\bar{X}})}{\sigma_{\bar{X}}} = \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

这里:
 \bar{X} = 样本均值
 μ = 总体均值
 σ = 总体标准差
 n = 样本容量

抽样分布



习题

某总体服从正态分布 ($\mu = 400$, $\sigma = 20$) 。

- 假如从总体中随机抽取 $n = 25$ 的样本。求样本均值 95% 的范围区间上下限。 (Z 分数分别是 -1.96 及 1.96 时, \bar{X} 的取值)
- 假如从总体中随机抽取 $n = 100$ 的样本。求样本均值 95% 的范围区间上下限。 (Z 分数分别是 -1.96 及 1.96 时, \bar{X} 的取值)

谢谢！