第十一讲
统计学（II）
Lecture 11
Statistics (II)

明玉瑞 Yurui Ming

yrming@gmail.com

声明
Disclaimer

- 本讲义在准备过程中由于时间所限，所用材料来源并未规范标示引用来源。所引材料仅用于教学所用，作者无意侵犯原著者之知识版权，所引材料之知识版权均归原著者所有；若原著者介意之，请联系作者更正及删除。

  The time limit during the preparation of these slides incurs the situation that not all the sources of the used materials (texts or images) are properly referenced or clearly manifested. However, all materials in these slides are solely for teaching and the author is with no intention to infringe the copyright bestowed on the original authors or manufacturers. All credits go to corresponding IP holders. Please address the author for any concern for remedy including deletion.

四分位数
Quartiles

- 使用一个叫做四分位数的概括性度量，任何数据集都可以分成四个相等的部分。四分位数是将一个排名数据集分成四个相等部分的三个概括性度量。第一个四分位数是所有小于中位数的观察值中的中间值，第二个四分位数与中位数相同，第三个四分位数是所有大于中位数的观察值中的中间值。

  Any data set can be divided into four equal parts by using a summary measure called quartiles. Quartiles are three summary measures that divide a ranked data set into four equal parts. The first quartile is the middle term among the observations that are less than the median, the second quartile is the same as the median, and the third quartile is the value of the middle term among the observations that are greater than the median.

- 依赖于第一四分位数和第三四分位数的变异度度量被称为四分位距（IQR），并定义为第三四分位数与第一四分位数之间的差值，即：$IQR = Q_3 - Q_1$。

  The measure of variation that depends on the first and the third quartiles is called the interquartile range (IQR) and defined as the difference between the third and the first quartiles, that is: $IQR = Q_3 - Q_1$.

箱线图
Box Plot

- 使用五个度量：中位数（Q_2）、Q_1和Q_3、数据集中的最小值和最大值的图表展示，称为箱线图。这是通过从最小数据值绘制水平线到Q_1，从Q_3到最大数据值绘制水平线，绘制一个垂直线通过Q_1和Q_3的箱子的图表，箱子内部通过中位数Q_2绘制的垂直线获得的数据集的图表。

  The graph presentation of data using the five measures: the median (Q_2), Q_1 and Q_3, the smallest and the largest values in a data set, is called a box plot. It is a graph of a data set obtained by drawing a horizontal line from the minimum data value to Q_1, drawing a horizontal line from Q_3 to the maximum data value, and drawing a box whose vertical sides pass through Q_1 and Q_3 with a vertical line inside the box passing through the median Q_2.

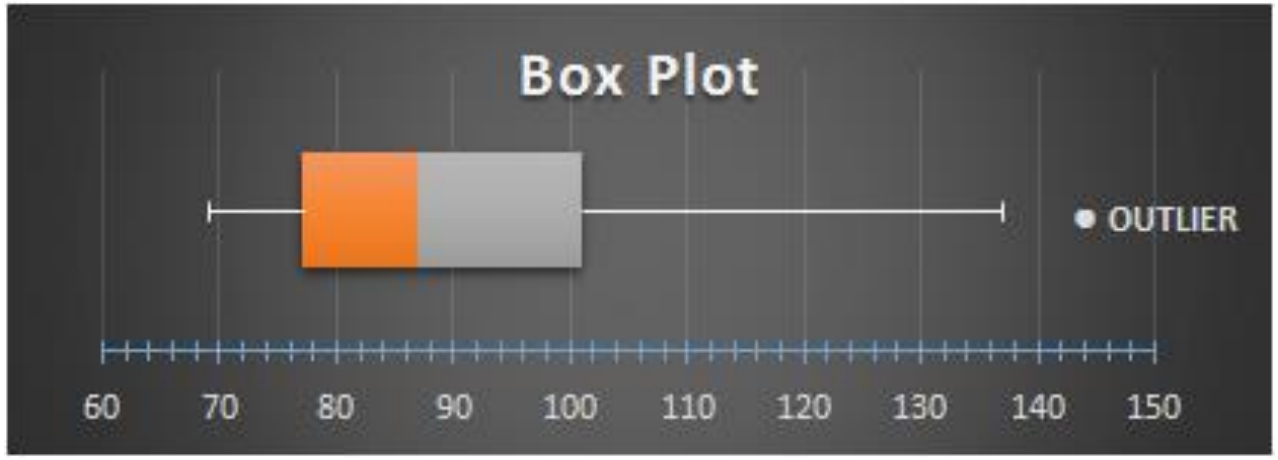
- 箱线图也被称为箱须图。它可以用来展示数据集中的一些特征，如数据集的分散程度、异常值和偏斜程度。

  A box plot is also called box-and-whisker plot. It can be used to illustrate some features in the data set such as the spread, the outlier, and the skewness of a data set.

箱线图
Box Plot

- 异常值是与数据值相比极高或极低的数据值。确定异常值的标准是，如果它大于最高界限：$Q_3 + 1.5 \times IQR$，或小于最低界限：$Q_1 - 1.5 \times IQR$ 。

  An outlier is an extremely high or an extremely low data value, when compared with the data values. The criteria for identifying an outlier if it is greater than the highest fence: $Q_3 + 1.5 \times IQR$, or less than the lowest fence: $Q_1 - 1.5 \times IQR$.



采样
Sampling

- 现在我们再次转向推论统计学，它强调对感兴趣的变量的总体特征进行结论或推断。实现这一目标的一种方法是进行普查，即为总体中的每个个体收集数据。然而，由于以下一种或多种原因，通常这是不可行的：
  - 这可能会过于昂贵或耗时
  - 测试可能是破坏性的
  - 总体可能仅仅是概念性的

  相反，我们可以收集样本，即人口的一个子集的数据。然后，我们使用样本的特征来估计总体的特征。为了使这个过程能够给出一个良好的估计，样本必须代表总体。否则，如果使用不具有代表性或“偏倚”的样本，结论将会系统性地不正确。

- Now we turn to Inferential Statistics again, which emphasizes on making conclusions, or inferences, about the population characteristics of variables of interest. One way to do so is to conduct a census, i.e. to collect data for each individual in the population. However often this is not feasible, due to one or more of the following:
  - It may be too expensive or time consuming to do so.
  - Testing may be destructive
  - The population may be purely conceptual

  Instead, we collect data only for a sample, i.e. a subset of the population. We then use the characteristics of the sample to estimate the characteristics of the population. In order for this procedure to give a good estimate, the sample must be representative of the population. Otherwise, if an unrepresentative or `biased' sample is used the conclusions will be systematically incorrect.

变量
Variables

- 对于给定的总体，通常会有一个或多个我们感兴趣的变量。为了更好地理解，我们考虑以下总体以及相应的感兴趣变量：
  - 北京市所有全职工作的成年男性；感兴趣的变量是个人的总收入。
  - 所有有权投票的美国成年人；感兴趣的变量是支持的政党。
  - 特定公司生产的特定类型汽车电池；感兴趣的变量是电池在故障前的使用寿命。
  - 实验室计划中的所有潜在可能结果；感兴趣的变量是特定测量值的数值。

- For a given population there will typically be one or more variables in which we are interested. To have a better understanding, we consider the following populations together with corresponding variables of interest:
  - All adult males working full-time in Beijing; the variable of interest is the person's gross income.
  - All adults in the US who are eligible to vote; the variable of interest is the political party supported.
  - Car batteries of a particular type manufactured by a particular company; the variable of interest is the lifetime of the battery before failure.
  - All potential possible outcomes of a planned laboratory experiment; the variable of interest is the value of a particular measurement.

变量
Variables

- 在现代统计学中，确保代表性的最常见方法是根据概率抽样规则选择一个大小为n的随机样本。这种概率抽样是客观的，消除了调查者的偏见。对于有限大小为N的总体，最常用的方法是使用简单随机抽样。它有两种主要形式：无放回抽样和有放回抽样。

  In modern Statistics, the most common way of guaranteeing representativeness is to use a random sample of size n chosen according to a probabilistic sampling rule. This probabilistic sampling is objective and eliminates investigator bias. For a population of finite size N, the most common method is to use simple random sampling. This takes two main forms: sampling without replacement and sampling with replacement.

- 假设v_1，v_2，…，v_N表示总体中第1个、第2个、第N个个体的变量X的值。假设我们有兴趣估计X的总体均值，$\mu = \left(\sum_{i=1}^{N} v_i\right)/N$。让$X_1$、$X_2$、…、 X_n表示通过无放回抽样选择的大小为n的样本中X的值。总体均值μ可以通过下式估计：$\bar{X} = \left(\sum_{i=1}^{N} X_i\right)/n$。

  Let v_1, v_2, …, v_N denote the values of the variable X for the 1st, 2nd, …, Nth individuals in the population. Suppose that interest lies in estimating the population mean of X:$\mu = \left(\sum_{i=1}^{N} v_i\right)/N$. Let X_1, X_2, …, X_n be the values of X in a sample of size n chosen by sampling without replacement. The population mean μ can be estimated by: $\bar{X} = \left(\sum_{i=1}^{N} X_i\right)/n$.

估计
Estimating

- $\bar{X}$的值会因不同的样本而不同，因此$\bar{X}$是一个随机变量，因为样本是随机选择的。因此，$\bar{X}$有自己的概率分布，即其抽样分布。我们如何衡量上述估计总体均值μ的方法的性能？一种方法是计算$\bar{X}$的抽样分布的期望和方差。特别地，可以证明在无放回抽样的情况下，$\mathrm{E}(\bar{X}) = \mu$，因此，$\bar{X}$被认为是无偏的。

  The value of $\bar{X}$ will be different for different samples, and so $\bar{X}$ is a random variable because the sample is chosen randomly. Thus, $\bar{X}$ has its own probability distribution, which is known as its sampling distribution. How can we measure the performance of the above method of estimating μ? One way is to calculate the expectation and variance of the sampling distribution of $\bar{X}$. In particular, it can be shown that under sampling without replacement, $\mathrm{E}(\bar{X}) = \mu$. As a result, $\bar{X}$ is said to be unbiased.

- 实际上，我们可以证明如下：

$$\mathrm{E}(\bar{X}) = \frac{\sum_{i=1}^{\binom{N}{n}} \bar{X}^i}{\binom{N}{n}} = \frac{\sum_{i=1}^{\binom{N}{n}} \frac{1}{n} \sum_{j=1}^{n} X_j^i}{\binom{N}{n}} = \frac{\frac{1}{n} \sum_{i=1}^{\binom{N}{n}} \sum_{j=1}^{n} X_j^i}{\binom{N}{n}}$$

估计
Estimating

- 对于$\sum_{i=1}^{\binom{N}{n}}\sum_{j=1}^{n}X_j^i$，实际上是将所有可能的采样相加，求$X$的任一取值在和中的系数。例如对于个体取值$X_1$，我们只需求包含$X_1$的所有可能采样的个数。显然，除去$X_1$外，在$X$的所有可能的$N$个取值中，还有$N-1$个值，我们必须无放回地采样$n-1$个样本，才能保证加上$X_1$后，正好是具有$n$元素的样本，即共有$\binom{N-1}{n-1}$个。以此类推，所以：

  For $\sum_{i=1}^{\binom{N}{n}}\sum_{j=1}^{n}X_j^i$, it is actually equivalent to summing up all possible samples and calculating the coefficient of any value of X in the sum. For example, for the value X_1, we only need to count the number of all possible samples that include X_1. Clearly, among all the possible N values of X, there are $N-1$ values if we exclude X_1 prior to sampling. Hence, we must sample $n-1$ samples without replacement to ensure that when X_1 is included, it exactly forms a sample with n elements, so there are a total of $\binom{N-1}{n-1}$ of them. So we deduce as follows:

$$\mathrm{E}(\bar{X})=\frac{\sum_{i=1}^{\binom{N}{n}}\bar{X}^i}{\binom{N}{n}}=\frac{\sum_{i=1}^{\binom{N}{n}}\frac{1}{n}\sum_{j=1}^{n}X_j^i}{\binom{N}{n}}=\frac{\frac{1}{n}\sum_{i=1}^{\binom{N}{n}}\sum_{j=1}^{n}X_j^i}{\binom{N}{n}}=\frac{\frac{1}{n}\sum_{k=1}^{N}\binom{N-1}{n-1}X_k}{\binom{N}{n}}=\frac{\frac{1}{n}\sum_{k=1}^{N}\binom{N}{n}\frac{n}{N}X_k}{\binom{N}{n}}$$

$$=\frac{\frac{1}{n}\binom{N}{n}\frac{n}{N}\sum_{k=1}^{N}X_k}{\binom{N}{n}}=\frac{\sum_{k=1}^{N}X_k}{N}=\mu$$

估计
Estimating

- 此外，可以证明在无放回抽样的情况下：

  Moreover, it is possible to show that under sampling without replacement:

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

  其中总体方差σ^2定义为：

  Where the population variance σ^2 is defined as:

$$\sigma^2 = \frac{1}{N}\sum_{j=1}^{N}\left(v_j - \mu\right)^2 = \frac{1}{N}\sum_{j=1}^{N} {v_j}^2 - \mu^2$$

- 上式说明，采用较大的n会使对总体的平均值的估计更准确。

  The equation above indicates that using a larger n will result in a more accurate estimate of the population mean.

估计
Estimating

- 回想一下，对于一般（不一定是有限的）总体，随机选取的个体的定量变量值可以由实值随机变量 X 描述，其累积分布函数 (c.d.f.) 为$F_X(x) = P(X \leq x)$。

  Recall that for a general (i.e. not necessarily finite) population, the value of a quantitative variable for a randomly selected individual can be described by a real-valued random variable X with cumulative distribution function (c.d.f.) $F_X(x) = P(X \leq x)$.

- 如果X_1、X_2、…、X_n服从$F_X(x)$且相互独立，即$P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n) = F_X(x)$，我们称$X_1$、$X_2$、…、$X_n$为从$X$中随机抽取的样本。

  We say that X_1, X_2, …, X_n are a random sample from X if X_1, X_2, …, $X_n \sim F_X(x)$ independently, aka, $P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n) = F_X(x_1)F_X(x_2)\cdots F(x_n)$.

- 设x_1、x_2、…、x_n为某个特定随机变量X的一个随机样本中的观测值，其分布未知。我们可能希望使用这些数据来估计事件$\{X \in A\}$的概率。一种方法是使用事件的经验概率，换句话说，就是样本值中落在A内的比例：

$$\hat{P}(\{X \in A\}) = \hat{P}(X \in A) = \frac{\#\{i: x_i \in A\}}{n}$$

估计
Estimating

Let $x_1, x_2, \ldots, x_n$ be the observed values in a particular random sample of the random variable X, whose distribution is unknown. We may wish to use these data to estimate the probability of an event $\{X \in A\}$. One way is to use the empirical probability of the event, in other words the proportion of the sample values that lie in A.

- 另一种方法是假设数据是从特定参数概率模型中生成的随机样本，例如$\mathrm{N}(\mu, \sigma^2)$。这种模型通常包含未知参数，例如在前面的例子中，参数μ和σ^2是未知的。我们可以使用样本来估计分布的参数，从而将模型拟合到数据中。拟合好的模型可以用来计算感兴趣事件的概率。

  An alternative approach is to assume that the data were generated as a random sample from a particular parametric probability model, e.g. $\mathrm{N}(\mu, \sigma^2)$. Such models usually contain unknown parameters, e.g. in the previous example the parameters μ and σ^2 are unknown. We can use the sample to estimate the parameters of the distribution, thereby fitting the model to the data. A fitted model can be used to calculate probabilities of events of interest.

- 如果所选择的模型适合，那么基于经验和基于模型的事件概率估计应该是相似的。由于我们只观察了一个随机样本而不是整个总体，经验和基于模型的估计概率之间会频繁出现小差异。因此，这两个估计都在真实总体概率周围表现出随机变化。然而，经验和基于模型的概率之间的较大差异可能表明所选择的参数模型与真实的数据生成过程不太匹配。

统计量
Statistic

If the chosen model is a good fit then the empirical and model-based estimated probabilities of the event should be similar. Small differences between the empirical and model-based estimated probabilities will occur frequently due to the fact that we have only observed a random sample and not the entire population. Thus, both estimates exhibit random variation around the true population probability. However, large differences between empirical and model-based probabilities may be indicative that the chosen parametric model is a poor approximation of the true data generating process.

- 设X_1、X_2、…、X_n是从分布$F_X(x)$中抽取的随机样本，统计量被定义为数据的一个函数，即$h(X_1,X_2,\cdots,X_n)$。通常情况下，这个统计量基于不同的样本会有不同的值。由于样本数据是随机的，所以统计量也是一个随机变量。如果我们反复抽取大小为n的样本，每次计算并记录样本统计量的值，那么我们将得到其概率分布。样本统计量的概率分布称为抽样分布。

  Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution $F_X(x)$. A statistic is a function of the data, $h(X_1,X_2,\cdots,X_n)$. The value of this statistic will usually be different for different samples. As the sample data is random, the statistic is also a random variable. If we repeatedly drew samples of size n, calculating and recording the value of the sample statistic each time, then we would build up its probability distribution. The probability distribution of a sample statistic is referred to as its sampling distribution.

独立同分布
Independent and Identically Distributed

- 通常，随机变量X_1、X_2、…、X_n被假定为独立同分布（通常简称为i.i.d.）的随机变量，每个随机变量都服从$F_X(x)$的分布。这意味着对于$i = 1, 2, \cdots, n$，有$\mathrm{E}(X_i) = \mu$和$\mathrm{Var}(X_i) = \sigma^2$。

  Usually, the random variables X_1, X_2, …, X_n are assumed to be independent and identically distributed (often abbreviated to i.i.d.) random variables, each being distributed as $F_X(x)$. This means that $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ for $i = 1, 2, \cdots, n$.

- 计算X的抽样（概率）分布的均值是非常直观的，如下所示：

  It is straightforward to calculate the mean of the sampling (probability) distribution of X as follows:

$$\mathrm{E}(\bar{X}) = \mathrm{E}\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] = \frac{1}{n}[\mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_n)] = \frac{n\mu}{n} = \mu$$

- 由于X_i是独立的，所以有$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$，因此方差计算如下：

  Because the X_i are independent, it holds that $\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$, so the variance is calculated as follows:

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left[\frac{1}{n}(X_1 + \cdots + X_n)\right] = \frac{1}{n^2}[\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

估计
Estimating

- 上述结果告诉我们，样本均值$\bar{X}$的抽样分布以每个样本变量X_1、X_2、…、X_n的共同均值μ为中心（即从中获得样本的分布的均值），并且方差等于X_i的共同方差除以n。因此，随着样本大小n的增加，样本均值$\bar{X}$的抽样分布会更加集中在真实均值μ周围。

  The above results tell us that the sampling distribution of the sample mean $\bar{X}$ is centered on the common mean μ of each of the sample variables X_1, X_2, …, X_n (i.e. the mean of the distribution from which the sample is obtained) and has variance equal to the common variance of the X_i divided by n. Thus, as the sample size n increases, the sampling distribution of $\bar{X}$ becomes more concentrated around the true mean μ.

- 如果样本X_i是i.i.d.且服从$\mathrm{N}(\mu, \sigma^2)$的随机变量，且样本均值$\bar{X}$是$X_i$的线性组合（系数$c_i = 1/n$, $i = 1, 2, \cdots, n$）。因此，$\bar{X}$服从均值为μ，方差为σ^2/n的正态分布，即$\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$。

  If now the X_i in the sample are i.i.d. $\mathrm{N}(\mu, \sigma^2)$ random variables then the sample mean, $\bar{X}$, is a linear combination of the X_i (with $c_i = 1/n$, $i = 1, 2, \cdots, n$, using the notation above). Thus, $\bar{X}$ is normally distributed with mean μ and variance σ^2/n, i.e. $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$.

$$\sum_{i=1}^{n} c_i X_i \sim \mathrm{N}\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$$

中心极限定理
The Central Limit Theorem

- 在前面讲解中，我们看到随机量$\bar{X}$服从一个均值为μ和方差为σ^2/n的抽样分布。在特殊情况下，比如从正态分布中抽样时，$\bar{X}$也服从正态分布。然而，在许多情况下，我们无法确$\bar{X}$的精确分布形式。在这种情况下，我们可以借助中心极限定理并获得一个近似分布。

  In the previous lecture, we saw that the random quantity $\bar{X}$ has a sampling distribution with mean μ and variance σ^2/n. In the special case when we are sampling from a normal distribution, $\bar{X}$ is also normally distributed. However, there are many situations when we cannot determine the exact form of the distribution of $\bar{X}$. In such circumstances, we may appeal to the central limit theorem and obtain an approximate distribution.

- 中心极限定理：设X是一个随机变量，具有均值μ和方差σ^2。如果$\bar{X}$是从 X 的分布中抽取的大小为n的随机样本的均值，则当$n \to +\infty$，下面统计量的分布趋向于标准正态分布：

  The central limit theorem: Let X be a random variable with mean μ and variance σ^2. If $\bar{X}$ is the mean of a random sample of size n drawn from the distribution of X, then the distribution of the statistic below tends to the standard normal distribution as $n \to +\infty$:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

中心极限定理
The Central Limit Theorem

- 上面结论意味着，对于一个来自均值为μ方差为σ^2的大样本，样本均值$\bar{X}$近似地服从均值为μ方差为σ^2/n的正态分布。因此，对于大的n，近似有$\bar{X}\sim\mathrm{N}(\mu,\sigma^2/n)$，从而$\sum_{i=1}^{n}X_i\sim\mathrm{N}(n\mu,n\sigma^2)$。同时，不需要明确指定潜在分布$F_X$的形式，无论它是离散的还是连续的，都可以使用这个结果。因此，这个结果具有重要的实际意义。

  This means that, for a large random sample from a population with mean μ and variance σ^2, the sample mean $\bar{X}$ is approximately normally distributed with mean μ and variance σ^2/n. Since, for large n, $\bar{X}\sim\mathrm{N}(\mu,\sigma^2/n)$ approximately we have that $\sum_{i=1}^{n}X_i\sim\mathrm{N}(n\mu,n\sigma^2)$ approximately. In addition, there is no need to specify the form of the underlying distribution FX, which may be either discrete or continuous, in order to use this result. As a consequence it is of tremendous practical importance.

- 一个常见的问题是，“在样本$\bar{X}$的正态性断言合理之前，n需要多大？”答案取决于从中抽取样本的基础分布F_X的非正态程度。分布F_X越不正态，所需的n就越大。一个有用的经验法则是n至少应该为30。

  A common question is `how large does n have to be before the normality of $\bar{X}$ is reasonable?' The answer depends on the degree of non-normality of the underlying distribution from which the sample has been drawn. The more non-normal F_X is, the larger n needs to be. A useful rule-of-thumb is that n should be at least 30.

方差
Variance

- 我们接下来研究样本方差的抽样分布，S^2，其定义如下，其中X_1、X_2、…、X_n是从具有累积分布函数$F_X(\cdot)$、均值为μ和方差为σ^2的分布中抽取的一个随机样本：

  we will look at the sampling distribution of the sample variance, S^2, defined as follows, where X_1, X_2, …, X_n are a random sample from the distribution with c.d.f. $F_X(\cdot)$ with mean μ and variance σ^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- 下面我们证明S^2是对σ^2的无偏估计：

  We demonstrate S^2 is an unbiased estimation of σ^2:

$$\mathrm{E}(S^2) = \frac{1}{(n-1)}\mathrm{E}\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{1}{(n-1)}\mathrm{E}\left[\sum_{i=1}^{n}[(X_i - \mu) - (\bar{X} - \mu)]^2\right]$$

$$= \frac{1}{(n-1)}\mathrm{E}\left[\sum_{i=1}^{n}[(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2]\right]$$

$$= \frac{1}{(n-1)}\mathrm{E}\left[\sum_{i=1}^{n}(X_i - \mu)^2 - \sum_{i=1}^{n}2(X_i - \mu)(\bar{X} - \mu) + \sum_{i=1}^{n}(\bar{X} - \mu)^2\right]$$

方差
Variance

$$\sum_{i=1}^{n} 2(X_i-\mu)(\bar{X}-\mu) = 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) = 2(\bar{X}-\mu)\left(\sum_{i=1}^{n}X_i - \sum_{i=1}^{n}\mu\right) =$$
$$= 2(\bar{X}-\mu)(n\bar{X}-n\mu) = 2n(\bar{X}-\mu)(\bar{X}-\mu)$$

$$\sum_{i=1}^{n}(\bar{X}-\mu)^2 = n(\bar{X}-\mu)^2$$

$$\mathrm{E}(S^2) = \frac{1}{(n-1)}\mathrm{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2 - 2n(\bar{X}-\mu)(\bar{X}-\mu) + n(\bar{X}-\mu)^2\right]$$

$$= \frac{1}{(n-1)}\mathrm{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2\right] = \frac{1}{(n-1)}\left[\sum_{i=1}^{n}\mathrm{E}(X_i-\mu)^2 - n\mathrm{E}(\bar{X}-\mu)^2\right]$$

$$\mathrm{E}(X_i-\mu)^2 = \sigma^2$$

$$\mathrm{E}(\bar{X}-\mu)^2 = \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$\mathrm{E}(S^2) = \frac{1}{(n-1)}\left[\sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n}\right] = \frac{1}{(n-1)}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right] = \frac{1}{(n-1)}[n\sigma^2-\sigma^2]$$

$$= \frac{1}{(n-1)}[(n-1)\sigma^2] = \sigma^2$$

方差
Variance

- 因此，我们可以看出，在S^2的定义中使用除数$(n-1)$，我们得到的统计量的抽样分布以真实的分布值σ^2为中心，即我们得到的估计是无偏的。如果我们使用了可能更直观明显的n值，情况就不同了。

  Hence, we can see that by using divisor $(n-1)$ in the definition of S^2, we obtain a statistic whose sampling distribution is centered on the true distribution value of σ^2. It means the statistic is an unbiased one by definition. This would not be the case if we had used the perhaps more intuitively obvious value of n.

- 统计分析的目标是基于样本对总体进行推断。通常，我们首先假设数据是由总体的概率模型生成的。这样的模型通常包含一个或多个参数θ，其值是未知的，需要使用样本数据来估计θ的值。例如，之前我们使用样本均值来估计总体均值，使用样本比例来估计总体比例。

  The objective of a statistical analysis is to make inferences about a population based on a sample. Usually we begin by assuming that the data were generated by a probability model for the population. Such a model will typically contain one or more parameters θ whose value is unknown. The value of θ needs to be estimated using the sample data. For example, previously we have used the sample mean to estimate the population mean, and the sample proportion to estimate the population proportion.

点估计
Point Estimation

- 一个给定的估计过程通常会在不同样本中产生不同的结果，因此在从总体中随机抽样时，估计的结果将成为一个具有自己抽样分布的随机变量。为了进一步讨论估计过程应具备的特性，我们从以下问题开始：
  - 估计过程是好还是不好？
  - 抽样分布期望具备什么样的特性？

  A given estimation procedure will typically yield different results for different samples, thus under random sampling from the population the result of the estimation will be a random variable with its own sampling distribution. To discuss further the properties an estimation procedure to have. We begin from the questions below:

  - Is the estimation procedure a good one or not?
  - What properties would the sampling distribution be supposed to have?
- 设X_1、X_2、…、X_n是从具有累积分布函数$F_X(x;\theta)$的分布中随机抽取的样本，其中θ是一个未知参数。参数θ的（点）估计器，记作$\hat{\theta}$，是样本的一个实数、单值函数，即$\hat{\theta} = h(X_1, X_2, \ldots, X_n)$。由于$X_i$是随机变量，估计器$\hat{\theta}$也是一个随机变量，其概率分布称为其抽样分布。

点估计
Point Estimation

Let X_1, X_2, ..., X_n be a random sample from a distribution with c.d.f. $F_X(x;\theta)$, where θ is a parameter whose value is unknown. A (point) estimator of θ, denoted by $\hat{\theta}$ is a real, single-valued function of the sample, i.e. $\hat{\theta} = h(X_1, X_2, ..., X_n)$. As we have seen already, because the X_i are random variables, the estimator $\hat{\theta}$ is also a random variable whose probability distribution is called its sampling distribution.

- 对于特定的观察数据样本x_1、...、x_n，所假设的估计值$\hat{\theta} = h(x_1, ..., x_n)$被称为参数$\theta$的（点）估计。需要注意的是，由于抽样误差，点估计几乎永远不会完全等于参数θ的真实值。实际上，θ可能是一个包含p个标量参数的向量。在这种情况下，我们需要针对θ的每个分量使用p个独立的估计器。例如，正态分布有两个标量参数μ和σ^2。这些参数可以合并成一个参数向量$\theta = (\mu, \sigma^2)$，其中一个可能的估计器是$\hat{\theta} = (\bar{X}, S^2)$。

  The value $\hat{\theta} = h(x_1, ..., x_n)$ assumed for a particular sample x_1, ..., x_n of observed data is called a (point) estimate of θ. Note the point estimate will almost never be exactly equal to the true value of θ, because of sampling error. Often θ may in fact be a vector of p scalar parameters. In this case, we require p separate estimators for each of the components of θ. For example, the normal distribution has two scalar parameters μ and σ^2. These could be combined into a single parameter vector,$\theta = (\mu, \sigma^2)$, for which one possible estimator is $\hat{\theta} = (\bar{X}, S^2)$.

点估计
Point Estimation

- 我们希望估计器$\hat{\theta}$对参数θ满足以下条件：
  - (i) 估计器$\hat{\theta}$的抽样分布以目标参数θ为中心。(ii) 估计器$\hat{\theta}$的抽样分布扩展较小。

  如果一个估计器具备以上的性质(i)和(ii)，那么我们可以期待通过统计实验得出的估计值会接近我们试图估计的总体参数的真实值。

  We would like an estimator $\hat{\theta}$ of θ to be such that:

  - (i) the sampling distribution of $\hat{\theta}$ is centered about the target parameter, θ. (ii) the spread of the sampling distribution of $\hat{\theta}$ is small.

  If an estimator has properties (i) and (ii) above then we can expect estimates resulting from statistical experiments to be close to the true value of the population parameter we are trying to estimate.

- 点估计器$\hat{\theta}$的偏差为$\text{bias}(\hat{\theta}) = \text{E}(\hat{\theta}) - \theta$。如果$\text{E}(\hat{\theta}) = \theta$，即$\text{bias}(\hat{\theta}) = 0$，则称估计器为无偏估计器。

  The bias of a point estimator $\hat{\theta}$ is $\text{bias}(\hat{\theta}) = \text{E}(\hat{\theta}) - \theta$. The estimator is said to be unbiased if $\text{E}(\hat{\theta}) = \theta$; i.e. if $\text{bias}(\hat{\theta}) = 0$.

点估计
Point Estimation

- 我们希望估计器$\hat{\theta}$对参数θ满足以下条件：
  - (i) 估计器$\hat{\theta}$的抽样分布以目标参数θ为中心。(ii) 估计器$\hat{\theta}$的抽样分布散布较窄。

  如果一个估计器具备以上的性质(i)和(ii)，那么我们可以期待通过统计实验得出的估计值会接近我们试图估计的总体参数的真实值。

  We would like an estimator $\hat{\theta}$ of θ to be such that:

  - (i) the sampling distribution of $\hat{\theta}$ is centered about the target parameter, θ. (ii) the spread of the sampling distribution of $\hat{\theta}$ is small.

  If an estimator has properties (i) and (ii) above then we can expect estimates resulting from statistical experiments to be close to the true value of the population parameter we are trying to estimate.

- 点估计器$\hat{\theta}$的偏差为$\text{bias}(\hat{\theta}) = \text{E}(\hat{\theta}) - \theta$。如果$\text{E}(\hat{\theta}) = \theta$，即$\text{bias}(\hat{\theta}) = 0$，则称估计器为无偏估计器。

  The bias of a point estimator $\hat{\theta}$ is $\text{bias}(\hat{\theta}) = \text{E}(\hat{\theta}) - \theta$. The estimator is said to be unbiased if $\text{E}(\hat{\theta}) = \theta$; i.e. if $\text{bias}(\hat{\theta}) = 0$.

点估计
Point Estimation

- 无偏性对应于上述的性质(i)，通常被视为估计器的一个理想特性。需要注意的是，有时候有偏的估计器可以被修改为无偏估计器。例如，如果$\mathrm{E}(\hat{\theta}) = k\theta$，其中$k \neq 1$是一个常数，那么$\mathrm{bias}(\hat{\theta}) = (k-1)\theta$。然而，$\hat{\theta}/k$是参数$\theta$的一个无偏估计器。

  Unbiasedness corresponds to property (i) above, and is generally seen as a desirable property for an estimator. Note that sometimes biased estimators can be modified to obtain unbiased estimators. For example, if $\mathrm{E}(\hat{\theta}) = k\theta$, where $k \neq 1$ a constant, then $\mathrm{bias}(\hat{\theta}) = (k-1)\theta$. However, $\hat{\theta}/k$ is an unbiased estimator of θ.

- 抽样分布的散布程度可以通过$\mathrm{Var}(\hat{\theta})$来衡量。估计器$\hat{\theta}$的标准差$\sqrt{\mathrm{Var}(\hat{\theta})}$，被称为标准误差。假设我们有两个不同的无偏估计器$\hat{\theta}_1$和$\hat{\theta}_2$，它们都是基于样本大小为$n$的样本。根据上述第二个原则，我们更倾向于使用方差较小的估计器，即如果$\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$，则选择$\hat{\theta}_1$，否则选择$\hat{\theta}_2$。

  The spread of the sampling distribution can be measured by $\mathrm{Var}(\hat{\theta})$. The standard deviation of $\hat{\theta}$, i.e. $\sqrt{\mathrm{Var}(\hat{\theta})}$, is called the standard error. Suppose that we have two different unbiased estimators of θ, called $\hat{\theta}_1$ and $\hat{\theta}_2$, which are both based on samples of size n. By principle (ii) above, we would prefer to use the estimator with the smallest variance, i.e. choose $\hat{\theta}_1$ if $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$, otherwise choose $\hat{\theta}_2$.

点估计
Point Estimation

- 点估计的关键要素包括：
  - 数据的概率模型
  - 待估计的未知模型参数
  - 估计过程或估计器
  - 估计器的抽样分布
- The key ingredients of point estimation are:
  - A probability model for the data.
  - Unknown model parameter(s) to be estimated.
  - An estimation procedure, or estimator.
  - The sampling distribution of the estimator.
- 点估计的主要要点包括：
  - 将估计过程或估计器应用于特定的观察数据集，得出未知参数值的估计。在不同的随机数据集中，估计值会有所不同。
  - 抽样分布的性质（偏差、方差）告诉我们估计器的优劣，从而也影响我们的估计值可能有多好。
  - 由于随机抽样误差，估计过程有时可能给出较差的估计。对于良好的估计器，获得较差估计的概率较低。

点估计
Point Estimation

- The main points for point estimation are:
  - Application of the estimation procedure, or estimator, to a particular observed data set results in an estimate of the unknown value of the parameter. The estimate will be different for different random data sets.
  - The properties of the sampling distribution (bias, variance) tell us how good our estimator is, and hence how good our estimate is likely to be.
  - Estimation procedures can occasionally give poor estimates due to random sampling error. For good estimators, the probability of obtaining a poor estimate is lower..