概率论与统计相关知识

李德山

[1] 西南科技大学 经济管理学院

[2]School of Economics and Management
Southwest University of Science and Technology

2021 年 3 月 15 日

主要内容

1. 总体、随机变量与概率分布

主要内容

1. 总体、随机变量与概率分布
2. 主要统计量

主要内容

1. 总体、随机变量与概率分布
2. 主要统计量
3. 多变量分布

主要内容

主要内容

1. 总体、随机变量与概率分布
2. 主要统计量
3. 多变量分布
4. 常见分布
5. 随机抽样与大样本

主要内容

1. 总体、随机变量与概率分布
2. 主要统计量
3. 多变量分布
4. 常见分布
5. 随机抽样与大样本
6. 估计、检验与置信区间

总体、随机变量与概率分布

- **结果** (Outcome)是一个随机过程中许多相互排斥的潜在结果。例如，明天某一时刻可能是晴天，可能是阴天，也有可能是暴雨。这些不同的天气情况就是结果，但只有其中一个结果会发生。
- **概率** (Probability)指长期观测的结果发生的次数比例。
- 所有可能的结果的集合称为**样本空间**。事件是样本空间的子集。
- **总体**（Population）感兴趣的所有可能个体的集合。我们将总体视为无限大 (∞ 近似于“非常大”)。
- **随机变量** (Random Variable)一个随机结果的数值概括。（如，一个班的数学期末考试平均成绩）。分为离散型随机变量和连续型随机变量。

总体、随机变量与概率分布

- 离散随机变量的**概率分布**是所有可能的变量值及其发生的概率列表，且所有概率之和相加等于 1。
- **累积概率分布**（c.d.f.）指随机变量小于或者等于某个特定值的概率。
- 连续型随机变量的概率用**概率密度函数** (p.d.f.) 表示。任何两点之间的概率密度函数所形成的区域就是该随机变量落在这两点之间的概率。

表 2.1　你电脑死机 *M* 次的概率

	结果(死机次数)				
	0	1	2	3	4
概率分布	0.80	0.10	0.06	0.03	0.01
累积概率分布	0.80	0.90	0.96	0.99	1.00

主要统计量

- **定义** 对于分布律为 $p_k \equiv \mathrm{P}\left(X = x_k\right)$ 的离散型随机变量 X, **其期望** (expection) 为
$$\mathrm{E}(X) \equiv \mu \equiv \sum_{k=1}^{\infty} x_k p_k$$
期望的直观含义就是对 x_k 进行加权平均，权重为概率 p_k。也就是说一个随机变量 X 的期望也称为 X 的均值。

- **定义** 对于概率密度函数为 $f(x)$ 的连续型随机变量 X, 其期望为
$$\mathrm{E}(X) \equiv \mu \equiv \int_{\infty}^{\infty} x f(x) d(x)$$
上式也是对 x 进行加权平均，权重为概率密度 $f(x)$。

  - 期望满足线性性 (linearity)，即对于任意常数 k 都有
$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y), \mathrm{E}(kX) = kX$$

主要统计量

- **定义** 随机变量 X 的**方差**（variance）为

$$Var(X) \equiv \sigma^2 \equiv E[X - E(X)]^2$$

  - 方差越大，则随机变量取值的波动幅度越大。方差的平方根称之为**标准差**（Standard deviation），记为 σ。
  - 在计算方差时，常利用以下简便公式：

$$Var(X) = E(X^2) - [E(X)]^2$$

- **定义** 随机变量 X 与 Y 的**协方差** (covariance) 为

$$Cov(X, Y) = \sigma_{XY} \equiv E[(X - E(X))(Y - E(Y))]$$

  - 如果当随机变量 X 的取值大于（小于）其期望值 $E(X)$ 时，随机变量 Y 的取值也大于（小于）其期望值 $E(Y)$，则 $Cov(X, Y) > 0$, 二者存在正相关。反之，当 $Cov(X, Y) < 0$, 二者存在负相关。当 $Cov(X, Y) = 0$, 则说明二者线性不相关，但不一定是相互独立, 可能是非线性关系。

主要统计量

- 在计算协方差时，常利用以下简便公式：

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

- 协方差的运算也满足线性性，可以证明：

$$Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$$

- 协方差同方差一样存在缺点，就是它受 X 与 Y 的计量单位的影响。为将其标准化，引入相关系数。

• **定义** 随机变量 X 的与 Y 的**相关系数**（correlation）为

$$\rho \equiv Corr(X,Y) \equiv \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- 相关系数介于 -1 到 1 之间，即 $-1 \leq \rho \leq 1$。

主要统计量

- 更一般地，对于随机变量 X，可以定义一系列的统计数字特征，即各阶**矩** (moment) 的概念。

- **定义** 一阶原点矩为 $E(X)$（即期望），二阶原点矩为 $E(X^2)$, 三阶原点矩为 $E(X^3)$, 四阶原点矩为 $E(X^4)$, 等等。

- 一阶原点矩（期望）表示随机变量的平均值。
- 二阶中心矩（方差）表示随机变量的波动程度。
- 三阶中心矩表示随机变量密度函数的不对称性（偏度）。
- 四阶中心矩表示随机变量密度函数的最高处（山峰）有多“尖”及尾部有多“厚”（峰度）。

主要统计量

- **定义** 随机变量 X 的**偏度**（skewness）为 $\frac{E[(X-\mu)]^3}{\sigma_X^3}$。
- **定义** 随机变量 X 的**峰度**（kurtosis）为 $\frac{E[(X-\mu)]^4}{\sigma_X^4}$。
  - 对于正态分布，其峰度为 3。如果随机变量 X 的峰度大于 3（如 t 分布），则其密度函数的最高处比正态分布更“尖”，而两侧尾部则更“厚”，称为“厚尾”（fat tails）。存在厚尾的概率分布更容易在尾部取值，称为**极端值**（outlier）。

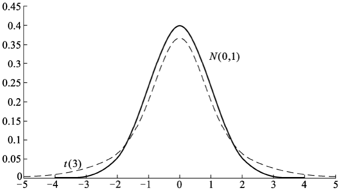


Figure: N(0,1) 和 t(3) 的概率密度

- **定义 联合概率分布**：两个离散随机变量的联合概率分布就是两个随机变量同时取得某个值时的概率。可以写成 $\Pr(X=x, Y=y)$。
- **定义 边缘概率分布**：随机变量 Y 的边缘概率分布仅仅只是 Y 的概率分布的另一种表述。它是为了区分单一变量 Y 的概率分布和 Y 与其他变量的联合概率分布。从联合概率分布中计算 Y 的边缘概率分布，就是把 Y 取某个特定值的所有概率相加。假设 X 可以取 m 个不同的值，则 Y 取 y 值的边缘概率为

$$\Pr(Y=y)=\sum_{i=1}^{m}\Pr(X=x_i, Y=y)$$

- **定义 条件分布**：给定另一随机变量 X 的特定值，随机变量 Y 的分布称为**给定 X 时 Y 的条件分布**。可以写成 $\Pr(Y=y|X=x)$。条件概率的计算公式为

$$\Pr(Y=y|X=x)=\frac{\Pr(X=x, Y=y)}{\Pr(X=x)}$$

多变量分布

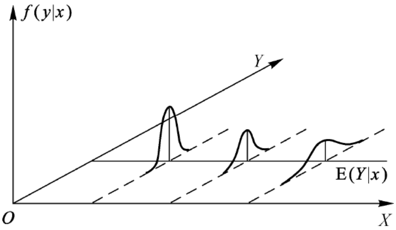
表 2.3　电脑死机次数(M)和电脑年龄(A)的联合分布和条件分布

A. 联合分布

	$M=0$	$M=1$	$M=2$	$M=3$	$M=4$	总和
旧电脑($A=0$)	0.35	0.065	0.05	0.025	0.01	0.50
新电脑($A=1$)	0.45	0.035	0.01	0.005	0.00	0.50
总和	0.8	0.1	0.06	0.03	0.01	1.00

B. 给定 A 时 M 的条件分布

	$M=0$	$M=1$	$M=2$	$M=3$	$M=4$	总和
$\Pr(M \mid A=0)$	0.70	0.13	0.10	0.05	0.02	1.00
$\Pr(M \mid A=1)$	0.90	0.07	0.02	0.01	0.00	1.00



- **定义 条件期望**（conditional expection），就是条件分布 $Y|x$ 的期望，也称**给定 X 时 Y 的条件分布的均值**。即

$$\mathrm{E}(Y|X=x)=\mathrm{E}(Y|x)=\int_{-\infty}^{+\infty} y f(y|x)dy$$

  - 由于 y 已经被积分积掉，故 $\mathrm{E}(Y|x)$ 只是 x 的函数，见图。

- **定义 条件方差**（conditional variance），就是条件分布 $Y|x$ 的方差，也称**给定 X 时 Y 的条件分布的方差**。即

$$Var(Y|X=x) = Var(Y|x) = \int_{-\infty}^{+\infty} [y - \mathrm{E}(Y|x)]^2 f(y|x)dy$$

  - 由于 y 已经被积分积掉，故 $Var(Y|x)$ 只是 x 的函数，见图。

- **定义 期望迭代法则**（Law of iterated expectation）。Y 的均值是给定 X 的条件下 Y 的条件期望以 X 的概率分布为权重的加权平均值。数学表达式：

$$\mathrm{E}(Y) = \sum_{i=1}^{m} \mathrm{E}(Y|X=x_i)\mathrm{Pr}(X=x_i)$$

  - 换句话说，Y 的期望就是在给定 X 条件下，Y 的期望的期望，即

$$\mathrm{E}(Y) = E[\mathrm{E}(Y|X)]$$

- 或者通俗的说，条件期望的期望等于无条件期望。上式的右边的内部期望是给定 X 条件下 Y 的条件期望，而外部的期望是利用 X 的边缘分布计算得到的。需要注意的是，如果给定 X 条件下 Y，那么 Y 的均值也就为 0。

- **证明**: 连续型情形

$$
\begin{aligned}
E[\mathrm{E}(Y|X)] &= \int \mathrm{E}(Y|X)]f_X(x)dx \\
&= \int (\int yf_{Y,X}(y|x)dy)f_X(x)dx \\
&= \int \int yf_{X,Y}(x,y)dxdy \\
&= \int yf_Y(y)dy \\
&= \mathrm{E}(\mathrm{Y})
\end{aligned}
$$

- 所以，回归模型的基本假设可知 $E(\mu|X) = 0$。根据期望迭代法则：$E(\mu) = E(E(\mu|X)) = 0$。

多变量分布

- 一个随机变量 X 的条件期望 $\mathrm{E}(\mathrm{X}|Y=y)$ 的值，依赖于 Y 的值 y(随着 y 的取值不同而不同)。因为 $\mathrm{E}(\mathrm{X}|Y=y)$ 是 y 的函数，所以 $\mathrm{E}(\mathrm{X}|Y)$ 是 Y 的函数，因此它也是一个随机变量。它的分布依赖于 Y 的分布。既然 $\mathrm{E}(\mathrm{X}|Y)$ 是一个随机变量，那么就应该有自己的期望 $\mathrm{E}(\mathrm{X}|Y=y)$。这个法则其实就是条件数学期望的“望远”性质，其中涉及到分割的概念，如需详细了解，参加施利亚耶夫《概率》第 79 页。
- **定义** 设 $X = (X_1, X_1, X_1 \cdots X_n)'$ 为 n 维随机向量，则其**协方差矩阵**（covariance matrix）为 $n \times n$ 的对称矩阵：

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{X}) &\equiv \mathrm{E}\left[(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))'\right] \\
&= \mathrm{E}\left[\begin{pmatrix} X_1-\mathrm{E}(X_1) \\ \vdots \\ X_n-\mathrm{E}(X_n) \end{pmatrix}\begin{pmatrix} X_1-\mathrm{E}(X_1) & \cdots & X_n-\mathrm{E}(X_n) \end{pmatrix}\right] \\
&= \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}
\end{aligned}$$

-

多变量分布

- 如果 X 和 Y 的变动趋势相同，则协方差为正。反之，如果 X 和 Y 的变动趋势相反，则协方差为负。如果 X 和 Y 独立，则协方差为 0。

- **定义** 两个随机变量 X 和 Y，如果在不提供一个随机变量的信息情况下，能得到另一个随机变量的值，那么就 X 和 Y 独立分布或者**相互独立**。$\Pr(X=x,Y=y)=\Pr(X=x)\Pr(Y=y)$。也就是说，两个独立随机变量的联合分布等于它们的边缘分布之积。

- 随机变量分布的一些性质：

$$\mathrm{E}(X+Y)=\mathrm{E}(X)+\mathrm{E}(Y)=\mu_X+\mu_Y$$

$$Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)=\sigma_X^2+\sigma_Y^2+2\sigma_X\sigma_Y$$

$$E(XY)=\sigma_{XY}+\mu_X\mu_Y$$

- **定义** 假设条件期望 $E(Y|x)$ 存在。如果 $E(Y|x)$ 不依赖于 X，则称 Y 均值独立于 X。均值独立不是一种对称的关系，即“Y 均值独立于 X”并不意味着“X 均值独立于 Y”。
- **命题** Y 均值独立于 X, 当且仅当 $E(Y|x) = E(Y)$。（证略）
- **命题** 如果 X 和 Y 相互独立，则 Y 均值独立与 X，且 X 均值独立与 Y。
- **定理** 如果 Y 均值独立于 X，或者 X 均值独立于 Y，则 $Cov(X,Y) = 0$。
  - “相互独立” $\Rightarrow$ “均值独立” $\Rightarrow$ “线性不相关”；反之不然。

- **正态分布**：如果随机变量 X 的概率密度函数为

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

则称 X 服从**正态分布**（normal distribution），记 $X\sim N(\mu,\sigma^2)$，μ 为期望，σ^2 为方差。

- 将 X 进行标准化，定义 $Z=\frac{X-\mu}{\sigma}$, 则 Z 服从标准正态分布，记 $Z\sim N(0,1)$，其概率密度函数为 $\phi(x)$。标准正态分布的概率密度以原点对称，呈钟型。其累计分布函数记为 $\Phi(x)$，见图。

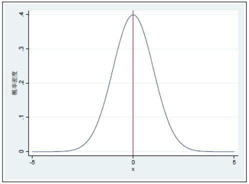


Figure: 标准正态分布的概率密度

常见分布

- χ^2(卡方分布，Chi-square)：如果 $\{Z_1, \cdots, Z_k\}$ 为独立同分布的标准正态，则其平方和服从自由度为 k 的卡方分布，记

$$\sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$$

参数 k 为自由度（df），因为上述分布由 k 个相互独立（自由）的随机变量所构成。

- χ^2 分布来自标准正态的平方和，故取值为正。$\chi^2(k)$ 分布的期望为 k, 而方差为 $2k$。

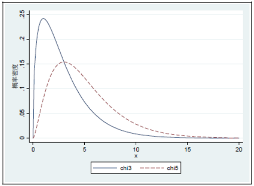


Figure: $\chi^2(3)$ 与 $\chi^2(5)$ 的概率密度

常见分布

- t **分布**：假设 $Z \sim N(0,1)$,$Y \sim \chi^2(k)$, 且 Z 与 Y 相互独立，则 $\frac{Z}{\sqrt{Y/k}}$ 服从自由度为 k 的 t 分布（也称 student 分布），记

$$\frac{Z}{\sqrt{Y/k}} \sim t(k)$$

k 为自由度。如果上式的分子与分母不相互独立，则一般不服从 t 分布。

- t 分布也以原点为对称。当自由度较小（20 及以下）时，t 分布尾部较厚。也就是说其与标准正态分布相比，中间的“山峰”更低（但更尖），而两侧有“厚尾”。当自由度大于等于 30 时，t 分布近似于正态分布。

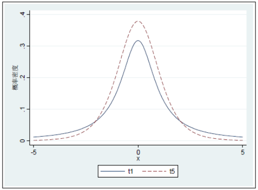


Figure: $t(1)$ 与 $t(5)$ 的密度函数

常见分布

- F **分布**：假设 $Y_1 \sim \chi^2(k_1)$,$Y_2 \sim \chi^2(k_2)$, 且 Y_1,Y_2 相互独立，则 $\frac{Y_1/k_1}{Y_2/k_2}$ 服从自由度为 k_1,k_2 的 F **分布**，记为

$$\frac{Y_1/k_1}{Y_2/k_2} \sim F(k_1, k_2)$$

  - F 分布的取值也只能为正数，其概率密度形状与 χ^2 分布相似。t 分布的平方就是 F 分布，即如果 $X \sim t(k)$, 则 $X^2 \sim F(1, k)$（证略）。

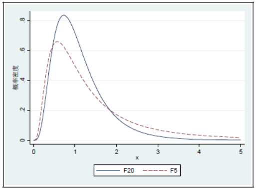


Figure: $F(10, 20)$ 与 $F(10, 5)$ 的概率密度

随机抽样与大样本

- 称我们感兴趣的研究对象全体为**总体**，其中的每个研究对象称为**个体**。
- 由于总体包含的个体可能很多，普查成本较高，故常从总体抽取部分个体，称为**样本**。
- 样本所包含的个体数目称为**样本容量**。

随机抽样与大样本

- 称我们感兴趣的研究对象全体为**总体**，其中的每个研究对象称为**个体**。
- 由于总体包含的个体可能很多，普查成本较高，故常从总体抽取部分个体，称为**样本**。
- 样本所包含的个体数目称为**样本容量**。

- **简单随机抽样**：从总体 N 个单位中任意抽 n 个单位作为样本，使每个可能的样本被抽中的概率相等的一种抽样方式。当每一个样本 Y_i 有相同的边缘分布时，我们称 Y_1、Y_2，…,Y_n 为**同分布**。
- 通常希望样本为随机样本（random sample），即总体中的每位个体都有相同的概率被抽中，且被抽中的概率相互独立，称为**独立同分布**，简记 $i.i.d.$。由于样本来自总体，必带有总体的信息。

随机抽样与大样本

- 由于抽取的每一个 Y_i 是随机的，因此它们的平均值 $\bar{Y}$ 也是随机的。

- $\bar{Y}$ 的均值、方差和标准差：

$$E(\bar{Y}) = \mu_Y$$

$$\text{var}(\bar{Y}) = \sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}$$

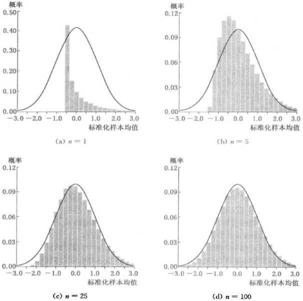
$$std.dev(\bar{Y}) = \sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$$

- 这里 $\sigma_{\bar{Y}}^2$ 是样本均值 $\bar{Y}$ 抽样分布的方差，σ_Y^2 表示抽样的总体分布方差。这些结论对 Y_i 的任意分布都成立，不需要 Y_i 服从某个特定分布。

随机抽样与大样本

- 目前，有两种方法刻画抽样分布：精确法和近似法。
- 精确分布又称为**有限样本分布**。近似法是样本容量较大时抽样分布的近似分布。常称抽样分布的这种大样本近似分布为**渐进分布**。说“渐进”是因为当 $n \to \infty$ 时这种近似式精确的。
- 当样本很大时，有两个很重要的法则：**大数定律和中心极限定理**。
- 大数定律是当样本规模很大时，$\bar{Y}$ 以非常高的概率接近 μ_Y。
- 中心极限定理是当样本容量很大时，标准化样本均值 $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$ 的抽样分布近似于正态分布。因此，渐进正态分布并不依赖于 Y 的分布。这一正态近似分布极大简化了问题，同时它也是回归理论的基础。

- 如果当 n 增大时，对任意常数 $c > 0$，$\bar{Y}$ 落入区间 $\mu_Y - c$ 和 $\mu_Y + c$ 的概率充分接近于 1，则样本均值 $\bar{Y}$ 依概率收敛域 μ_Y。记为 $\bar{Y} \xrightarrow{p} \mu_Y$。
- n“足够大”是多大呢？答案是“视情况而定”。在 Y_i 服从正态分布时，对所有的 n，$\bar{Y}$ 精确服从正态分布。当 Y_i 不是正态分布时，要求 $n \geq 30$。注：即使样本容量只有 30 个观测值时，近似也是非常精确的。因为应用微观计量经济学的样本容量通常是成千上万，故这些渐进分布能为精确抽样分布提供一个较好的近似。



估计、检验与置信区间

- 统计推断就是根据样本数据，对总体性质进行推断的科学。
- 计量经济学中使用的统计方法主要有：**估计、假设检验和置信区间**。
- 估计就是从样本数据中，为一个总体分布特征（均值、方差等）计算出一个“最佳猜测值”。
- 假设检验就是提出一个假设，然后利用样本证据来验证假设是否为真。
- 置信区间就是利用一组样本数据来估计未知总体分布特征的范围或区间。

估计、检验与置信区间

- **估计量**是从总体中随机抽取的样本数据的函数。**估计量**是基于某一特定的样本数据实际计算得到的估计量的数值。由于抽样的随机性，估计量是一个随机变量，而估计值是一个非随机的数。那么，在众多的估计量中，如何评价一个估计量比另外一个“更好”？

无偏、一致和有效性

假设 $\hat{\mu}_Y$ 是 μ_Y 的一个估计量，则

1. $\hat{\mu}_Y$ 的偏差（bias）为 $E(\hat{\mu}_Y) - \mu_Y$。
2. 若 $E(\hat{\mu}_Y) = \mu_Y$，则 $\hat{\mu}_Y$ 是 μ_Y 的无偏估计量。
3. 若 $\hat{\mu}_Y \xrightarrow{p} \mu_Y$, 则 $\hat{\mu}_Y$ 是 μ_Y 的一个一致估计量。
4. 若 $\tilde{\mu}_Y$ 是 μ_Y 的另外一个估计量，且设 $\tilde{\mu}_Y$ 和 $\hat{\mu}_Y$ 都是无偏的。若 $\mathrm{var}(\hat{\mu}_Y) < \mathrm{var}(\tilde{\mu}_Y)$, 则称 $\tilde{\mu}_Y$ 比 $\hat{\mu}_Y$ 有效。

- 1. **无偏性**: 如果你通过重复抽样来估计一个估计量，一般你会得到一个“真值”。因此，一个优良的估计量就是其抽样分布的均值等于总体均值。
- 2. **一致性**: 当样本容量很大时，由于样本随机变动引起 $\hat{\mu}_Y$ 的值不确定性就非常小。或者说，$\hat{\mu}_Y$ 落入真值 μ_Y 的一个小区间内的概率接近 1。
- 3. **有效性**: 如果你有两个无偏估计量，你应该选择最小方差的估计量。方差最小的估计量更有效（或者说方差最小的估计量更有效的利用了数据的信息）。

- 如果 $\bar{Y}$ 满足上述三个标准，此时，我们就可以把样本均值 $\bar{Y}$ 称之为**最佳线性无偏估计量**（Best Linear Unbiased Estimator，简称 BLUE）

- 待检验的假设称之为**原假设**（null hypothesis）。如果原假设不成立，那么备择假设成立。在统计学中，原假设通常为总体均值等于某一特定值，用 H_0 来表示，即：$H_0 : E(Y) = \mu_{Y,0}$。备择假设为：$H_1 : E(Y) \neq \mu_{Y,0}$，这种类型被称为双边备择假设。
- 现实中，我们不可能知道总体均值，只能用随机抽样的样本均值 $\bar{Y}$ 来替代，但由于 $\bar{Y}$ 不可能精确等于 $\mu_{Y,0}$。虽然样本数据无法提供有关原假设的确凿证据，但我们可以计算一个概率来允许检验原假设。

- p 值，也称为**显著性概率**（$sig.p$）：在原假设成立的情形下，抽到的统计量与原假设之间的距离至少等于其样本计算值与原假设之间距离的概率。当 p 值很小，则说明原假设成立时不太可能抽到这种样本，因此认为原假设不成立是合理的。反之，p 值较大时，不能拒绝原假设是合理的。

估计、检验与置信区间

- 数学语言描述 p 值的含义：

$$p\_\text{vale} = \Pr_{H_0}[|\bar{Y} - \mu_{Y,0}| \geq |\bar{Y}^{act} - \mu_{Y,0}|]$$

- $\bar{Y}^{act}$ 表示用实际数据计算的样本均值，$\Pr_{H_0}$ 表示原假设为真时计算出来的概率。p 值在原假设下 $\bar{Y}$ 的分布位于 $|\bar{Y}^{act} - \mu_{Y,0}|$ 之外的尾部面积。当 p 值较大时，观测到的 $\bar{Y}^{act}$ 与原假设相符，当 p 值较小时则不相符。
- 为了计算 p 值，必须知道原假设下 $\bar{Y}$ 的抽样分布，根据之前的知识，当样本容量很大时，无须知道 $\bar{Y}$ 的总体分布就可以用大样本正态近似来计算 p 值。但详细的计算依赖于 σ_Y^2 是否已知。

- 当 σ_Y^2 已知时，p 值为图中的阴影面积：

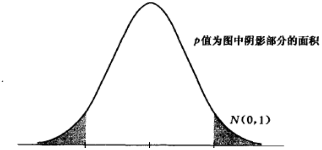


Figure: p 值

- 但是，在实践中，σ_Y^2 是未知的。由于样本方差 s_Y^2 为 σ_Y^2 的一致估计量，因此可以用标准误 $SE(\bar{Y}) = \hat{\sigma}_{\bar{Y}}$ 来替换 $\sigma_{\bar{Y}}$ 计算 p 值。
- $s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, 这里分母是 $n-1$ 而不是 n, 原因在于均值估计用完了数据中的某些信息，即用了一个“自由度”，所以只剩下 $n-1$ 个自由度。另外，当样本容量较大时，根据大数定律样本方差是总体方差的一致估计量。

估计、检验与置信区间

- t **统计量**：

$$t = \frac{\bar{Y} - \mu_{Y,0}}{SE(\bar{Y})}$$

  - 当样本容量很大时，t 分布近似服从标准正态分布 $N(0,1)$。
  - 统计假设检验常犯的两类错误：（1）第一类错误。原假设为真时却被拒绝；（2）第二类错误。原假设为假时却没有拒绝。
  - 称原假设为真时假设检验中事先给定的拒绝概率（即事先给定的犯第一类错误的概率）为检验的**显著性水平**。假设检验的临界值为给定显著水平下恰好拒绝原假设的统计量取值。拒绝原假设的统计量的取值集合为拒绝域。而不能拒绝原假设的统计量的取值集合为接受域。

- 实践中，常用的显著性水平有：10%、5%、1%。注：显著性水平不是一个固定不变的数值，依据拒绝区间所能承担的风险来决定；统计学所讲的统计显著性与实际生活工作中的经济显著性是不一样的。

置信区间

- 90% :$\mu_Y = [\bar{Y} \pm 1.64SE(\bar{Y})]$
- 95% :$\mu_Y = [\bar{Y} \pm 1.96SE(\bar{Y})]$
- 99% :$\mu_Y = [\bar{Y} \pm 2.58SE(\bar{Y})]$

- **定义** 以估计量 $\hat{\theta}$ 来估计参数 θ（估计值），则其均方误差（MSE）为

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

- 最优的估计量应在所有估计量中具有最小的均方误差。估计量 $\hat{\theta}$ 的均方误差可以分解为 $\hat{\theta}$ 的方差与偏差平方之和，即 (证略)：

$$MSE(\hat{\theta}) = \mathrm{var}(\hat{\theta}) + [Bias(\hat{\theta})]^2$$

- 因此，均方误差最小化，可以视为在“估计量方差”与“偏差”之间进行 trade-off。

参考文献：

Stock J.H., Watson M.W. Introduce to Econometrics, 3rd edition, 2011.

陈强. 计量经济学及 Stata 应用 [M]. 高等教育出版社，2015.