

内 容 简 介

本书记录了作者在山大数院的三年所学的数学知识，简明扼要地列举出每个学科所必须掌握的定理等重要知识点。此书不宜作为初学某一科目的参考资料，而适合已学完部分内容者查漏补缺。

根据山大本科开课顺序以及个人自学进度，大致收录下列学科：

大一上：数学分析 1，高等代数 1，解析几何

大一下：数学分析 2，高等代数 2

大二上：数学分析 3，复变函数，常微分方程

大二下：实变函数，偏微分方程，概率论

大三上：机器学习、数字图像处理、数理统计、矩阵论

大三下：时间序列分析、数据库系统、数据结构

目 录

第一章 概率论与数理统计	1
1.1 随机事件与概率	2
1.1.1 随机事件	2
1.1.2 概率	2
1.1.3 概率的性质	3
1.1.4 条件概率	3
1.2 随机变量及其分布	4
1.2.1 随机变量	4
1.2.2 常用分布及概率密度函数	5
1.2.3 数字特征	5
1.2.4 随机变量函数的分布	6
1.3 多元随机变量及其分布	6
1.3.1 多元随机变量	6
1.3.2 边缘分布	7
1.3.3 多元随机变量函数的分布	7
1.3.4 多元随机变量的特征数	8
1.3.5 条件分布与条件期望	9

1.4	大数定律和中心极限定理	9
1.4.1	随机变量的特征函数	9
1.4.2	大数定律	10
1.4.3	中心极限定理	11
1.5	数理统计基本概念	11
1.5.1	统计学基本思想	12
1.5.2	常用统计量	12
1.5.3	抽样分布	12
1.5.4	利用抽样分布统计推断	13
1.5.5	充分统计量	14
1.6	参数估计	14
1.6.1	矩估计	15
1.6.2	最大似然估计	15
1.6.3	点估计的评价标准	16
1.6.4	贝叶斯估计	17
1.6.5	区间估计	18
1.7	假设检验	19
1.7.1	基本思想	20
1.7.2	正态总体假设检验	20
1.7.3	广义似然比检验	22
1.7.4	拟合优度检验	22
1.7.5	正态性检验	22
1.7.6	游程检验	23
1.8	方差分析	23
1.8.1	基本思想	24

1.8.2	单因素方差分析	24
1.8.3	方差齐性检验	25
1.9	回归分析	25
1.9.1	基本思想	26
1.9.2	回归系数的最小二乘估计	26
1.9.3	区间估计与预测	26
1.9.4	显著性检验	27
1.9.5	多元线性回归	27

第一章 概率论与数理统计

Mathematical Analysis

概率论与数理统计是由数分高代派生出来的应用学科,用于刻画日常生活中随机发生的事件,具有很高的应用价值. 其中,概率论主要研究随机变量的分布与特征,而数理统计主要研究通过样本对未知分布进行估计.

概率论的重点: 概率的定义, 条件概率与独立性, 一元或多元随机变量分布, 常用分布函数, 随机变量的特征数, 大数定律和中心极限定理

数理统计的重点: 基本概念与三大分布, 参数估计, 假设检验, 方差分析, 回归分析

1.1 随机事件与概率

之前数学分析研究的内容都是具有确定解析式或约束条件的函数, 但概率论引进了随机因素, 即实验和结果并不是一一对应的, 一次实验可能会出多种结果. 这一部分的任务是使用概率这一量化方式, 将随机性规范化.

1.1.1 随机事件

1. 随机现象: 重复实验会出现不同结果的现象.
2. 样本空间: 随机现象可能出现的结果组成的集合.
3. 随机事件: 样本空间的子集. 当实验结果属于此子集时, 称随机事件发生.
4. 随机变量: 用于描述随机事件的人为设定变量 (非正式定义).
5. 事件的运算: 和集合一致, 有交并补余四大运算. 有两个公式很重要.
 - (1) 集合减法公式: $A - B = A \cap \bar{B}$.
 - (2) 德摩根律: $\overline{A \cup B} = \bar{A} \cap \bar{B}; \overline{A \cap B} = \bar{A} \cup \bar{B}$.
6. 事件域: 令 Ω 为样本空间, 定义事件域 \mathcal{F} 符合下列性质:
 - (1) $\Omega \in \mathcal{F}$; (2) $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$; (3) $A_n \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

1.1.2 概率

1. 公理化定义: 在事件域 (Ω, \mathcal{F}) 上定义可测函数 $P(A)$ 满足:
 - (1) 非负性: $P(A) \geq 0$; (2) 正则性: $P(\Omega) = 1$;
 - (3) 可列可加性: 事件 A_1, \dots, A_n 互不相容时, $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_i)$.
2. 用频率定义概率: 令 $n(A)$ 为事件 A 发生的频数, 则可用大量重复事件的频率表示概率: $P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n}$.

3. 古典概型: 若样本空间有 n 个等可能发生的样本点, 则事件 A 包含 k 个样本点时, $P(A) = \frac{k}{n}$.

4. 几何概型: 若样本空间 Ω 的面积测度为 S_n , 事件 A 包含其中面积为 S_A 的一部分, 则 $P(A) = \frac{S_A}{S_n}$. (蒙特卡罗法的理论依据)

5. 贝叶斯概率: 对事件发生可能性的主观预测, 在机器学习中使用频率很高.

1.1.3 概率的性质

1. 有限可加性: 若 A_1, \dots, A_n 互不相容, 则 $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.
2. 单调性: 若 $A \subset B$, 则 $P(A) \leq P(B)$.
3. 加法公式: $P(A \cup B) = P(A) + P(B) - P(AB)$.

1.1.4 条件概率

1. 定义: $P(A|B)$ 表示已知 B 发生的条件下 A 发生的概率. $P(A|B) = \frac{P(AB)}{P(B)}$.
2. 乘法公式: $P(AB) = P(B)P(A|B)$, 即定义式的变种.
3. 全概率公式: 若 B_i 互不相容, 且 $\bigcup_{i=1}^n B_i = \Omega$, 则

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

4. 贝叶斯公式: 用先验概率推后验概率. 若 B_i 互不相容, 且 $\bigcup_{i=1}^n B_i = \Omega$, 则

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{k=1}^n P(A|B_k)P(B_k)}$$

5. 独立性: 若 $P(A|B) = P(A)$, 即 $P(AB) = P(A)P(B)$, 则称事件 A 和 B 相

互独立.

1.2 随机变量及其分布

用概率描述随机事件发生可能性的大小后, 为了更充分认识随机事件, 我们引入随机变量来刻画随机事件, 如抽奖是随机事件, 在此基础上可以定义随机变量“是否中奖”, 这是一个二值随机变量 (0/1).

使用随机变量来描述随机事件, 能更方便地研究随机事件中我们感兴趣的性质, 比如随机变量“灯泡坏掉的个数”能帮助我们衡量灯泡的寿命. 这些随机变量取值的规律可以用分布来描述, 离散随机变量和连续随机变量的刻画方式略有区别.

1.2.1 随机变量

1. 定义: 样本空间 Ω 上的实值函数 $X(\omega)$.
2. 离散随机变量的确定: 使用分布列描述.

X	X_1	X_2	\cdots	X_n
P	p_1	p_2	\cdots	p_n

其中 p_i 表示随机变量 X 取值 X_i 的概率, $\sum_{i=1}^n p_i = 1$.

3. 连续随机变量的描述: 使用分布函数与概率密度函数.

(1) 分布函数 $F(x)$: $F(x) = P(X \leq x)$, 是单调递增的右连续函数, 且 $F(+\infty) = 1, F(-\infty) = 0$.

(2) p.d.f 概率密度函数 $p(x)$: $p(x) = F'(x)$, 是非负函数且 $\int_{-\infty}^{+\infty} p(x)dx = 1$.

1.2.2 常用分布及概率密度函数

1. 离散分布

(1) 泊松分布: $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, 用于计数过程, 记作 $X \sim P(\lambda)$.

(2) 伯努利分布: $P(X = 1) = p, P(X = 0) = 1 - p$, 又称两点分布.

(3) 二项分布: $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, 即 n 重伯努利分布中事件发生的次数, 记作 $X \sim b(n, p)$.

(4) 几何分布: $P(X = k) = (1 - p)^{k-1} p$, 具有无记忆性.

2. 连续分布

(1) 正态分布: $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, 是最常用的分布. 记作 $X \sim N(\mu, \sigma^2)$, 标准正态分布即 $N(0, 1)$.

(2) 均匀分布: $p(x) = \frac{1}{b-a}$, 其中 $x \in (a, b)$, 记作 $X \sim U(a, b)$.

(3) 指数分布: $p(x) = \lambda e^{-\lambda x}$, 其中 $x \geq 0$, 记作 $X \sim \epsilon(\lambda)$, 具有无记忆性.

(4) 伽马分布: $p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$, 其中 $x \geq 0$, 记作 $X \sim Ga(\alpha, \lambda)$. 特殊地, $Ga(\frac{n}{2}, \frac{1}{2}) = \chi^2(n)$ 为卡方分布, 统计中常用.

1.2.3 数字特征

1. 数学期望: X 在不同取值数按概率的加权平均数, 是消除随机性的主要手段, 记作 Ex . 在离散场合, $EX = \sum_{n=1}^{\infty} p_i x_i$. 在连续场合, $EX = \int_{-\infty}^{+\infty} xp(x)dx$.

2. 方差: $DX = E[(X - EX)^2]$, 也记作 $\text{Var}(X)$, 用于衡量数据的集中程度. 常用的计算公式为

$$DX = E(X^2) - (EX)^2$$

3. 标准差: $\sigma(x) = \sqrt{DX}$, 也记作 $Std(X)$, 好处是与 X 的量纲一致.

4. 切比雪夫不等式: $P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}$.

1.2.4 随机变量函数的分布

1. 离散情形: 先求各项的像 $g(x_1), \dots, g(x_n)$, $g(x_i)$ 对应概率仍为 p_i , 再合并相同项.

2. 连续情形: 若 $Y = g(x)$ 严格单调, 反函数为 $x = h(y)$, X 的概率密度函数为 $p(x)$, 则 Y 的概率密度函数为 $p_Y(y) = p_X(h(y)) \cdot |h'(y)|$. 一般情况下, 需要根据 $P(g(x) \leq y)$ 反解出 x 的范围, 再利用 X 的分布函数求解.

1.3 多元随机变量及其分布

若样本点含有不止一个我们感兴趣的属性, 如身体指标包含身高和体重, 则可定义多元随机变量来刻画这些指标的分布. 研究多元随机变量, 除了明确各分量的分布外, 还需要研究各分量间的相关关系, 以及给定某条件后的分布情况.

事实上, 只要给定多元随机变量的联合分布, 就能得到所有信息, 该部分的目的就是掌握将信息从联合分布中提取出来的方法.

1.3.1 多元随机变量

1. 定义: 样本空间 Ω 上的向量值函数 $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$.
2. 联合分布函数: $F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$
3. 离散情形的联合分布列: 仅用于二元分布 (X, Y) , 用 i 行 j 列元素 p_{ij} 表示 $X = X_i, Y = Y_j$ 的概率, 其中 $\sum_{i,j} p_{ij} = 1$.
4. 连续情形的联合密度函数: $p(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$.
5. 多元正态分布: 最重要的多元连续分布. 令 $x = (x_1, \dots, x_n)$, 均值向量为

μ , 协方差矩阵为 Γ , 则 n 元正态分布的联合密度函数

$$p(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Gamma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Gamma^{-1}(x-\mu)}$$

特殊地, 当 $n = 2$ 时, 二元正态分布为

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2 - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + (\frac{y-\mu_2}{\sigma_2})^2]}$$

记作 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

1.3.2 边缘分布

1. 边缘分布函数: $F_x(x) = F(x, \infty)$, $F_y(y) = F(\infty, y)$.
2. 离散情形的边缘分布列: $P(X = X_i) = \sum_j P(X = X_i, Y = Y_j)$; $P(Y = Y_j) = \sum_i P(X = X_i, Y = Y_j)$
3. 连续情形的边缘密度函数: $p_X(x) = \int_{-\infty}^{\infty} p(x, y) dy$; $p_Y(y) = \int_{-\infty}^{\infty} p(x, y) dx$.
4. 随机变量的独立性: 若 $\prod_{i=1}^n p_i(X_i) = p(x_1, \cdots, x_n)$, 即联合密度函数为边缘密度函数之积, 则称 X_1, \cdots, X_n 相互独立.

1.3.3 多元随机变量函数的分布

1. $Z = X + Y$ 的分布: 可用后面提到的特征函数法, 也可用卷积公式, 即 $p_Z(z) = \int_{-\infty}^{\infty} p_X(x)p_Y(z-x)dx$.
2. 次序统计量分布: 若 $X_{(1)}, \cdots, X_{(n)}$ 独立同分布且升序排列, 则第 k 个次

序统计量 $X_{(k)}$ 的概率密度函数为

$$p_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} p(x) [1-F(x)]^{n-k}$$

特殊地, $\min X$ 即 $X_{(1)}$ 的概率密度函数为 $n[1-F(x)]^{n-1}p(x)$;

$\max X$ 即 $X_{(n)}$ 的概率密度函数为 $n[F(x)]^{n-1}p(x)$.

3. 变量变换法: 令 $u = u(x, y), v = v(x, y)$, 从中反解出 $x = x(u, v), y = y(u, v)$, 则 $p(u, v) = p(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|$.

1.3.4 多元随机变量的特征数

1. 数学期望: $g(x, y)$ 的期望为 $\int_R g(x, y)p(x, y)dxdy$.

2. 方差: 定义不变, 仍有 $\text{Var}(x) = E(X - EX), \text{Var}(y) = E(Y - EY)$.

3. 协方差: $\text{Cov}(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - EXEY$, 用于刻画两变量的相关程度.

4. 相关系数: $\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$. 当 $\text{Corr}(x) \in (0, 1]$ 时, 称 X 和 Y 正相关; $\text{Corr}(x) \in [-1, 0)$ 时, 称 X 和 Y 负相关; $\text{Corr}(x) = 0$ 时, 称 X 和 Y 不相关.

5. 方差运算性质: $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$

6. n 元随机变量的协方差矩阵:

$$\Gamma = \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Cov}(X_n, X_n) \end{pmatrix}$$

用于刻画各分量之间的总体相关性.

1.3.5 条件分布与条件期望

1. 离散条件分布: $P_{i|j} = P(X = X_i | Y = y_j) = \frac{p_{ij}}{\sum_i p_{ij}}$.
2. 连续条件分布: $p(y|x) = \frac{p(x,y)}{p_X(x)}$; $p(x|y) = \frac{p(x,y)}{p_Y(y)}$
3. 全概率公式: $p_Y(y) = \int_{-\infty}^{\infty} p_X(x)p(y|x)\mathrm{d}x$.
4. 贝叶斯公式:

$$p(x|y) = \frac{p(y|x)p_X(x)}{\int_{-\infty}^{\infty} p(y|x)p_X(x)\mathrm{d}x}$$

5. 条件数学期望: $E(X|Y = y) = \int_{-\infty}^{\infty} xp(x|y)\mathrm{d}x$.
6. 重期望公式: $E[E(X|Y)] = EX$.

1.4 大数定律和中心极限定理

这部分首先将傅里叶变换引入概率密度函数的求解中, 得到特征函数这个很好用的工具, 再借助特征函数推导大数定律和中心极限定理的一般结论, 为数理统计的展开做好铺垫.

大数定律的内容很简单, 就是抽样次数足够大时, 频率近似于概率, 均值近似于数学期望, 这给大样本统计提供了理论依据. 中心极限定理说明多个独立同分布随机变量之和近似于正态分布, 这鼓励我们在大样本统计中使用正态分布进行统计推断.

1.4.1 随机变量的特征函数

1. 定义: $\varphi(t) = E(e^{itX})$. 离散情形下, $\varphi(t) = \sum_{k=1}^{\infty} p_k e^{itX_k}$; 连续情形下, $\varphi(t) = \int_{-\infty}^{\infty} p(x)e^{itx}\mathrm{d}x$.
2. 性质: (1) 若 X 与 Y 独立, $Z = X + Y$, 则 $\varphi_Z(t) = \varphi_X(t) \cdot \varphi_Y(t)$.

(2) 求各阶矩的方式: $\varphi^{(k)}(0) = i^k E(X^k)$.

(3) 唯一性定理: 分布函数由特征函数唯一确定.

3. 逆转公式: $F(x_2) - F(x_1) = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t) dt$.

4. 连续随机变量的逆变换公式: $p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt$.

1.4.2 大数定律

1. 一般形式: 对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow +\infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n EX_i \right| < \varepsilon \right) = 1$$

2. 伯努利大数定律: 令 S_n 为 n 重伯努利试验中事件发生的次数, p 为事件发生的概率, 则对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow +\infty} P \left(\left| \frac{S_n}{n} - p \right| < \varepsilon \right) = 1$$

3. 切比雪夫大数定律: 当 $\{X_n\}$ 两两不相关且 $Var(X_i)$ 有界时, 大数定律成立.

4. 辛钦大数定律: 当 X_1, \dots, X_n 独立同分布且 EX_i 存在时, 大数定律成立.

5. 辛钦大数定律的证明: 令 $\varphi(t)$ 为 X_i 共同的特征函数, 数学期望为 a , 将 $\varphi(t)$ 在 $t = 0$ 处泰勒展开: $\varphi(t) = 1 + iat + o(t)$. 故 $\varphi_{\frac{1}{n} \sum_{i=1}^n X_i}(t) = [\varphi(\frac{t}{n})]^n \sim e^{iat}$, 恰是退化分布的特征函数.

1.4.3 中心极限定理

1. 林德伯格-莱维中心极限定理: 令 X_n 独立同分布, $EX_i = \mu$, $DX_i = \sigma^2$, 则 $n \rightarrow \infty$ 时, $\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$ 的分布弱收敛于标准正态分布.

2. 棣莫弗-拉普拉斯中心极限定理: 令 S_n 为 n 重伯努利试验中事件发生的次数, p 为事件发生的概率, 则 $\frac{S_n - np}{\sqrt{np(1-p)}}$ 在 $n \rightarrow \infty$ 时的分布弱收敛于标准正态分布.

3. 中心极限定理的证明: 将 X_i 标准化: $Y_i = \frac{X_i - \mu}{\sigma}$, 则 $EY_i = 0$ 且 $DY_i = 1$. 令 Y_N 的特征函数为 $\varphi(t)$, 故 $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$ 的特征函数为 $\left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n$. 由泰勒展开: $\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o(t^2)$, 在 $n \rightarrow \infty$ 时, $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$ 的特征函数趋近于 $e^{-\frac{t^2}{2}}$, 恰为 $N(0, 1)$ 的特征函数.

1.5 数理统计基本概念

概率论研究的随机变量都有确定的总体, 而现实生活中, 我们通常需要推断某一总体服从何种分布. 这就是数理统计的核心任务: 推断总体服从的分布族, 以及通过样本估计分布中的未知参数. 由于总体的性质只能通过抽样反馈, 因此需要研究由样本推断总体的方法.

在研究过程中, 通常假设各样本与总体同分布且相互独立, 并利用统计量的分布进行推断, 这一思想贯穿了后面的所有章节. 而这部分的任务是打好基础, 理清数理统计的基本概念, 并初步介绍最常用的统计量及其抽样分布, 为参数估计和假设检验打好基础.

1.5.1 统计学基本思想

1. 任务: 收集受随机因素影响的数据, 并根据样本推断总体分布.
2. 总体: 研究对象的全体. 具体分布未知, 一般认为分布族已知, 即推断分布中的未知参数.
3. 样本: 从总体中随机抽取的 n 个数据, 记作 x_1, x_2, \dots, x_n . 若这些样本独立同分布 (i.i.d.), 则称为简单随机样本.
4. 统计量: 当总体分布族已知而参数未知时, 可以构造只与样本有关而与未知参数无关的函数 $T = T(x_1, \dots, x_n)$, 利用统计量的特征估计未知参数.

1.5.2 常用统计量

1. 样本均值: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.
2. 样本方差: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$. 计算时常用公式

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

3. 样本标准差: $s = \sqrt{s^2}$.
4. 样本 k 阶矩: $a_k = \frac{1}{n} \sum_{i=1}^n x_i^k$.

1.5.3 抽样分布

1. 定义: 统计量的分布称为抽样分布.
2. χ^2 分布: 若简单随机样本 $x_1, \dots, x_n \sim N(0, 1)$, 则 $\sum_{i=1}^n x_i^2$ 服从自由度为 n 的 χ^2 分布, 记作

$$\sum_{i=1}^n x_i^2 \sim \chi^2(n)$$

其概率密度函数只在第一象限定义且非对称. 不用刻意记忆其具体 p.d.f., 但需注意 $\chi^2(2)$ 的概率密度函数为 $\frac{1}{2}e^{-\frac{1}{2}x} (x > 0)$.

3. F 分布: 令独立随机变量 $X \sim \chi^2(m), Y \sim \chi^2(n)$, 则 $\frac{X/m}{Y/n}$ 服从自由度为 m 与 n 的 F 分布, 记作

$$\frac{X/m}{Y/n} \sim F(m, n)$$

F 分布具有的特殊性质: 若 $X \sim F(m, n)$, 则 $\frac{1}{X} \sim F(n, m)$.

4. t 分布: $X \sim N(0, 1), Y \sim \chi^2(n)$, 则 $\frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布, 记作

$$\frac{X}{\sqrt{Y/n}} \sim t(n)$$

t 分布与 F 分布间存在联系: 若 $X \sim t(n)$, 则 $X^2 \sim F(1, n)$.

1.5.4 利用抽样分布统计推断

1. 前置条件: x 服从正态分布, 即 $x \sim N(\mu, \sigma^2)$.

2. σ^2 已知, 对 μ 统计推断: 构造统计量

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

3. σ^2 未知, 对 μ 统计推断: 由 \bar{x} 与 s^2 相互独立, 可构造统计量

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

4. 对 σ^2 统计推断:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

5. 双正态总体方差比推断: 令 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, 构造统计量

$$\frac{s_X^2}{s_Y^2} \sim F(m-1, n-1)$$

1.5.5 充分统计量

1. 定义: 若统计量 T 包含了样本的全部信息, 即给定 T 的取值后, x_1, \dots, x_n 的分布与未知参数 θ 无关, 则称 T 为 θ 的充分统计量.

2. 因子分解定理: 若总体分布为 $f(x; \theta)$, 存在函数 $g(T, \theta)$ 与 $h(x_1, \dots, x_n)$ 使得 $f(x_1, \dots, x_n; \theta) = g(T, \theta) \cdot h(x_1, \dots, x_n)$, 则 T 为 θ 的充分统计量.

1.6 参数估计

参数估计的目的是对分布族已知, 但含有未知参数的总体, 通过样本估计其中的未知参数. 一种思路为点估计, 即给出参数的确切估计值; 另一种思路为区间估计, 即给出一个大致范围, 有很大的可能性包含参数的真实值. 点估计的评价标准为无偏性和有效性, 即样本越多, 估计值越接近真实值, 且波动尽可能小; 区间估计的手段是通过抽样分布的分位数, 划定统计量所处的范围以包含分布中比例为 $1 - \alpha$ 的部分, 再解出参数所处的范围.

考虑到与机器学习接轨, 这一部分列举了很多超纲的内容, 如用先验推后验的贝叶斯估计, 以及求 ML 估计的 EM 算法, 可视自身需要加以取舍.

1.6.1 矩估计

1. 思想: 点估计的一种, 另一种即下面讨论的最大似然估计. 令总体分布为 $X(\theta)$. 用样本矩 $a_k = \sum_{i=1}^n X_i^k$ 代替总体矩 $EX^k(\theta)$, 列方程求解未知参数.

2. 以正态分布的矩估计为例: 令 $X \sim N(\mu, \sigma^2)$, 则

$$\begin{cases} EX = \mu &= \frac{1}{n} \sum_{i=1}^n x_i \\ EX^2 = \sigma^2 + \mu^2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 \end{cases}$$

从中解出 $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s_n^2$.

3. 矩估计的相合性: 若 $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$, $\lim_{n \rightarrow \infty} D(\hat{\theta}) = 0$, 则称 $\hat{\theta}$ 为 θ 的相合估计. 矩估计通常为相合估计.

1.6.2 最大似然估计

1. 思想: 选取参数 θ , 使得样本概率 $f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$ 最大.

2. 求解方式: 令似然函数 $L(\theta) = \prod_{i=1}^n f(x_i; \theta)$, 求解方程以解出 $\hat{\theta}$:

$$\frac{d \ln L(\theta)}{d\theta} = 0$$

3. 以正态分布的 ML 估计为例: 令 $X \sim N(\mu, \sigma^2)$, 则未知参数 θ 由 μ 和 σ^2

构成. 求解下列方程组:

$$\begin{cases} \frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = -\sum_{i=1}^n \frac{\mu - x_i}{\sigma^2} = 0 \\ \frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(\mu - x_i)^2}{2\sigma^4} = 0 \end{cases}$$

解得 $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = s_n^2$, 恰好与矩估计结果一致.

4. EM 算法: Expectation Maximization, 针对似然函数中存在不可观测的隐变量 z 时的局部 ML 优化.

(1) E 步: 构造似然函数 $Q(\theta|x, \theta^{(i)}) = E_z[\ln L(\theta; x, z)]$, 目的是消除隐变量 z 的随机性;

(2) M 步: 在已知上轮迭代值 $\theta^{(i)}$ 和样本 x 的情况下, 寻找使似然函数最大的局部最优解:

$$\theta^{(i+1)} = \arg \max_{\theta} Q(\theta|x, \theta^{(i)})$$

(3) 迭代: 设定初始值 $\theta^{(0)}$, 重复 E 步和 M 步直至收敛.

1.6.3 点估计的评价标准

1. 无偏性: 若 $E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 为 θ 的无偏估计.
2. 有效性: 若 $\hat{\theta}_1, \hat{\theta}_2$ 均为 θ 的无偏估计, 且 $D\hat{\theta}_1 > D\hat{\theta}_2$, 则估计 $\hat{\theta}_2$ 比估计 $\hat{\theta}_1$ 更有效.
3. Fisher 信息量: $I(\theta) = E \left[\frac{\partial}{\partial \theta} \ln f(x; \theta) \right]^2$. $I(\theta)$ 越大, 表示总体分布中包含未知参数 θ 的信息越多.

4. Cramer-Rao 不等式: 若 T 为 $g(\theta)$ 的无偏估计, 则

$$DT \geq \frac{[g'(\theta)]^2}{nI(\theta)}$$

若 DT 取到 C-R 下界, 则称 T 为 $g(\theta)$ 的有效估计.

1.6.4 贝叶斯估计

1. 思想: 在抽样之前, 便有关于 θ 的先验信息, 即 θ 服从先验分布 $\pi(\theta)$. 以后验信息

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\theta} f(x|\theta)\pi(\theta)d\theta}$$

以后验分布的最大值点作为 θ 的点估计.

2. 朴素贝叶斯分类器: 假设各属性 A_1, \dots, A_n 相互独立, 在得到新的样本 $A_1 = a_1, \dots, A_n = a_n$ 后, 尝试将样本归类: $Y \in y_1, \dots, y_m$.

朴素贝叶斯分类器的最大化目标为: 以样本信息为先验, 寻找可能性最大的分类结果, 即

$$k = \arg \max_k P(Y = y_k | A_1 = a_1, \dots, A_n = a_n)$$

由贝叶斯公式以及朴素假设, 最终优化目标为

$$k = \arg \max_k \prod_{i=1}^n P(A_i = a_i | Y = y_k)$$

每一项都可以通过现有样本点在 $Y = y_k$ 时 $A_i = a_i$ 的占比求出, 选出使优化目标最大的 k , 便可作出最优决策 $Y = y_k$.

3. 共轭先验: 若 $\pi(\theta)$ 与 $\pi(\theta|x)$ 同属一个分布族, 则称该分布族为 θ 的共轭先验分布族, 此时样本的作用仅是将分布族中的未知参数作调整.

1.6.5 区间估计

1. 思想: 区别于点估计, 区间估计的目标是给出 θ 可能的所在区间 $[\hat{\theta}_1, \hat{\theta}_2]$, 使 θ 有 $1 - \alpha$ 的概率落入该区间. 通常使用枢轴量法, 即构造合适的统计量, 利用抽样分布的分位数划定置信限.

2. 单侧区间估计: 若给出 $\hat{\theta}$, 使得 $P(\theta \geq \hat{\theta}) \geq 1 - \alpha$, 则 $\hat{\theta}$ 称为单侧置信下限; 若给出 $\hat{\theta}$, 使得 $P(\theta \leq \hat{\theta}) \geq 1 - \alpha$, 则 $\hat{\theta}$ 称为单侧置信上限. 求解方法与双侧区间估计类似.

3. 单正态分布总体区间估计: 设 $X \sim N(\mu, \sigma^2)$. 主要依据为 1.5.4 节给出的统计量.

(1) σ^2 已知, 对 μ 区间估计: 构造统计量

$$u = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

则通过 $|u| \leq u_{1-\frac{\alpha}{2}}$ 反解出置信度为 $1 - \alpha$ 时 μ 的置信区间.

(2) σ^2 未知, 对 μ 区间估计: 构造统计量

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

则通过 $|t| \leq t_{1-\frac{\alpha}{2}}(n-1)$ 反解出置信度为 $1 - \alpha$ 时 μ 的置信区间.

(3) 对 σ^2 区间估计: 构造统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

则通过 $\chi^2 \in [\chi_{\frac{\alpha}{2}}^2(n-1), \chi_{1-\frac{\alpha}{2}}^2(n-1)]$ 反解置信度为 $1 - \alpha$ 时 σ^2 的置信区间.

4. 双独立正态分布总体区间估计: $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$. X 有 m

个样本, 且 Y 有 n 个样本.

(1) σ_1^2 与 σ_2^2 已知, 对 $\mu_1 - \mu_2$ 估计: 构造统计量

$$\frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1)$$

(2) $\sigma_1^2 = \sigma_2^2$ 未知, 对 $\mu_1 - \mu_2$ 估计: 构造统计量

$$\sqrt{\frac{m+n-2}{\frac{1}{m} + \frac{1}{n}}} \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{(m-1)s_X^2 + (n-1)s_Y^2}} \sim t(m+n-2)$$

(3) 对方差比 $\frac{\sigma_1^2}{\sigma_2^2}$ 估计: 构造统计量

$$\frac{s_X^2/\sigma_1^2}{s_Y^2/\sigma_2^2} \sim F(m-1, n-1)$$

1.7 假设检验

统计学中会有很多假设, 最常见的是假设总体服从正态分布. 但这些假设是不是准确呢? 对这些假设作检验的本质就是反证法: 如果假设是准确的, 则选定统计量应该服从某分布, 但统计量结果在该分布中出现可能性很小, 就推翻假设的正确性.

最常见的假设检验即检验分布的参数是否为某一定值, 如灯泡的寿命是否维持原状 (指数分布的参数是否为某一定值), 是该部分的主要内容. 也有其它的检验目标, 如总体分布的假设是否合理, 样本是否为简单随机样本等, 也将作一定介绍.

1.7.1 基本思想

1. 假设检验的基本问题: 选定原假设 H_0 与备择假设 H_1 , 若 H_0 发生的可能性非常小, 则拒绝原假设而接受备择假设; 若不能拒绝原假设, 则接受原假设. 假设检验的一般问题记作

$$H_0 : \text{_____} \text{ vs } H_1 : \text{_____}$$

2. 检验方法: 先假设 H_0 成立, 结合某一检验统计量的分布给出拒绝域. 若该统计量落入拒绝域, 则拒绝 H_0 , 否则接受 H_0 .

3. 两类错误: 若 H_0 为真, 但统计量落入拒绝域, 则犯了第一类错误 α ; 若 H_0 为假, 但接受了 H_0 , 则犯第二类错误 β .

4. 显著性水平 α : 控制犯第一类错误的可能性 $\leq \alpha$, 即假设 H_0 为真, 检验统计量落入拒绝域的概率应小于等于 α .

1.7.2 正态总体假设检验

1. 单正态总体假设检验: $X \sim N(\mu, \sigma^2)$, 样本量为 n , 构造的统计量依然如 1.5.4 节所述. 由于等式假设和不等式假设仅涉及双侧置信区间和单侧置信区间, 处理手法类似, 故仅以等式假设为例.

(1) σ^2 已知, 检验 $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$: 构造统计量

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

当 H_0 为真时, $u \sim N(0, 1)$, 即拒绝域为 $\{|u| \geq u_{1-\frac{\alpha}{2}}\}$.

(2) σ^2 未知, 检验 $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$: 构造统计量

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

当 H_0 为真时, $t \sim N(0, 1)$, 即拒绝域为 $\{|t| \geq t_{1-\frac{\alpha}{2}}\}$.

(3) 检验 $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$: 构造统计量

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

当 H_0 为真时, $\chi^2 \sim \chi^2(n-1)$, 即拒绝域为 $\{\chi^2 \leq \chi_{\frac{\alpha}{2}}^2(n-1), \text{ or } \chi^2 \geq \chi_{1-\frac{\alpha}{2}}^2(n-1)\}$.

2. 双正态总体假设检验: 设 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 且 X 的样本数为 m , Y 的样本数为 n .

(1) σ_1^2 和 σ_2^2 已知, 检验 $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$: 取检验统计量

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1)$$

(2) $\sigma_1^2 = \sigma_2^2$ 未知, 检验 $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$: 取检验统计量

$$t = \frac{m+n-2}{\sqrt{\frac{1}{m} + \frac{1}{n}}} \frac{\bar{x} - \bar{y}}{(m-1)s_X^2 + (n-1)s_Y^2} \sim t(m+n-2)$$

(3) $m = n$ 且方差未知的成对样本检验, 检验 $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$: 取检验统计量

$$t = \frac{\bar{x} - \bar{y}}{(s_X^2 + s_Y^2)/\sqrt{2n}} \sim t(2n-2)$$

(4) 方差比检验, 检验 $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_1^2 \neq \sigma_2^2$: 取检验统计量

$$F = \frac{s_X^2}{s_Y^2} \sim F(m-1, n-1)$$

1.7.3 广义似然比检验

1. 思想: 检验 $H_0: \theta \in \Theta$ vs $H_1: \theta \notin \Theta$ 时, 取统计量

$$\Lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta} f(x_1, \dots, x_n; \theta)}{\sup_{\theta \notin \Theta} f(x_1, \dots, x_n; \theta)}$$

Λ 越大, 说明 H_0 越有可能成立.

2. 拒绝域. 尚未有统一的形式. 但可以用渐近分布 $2\Lambda \sim \chi^2(n)$, 其中 n 为独立参数的个数.

1.7.4 拟合优度检验

1. 分布拟合检验: 设总体被分为 r 个类 A_1, \dots, A_r , A_i 类中有 n_i 个样本, 检验原假设 $H_0: A_i$ 所占比率为 p_i , 其中 $\sum_{i=1}^r p_i = 1, \sum_{i=1}^r n_i = n$. 构造统计量

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(r-1)$$

拒绝域为 $\{\chi^2 \geq \chi_{1-\alpha}^2(r-1)\}$.

2. χ^2 拟合优度检验: 若 X 的分布函数为 $F(x)$, 将样本归为 r 类: $(-\infty, a_1], (a_1, a_2], \dots, (a_{r-1}, +\infty)$, 每一类理论占比 $p_i = F(a_i) - F(a_{i-1})$, 实际有 n_i 个样本落入第 i 类, 在此基础上应用分布拟合检验.

1.7.5 正态性检验

1. 目的: 检验总体是否服从正态分布.

2. 概率图纸法: 令样本从小到大排列为 $x_{(1)}, \dots, x_{(n)}$, 将点 $(x_{(i)}, \frac{i-0.375}{n+0.25})$ 描在图纸上, 若近似成一条直线, 则认为总体 X 服从正态分布.

1.7.6 游程检验

1. 目的: 检验样本是否随机抽取.
2. 游程检验: 设样本中位数为 M_e , 将样本按抽样时间顺序排列, 并将 $\geq M_e$ 的值替换为 1, $< M_e$ 的值替换为 0, 得到一串 0-1 序列.
3. 判断依据: 把以 0 为界的连续 1 串称为 1 游程, 以 1 为界的连续 0 串称为 0 游程. 若 0 游程数和 1 游程数之和过大或过小, 则拒绝采样随机性, 拒绝域通过查表得出.

1.8 方差分析

单因素方差分析用于解决这一类问题: 控制其它因素都一样, 就改变一个因素, 会不会造成很显著的影响? 换用统计语言来说, 设一个因素有不同的各个水平, 这些水平的均值是否相等? 若相等, 则因素 A 对实验结果没啥影响; 若不相等, 则因素 A 的影响显著. 方差分析作出一个假设: 各水平服从方差相等的正态分布.

样本的波动可由两部分构成: 一是随机性导致同一水平内的数据波动, 即组内误差; 二是因素 A 的作用使不同水平的样本发生了质变, 即组间误差. 显然组间误差占比越高, 因素 A 的影响越显著, 方差分析表也基于此思想得出.

1.8.1 基本思想

1. 检验问题: 设因素 A 有 r 个水平, 各水平均为正态总体 $N(\mu_i, \sigma^2)$ 且方差相等, 检验因素 A 对均值的影响是否显著, 即检验

$$H_0: \mu_0 = \mu_1 = \cdots = \mu_r \quad \text{vs} \quad H_1: \mu_0, \mu_1, \cdots, \mu_r \text{ 不全相等}$$

2. 统计模型: 令 y_{ij} 表第 i 个总体的第 j 次试验结果, m_i 为水平 A_i 的样本数, 总样本数 $n = \sum_{i=1}^r m_i$. 记 $\varepsilon_{ij} = y_{ij} - \mu_i$ 为随机误差, 则 ε_{ij} 相互独立, 且 $\varepsilon_{ij} \sim N(0, \sigma^2)$.

3. 组内偏差: 令 \bar{y}_i 表示第 i 个总体的组内均值, 则用 $S_e = \sum_{i=1}^r \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$ 表示第 i 个总体的组内偏差.

4. 组间偏差: 令 \bar{y} 为所有样本的均值, 则用 $S_A = \sum_{i=1}^r m_i (\bar{y}_i - \bar{y})^2$ 表示因素 A 导致的组间偏差.

5. 平方和分解公式: 令总偏差为 $S_T = \sum_{i=1}^r \sum_{j=1}^{m_i} (y_{ij} - \bar{y})^2$, 由 $y_{ij} - \bar{y} = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y})$ 可推导如下重要公式:

$$S_T = S_A + S_e$$

1.8.2 单因素方差分析

1. 基本思想: 由平方和分解公式, 若 S_A 远大于 S_e , 即偏差大部分由因素 A 导致, 则认为因素 A 影响显著. 构造检验统计量

$$F = \frac{S_A/(r-1)}{S_e/(n-r)} \sim F(r-1, n-r)$$

当 F 大于临界值 $F_{1-\alpha}(r-1, n-r)$ 时, 拒绝原假设, 认为因素 A 显著.

2. 单因素方差分析表

	平方和	自由度	均方	F 比	临界值
因素 A	S_A	$r - 1$	$S_A/(r - 1)$	$F = \frac{S_A/(r-1)}{S_e/(n-r)}$	$F_{1-\alpha}(r - 1, n - r)$
误差 e	S_e	$n - r$	$S_e/(n - r)$		
总和	S_T	$n - 1$			

3. 填表方法: 先计算 $S_T = \sum_{i=1}^r \sum_{j=1}^{m_i} (y_{ij} - \bar{y})^2$, 再计算 $S_e = \sum_{i=1}^r \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$, 由平方和分解公式计算 $S_A = S_T - S_e$, 再从左到右填写剩下内容. 最后比较 F 比和临界值, 若 F 比大于临界值, 则拒绝 H_0 , 认为因素 A 显著.

1.8.3 方差齐性检验

1. 目的: 检验 r 个总体是否符合方差相等的假设条件.
2. 哈特利检验: 令 $H = \frac{\max\{s_1^2, \dots, s_r^2\}}{\min\{s_1^2, \dots, s_r^2\}}$, 则 H 越接近 1, 越有可能认为方差相等. 查表得 H 分布的分位数和拒绝域.

1.9 回归分析

现实生活中, 很难有自变量和因变量的关系能和数学分析中研究的函数一样, 具有良好的性质. 但是我们可以用性质好的函数去拟合变量间的相关关系, 并综合运用前述统计方法, 评价这种拟合到底合不合理, 这就是回归分析的最基本思想.

对单变量关系的情形, 若将样本点 (x, y) 描在图纸上, 仅有散点图很像一条直线时, 我们才能猜测变量间存在线性关系, 其它形状的散点图都不能直接得出结论. 因此一元线性回归是回归分析中最重要的一环, 即用线性函数 $y = \beta_0 + \beta_1 x$ 拟合 x 和 y 间的相关关系.

1.9.1 基本思想

1. 目的: 令 x 为自变量, y 为因变量, 用函数关系 $y = f(x)$ 拟合 x 与 y 间的相关关系, 并要求误差尽可能小.

2 一元线性回归: 用 $y = \beta_0 + \beta_1 x$ 拟合相关关系, 统计模型为

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

1.9.2 回归系数的最小二乘估计

1. 目的: 令 n 组样本对为 (x_i, y_i) , 求 $\hat{\beta}_0, \hat{\beta}_1$, 使误差和 $Q = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$ 最小.

2. 求解: 由 $\frac{\partial Q}{\partial \beta_0} = 0$, 得 $2 \sum_{i=1}^n (\beta_0 + \beta_1 x_i - y_i) = 0$.

由 $\frac{\partial Q}{\partial \beta_1} = 0$, 得 $2 \sum_{i=1}^n x_i (\beta_0 + \beta_1 x_i - y_i) = 0$. 联立解得

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

3. 估计的性质: $\hat{\beta}_0, \hat{\beta}_1$ 均为无偏估计.

1.9.3 区间估计与预测

1. 区间估计目的: 给定 x_0 , 求 $E(y_0) = \beta_0 + \beta_1 x_0$ 的区间估计.

2. 区间估计方法: 令 $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$, 构造统计量

$$\frac{\hat{y}_0 - (\beta_0 + \beta_1 x_0)}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{l_{xx}}}} \sim t(n-2)$$

其中 $l_{xx} = \sum_{i=1}^n x_i^2 - \frac{1}{n}(\sum_{i=1}^n x_i)^2$.

3. 预测: 给定 x_0 的条件下, 求 y_0 的区间估计. 构造统计量

$$\frac{\hat{y}_0 - y_0}{\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{l_{xx}}}} \sim t(n-2)$$

从中反解出 y_0 所处的区间.

1.9.4 显著性检验

1. 目的: 检验 y 和 x 的相关性是否显著. 若 y 与 x 无关, 则 $\beta_1 = 0$, 即检验假设 $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$.

2. t 检验: 取检验统计量

$$t = \frac{\hat{\beta}_1}{\hat{\sigma} / \sqrt{l_{xx}}} \sim t(n-2)$$

3. F 检验: 取 $S_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, $S_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, $S_T = \sum_{i=1}^n (y_i - \bar{y})^2$, 则有平方和分解公式 $S_T = S_R + S_e$. 列方差分析表:

	平方和	自由度	均方	F 比	临界值
回归 R	S_R	1	S_R	$F = \frac{S_R}{S_e/(n-2)}$	$F_{1-\alpha}(1, n-2)$
误差 e	S_e	$n-2$	$S_e/(n-2)$		
总和	S_T	$n-1$			

与方差分析流程一致. 实际上, F 检验与 t 检验等价.

1.9.5 多元线性回归

1. 目的: 用 $y = \omega^T x + b$ 拟合向量 y 与 x 之间的相关关系.

2. 求解: 最小二乘法. 解为 $\theta = (X^T X)^{-1} X^T Y$, 其中 $(x_1, y_1), \dots, (x_m, y_m)$ 为样本点,

$$\theta = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \\ b \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} & 1 \\ x_{21} & x_{22} & \cdots & x_{2n} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

1.9.6 非线性回归

1. 确定函数形式: 根据散点图形状, 确定变换 $z = \varphi(y)$.
2. 作变换: 重新绘制 (x, z) 的散点图.
3. 线性回归: 若 z 与 x 的散点图近似直线, 则对 x 和 z 作一元线性回归.