

《概率论与数理统计》勘误表（第4次印刷）

第一章 事件及其概率

1. P1分赌本问题, “这里不妨简单地假设: 规则为五局三胜制且没有平局.” 应为“这里不妨简单地假设: 两人水平相同, 规则为五局三胜制且没有平局, 各局独立”; “但是赌局因故没有赌满 s 局, 其中赌徒 A 和 B 分别赢了 a 局和 b 局 ($a + b < 2s - 1$),” 应为“但是赌局因故停止且尚无人赢满 s 局, 其中赌徒 A 和 B 分别赢了 a 局和 b 局 ($0 \leq a, b < s, a + b < 2s - 1$),”
2. P4例1.1, “掷出三个正面所需掷的次数”的样本空间 应为 $\{3, 4, 5, \dots\}$.
3. P4例1.2, “ $\omega = \{t = 500\}$ ” 应为 $\{t = 500\}$
4. P12第2行, 应为记 x_k 为第 k 个人的生日, $k = 1, 2, \dots, r$, 则样本空间

$$\Omega = \{(x_1, x_2, \dots, x_r) : x_k = 1, 2, \dots, 365, 1 \leq k \leq r\},$$

由乘法原理 $|\Omega| = 365^r$

5. P12例1.7后的注, “长期赌的后果是此人一定输” 应为“在公平、各局独立且每局以固定单位输赢、直到一方破产的模型中, 此人最终破产的概率为 $b/(a+b)$ ”. “计算可以参见本章习题”删除.
6. P14例1.10解答, “样本空间...9” 应为记 r_i 表示第 i 个患者就诊的时间, 则样本空间
$$\Omega = \{(r_1, r_2, \dots, r_9) : r_i = 2, \dots, 7, i = 1, 2, \dots, 9\}$$
7. P16概率公理化定义脚注, “即 A 为 Ω 所生成的 σ 域 $\mathcal{F}$ 中的元素.” 应为“即 A 为 Ω 上的 σ 域 $\mathcal{F}$ 中的元素.”.
8. P29, “非空的两个互斥事件必然不相互独立” 应为“概率均为正的两个互斥事件必然不相互独立”.
9. P42习题18, “12:30半” 应为“12:30”, “1:00”为“13:00”
10. P42–43习题19, B_n 所表示的事件 应为“ $A_n, A_{n+1}, \dots$ 中至少有一个发生”, 原文的 $A_k, A_{k+1}, \dots$ 指标应改为 n . 提示(2)中 $\ln(1-x) > -x$ 应为 $\ln(1-x) < -x$ ($0 < x < 1$).
11. P44习题27, “从人群中” 应为 “从人群(男女比例为1:1)中”

第二章 随机变量及其分布

1. P48, “对定义2.5中的离散型随机变量 X ” 应为“对定义2.2中的离散型随机变量 X ”.
2. P48, “0-1分布随机变量 X 的分布函数也可以写为” 应为“0-1分布随机变量 X 的概率质量函数也可以写为”.
3. P49例2.3的三处“单字节” 应为“单比特”.
4. P49例2.5, 超几何分布的可取范围 $m = 0, \dots, n$ 应为

$$\max\{0, n - (N - M)\} \leq m \leq \min\{M, n\}.$$

5. P54定理2.1证明, 分母 q^n 应为 q^m , 即

$$\frac{P(X > m+n)}{P(X > m)} = \frac{q^{m+n}}{q^m} = q^n.$$

6. P61例2.16解答后的结论, “离散型随机变量的分布函数为阶梯函数” 应为“离散型随机变量的分布函数是纯跳跃型, 在每个具有正概率的取值点有跳跃, 跳跃高度等于该点概率”
7. P63例2.18, “也就是说我们得到了一个经验分布函数” 应为“得到了一个直方图密度估计”.
8. P68实验7, “进行均匀分布模拟实验” 应为“进行指数分布模拟实验”.
9. P68-69例2.20, “单位时间内失效的概率” 应为“单位时间内失效的条件概率”
10. P72第2.4节开头, 物体动能应写为 $E = \frac{1}{2}mV^2$, 其中 $m > 0$ 为物体质量.
11. P74逐段单调变换公式的注, 原末项 I_j 应为 $I_{g(I_j)}(y)$, 即仅在第 j 段的像集上使用该段反函数:

$$f_1(y) = \sum_j f(h_j(y))|h'_j(y)|I_{g(I_j)}(y).$$

12. P75例2.26, 所列 $I_{0,\infty}(y)$ 应为 $I_{(0,\infty)}(y)$.
13. P77正态近似推导, “具体地, $i = o(n)$ ” 应为 $i = o(n^{3/4})$, “再在二项分布的...”下一组公式首行中的

$$P\left(\left|X - \frac{n}{2}\right| \leq \frac{c}{\sqrt{n}}\right)$$

应为 $P(|X - n/2| \leq c\sqrt{n})$.

14. P79 由 $g(x) = f'(x)/f(x) = cx$ 应得

$$f(x) = M \exp\left(\frac{c}{2}x^2\right), \quad c < 0.$$

原 Me^{cx^2} 漏写 $1/2$;

15. P83习题18(2), 概率中的小写 x 应为随机变量 X , 即 $P(1/2 < X < 3/2)$.
16. P85 习题48中 Y 的分段最后一段改为 $X \geq 2$.

第三章 多维随机变量及其分布

1. P89联合分布函数性质(1), “单调不减” 应为“分别对 x, y 单调不减”; 性质(2) 应完整写为 $0 \leq F(x, y) \leq 1, F(\infty, \infty) = 1, F(-\infty, y) = F(x, -\infty) = 0, F(-\infty, \infty) = F(\infty, -\infty) = 0$.
2. P90例3.1, 分段函数中三个 $\leq$ 均改为 $<$, 即 $F(x, y) = 0$ 当 $x < 0$ 或 $y < 0$ 或 $x + y < 1$, 其他情形为1. 证明中的 $1 - 1 - 1 - 0$ 应为 $1 - 1 - 1 + 0$.
3. P91例3.3, “试求分布函数” 应为“试求联合概率质量函数”; 解答中“总的结果数为 N ” 改为“总试验次数为 N ”.
4. P93例3.4, “只能在对角线取非0值” 改为“取值只能落在对角线 $y = x$ 上”. “不存在非负函数满足性质(2)” 改为“不存在能表示此联合分布的二维概率密度函数”.
5. P93-94例3.7, 分布函数积分式中的 $I_{(u>0, v>0)}$ 应删除; 计算 $P(X + Y \leq 1)$ 的中间式 $\int_0^1 (1 - e^{1-x}) dx$ 应为 $\int_0^1 (e^{-x} - e^{-1}) dx$;

6. P94的 n 重积分同理由内到外**应为** $du_n \cdots du_1$.
7. P98, “则不难验证”**改为**“其中 $|\rho| < 1$. 则不难验证”
8. P103, 离散型独立性必要性的证明中, $F(x_{i-1}, y_i)$ **应为** $F(x_{i-1}, y_j)$.
9. P104例3.17, 标题“两个正态分布随机变量独立的充要条件”**应为**“二元联合正态随机变量独立的充要条件”.
10. P109例3.22, 第二步配平方后平方括号内的中心**应为**

$$x - \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1} \left(\frac{\mu_1}{\sigma_1^2} + \frac{z - \mu_2}{\sigma_2^2} \right).$$
原式第一个括号遗漏指数-1; 最终正态密度正确.
11. P110例3.23末, “独立正态分布随机变量的商服从柯西分布”**应为**“独立零均值同方差正态分布随机变量的商服从柯西分布”
12. P110例3.24, “极轴和极角 (R, Θ) ”**应为**“极径和极角 (R, Θ) ”.
13. P111例3.25(3), V 服从 $Ga(1, 3)$ **应为** $Ga(3, 1)$

第四章 随机变量的数字特征和极限定理

1. P127, 倒数第4行 $E(X_1 + X_2)$ **应为** $E(X + Y)$.
2. P132, “当 $f(m) > 0$ 时, 中位数唯一”**应为**“当 f 在中位数 m 的邻域连续且 $f(m) > 0$ ”.
3. P133例4.14, 中位数集合**应为**闭区间 $[0, 1]$, 原文的 $[0, 1)$ 漏掉了端点1, 因为 $P(X \geq 1) = P(X \leq 0) = 1/2$.
4. P133例4.15, 对数正态期望推导中的 $\exp\{-\frac{1}{2}(t-\sigma)^2\}$ 多出一个右圆括号,**应为** $\exp\{-\frac{1}{2}(t-\sigma)^2\}$.
5. P141-142偏度说明部分, “三阶矩的一个应用就是衡量分布是否有偏. 设 $X \sim f(x)$,”**改为**“三阶矩的一个应用就是衡量分布是否有偏. 设 $X \sim f(x)$, 三阶矩存在,”; “如果 $\mu_3 > 0, f(x)$ 的图形最高点偏左, 称为正偏或右偏; 如果 $\mu_3 < 0, f(x)$ 的图形最高点偏右, 称为负偏或左偏”**应为**: “如果 $\mu_3 > 0$, 称为正偏或右偏, $\mu_3 < 0$ 则称为负偏或左偏.”
6. P146相关系数说明, “不能反映随机变量之间有某种函数关系”**应为**“不能反映它们之间的非线性函数关系”.
7. P135例4.17, $= 0.81N$ 和 $= 0.49N$ 分别**应为** $\approx 0.81N$ 和 $\approx 0.49N$.
8. P138例4.20, 由独立性计算和的方差时引用的“方差性质(3)”**应为**“方差性质(4)”.
9. P145, 4.21式上面, M_1 和 M_2 应分别**改为** μ_1 和 μ_2 , 与后续公式一致.
10. P146-147, “为此要引入一个线性空间 $\mathcal{L} = \{r.v. X : E(X^2) < \infty\}$. 在 $\mathcal{L}$ 上, 对任意 $X, Y \in \mathcal{L}$ 定义内积

$$\langle X, Y \rangle = Cov(X, Y).$$

$\mathcal{L}$ 是内积空间可以用线性代数中内积空间的定义来验证.”**改为**“为此要引入中心化平方可积随机变量构成的线性空间

$$\mathcal{L} = \{X : E(X^2) < \infty, E(X) = 0\},$$

其中随机变量定义在同一概率空间上, 并将几乎必然相等的随机变量视为同一元

素. 在 $\mathcal{L}$ 上, 对任意 $X, Y \in \mathcal{L}$ 定义内积

$$\langle X, Y \rangle = E(XY) = Cov(X, Y).$$

由内积的定义可验证, $\mathcal{L}$ 是内积空间. 对方差有限且非零的一般随机变量 X, Y , 其相关系数就是中心化向量 $X - E(X)$ 与 $Y - E(Y)$ 在该空间中夹角的余弦.”

11. P148–149例4.27, 换元 $t = (u - \rho v)/\sqrt{1 - \rho^2}$ 后的中间积分漏了 $\sigma_1\sigma_2$ 因子, 应为

$$= \sigma_1\sigma_2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} v e^{-\frac{v^2}{2}} dv \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} (t + \rho v) e^{-\frac{t^2}{2}} dt$$

最终 $\rho\sigma_1\sigma_2$ 正确.

12. P149, “

$$S = k \log_2(|\Omega|).$$

k 表示玻尔兹曼常数, $|\Omega|$ 表示微观粒子数. 系统越乱, 熵就越大; 系统越有序, 熵就越小.” 应为

$$S = k \ln |\Omega|.$$

其中, S 为系统的物理熵, k 为玻尔兹曼常数, Ω 为与给定宏观状态相容的微观状态集合, $|\Omega|$ 为这些微观状态的数目. 在微观状态等概率的模型下, 对应的微观状态数越多, 仅凭宏观状态确定系统所处微观状态的不确定性就越大, 系统的物理熵也越大.

13. P151熵的性质(2), “如果取有限个值的随机变量 X 的概率分布 $\{p_k, k = 1, 2, \dots, n\}$ 不是离散均匀分布,” 改为“如果离散型随机变量 X 的概率分布为 $\{p_k, k = 1, 2, \dots, n\}$, ”

14. P153例4.31, c_0 中对数项的符号应为负, 即

$$c_0 = -\frac{m^2}{2\sigma^2} - \ln(\sqrt{2\pi}\sigma) + 1.$$

15. P154例4.32, “因此”后的积分式遗漏“0 =”, 应为 $0 = \int_0^\infty g(x)(-\ln f(x) + c_1x + c_0 - 1)dx$.

16. P162例4.38, 两个正态近似中间值顺序有误. 使用未取整的标准化数值时 应为

$$\Phi((3.5 - 6)/\sqrt{4.2}) - \Phi((0.5 - 6)/\sqrt{4.2}) \approx 0.111256 - 0.003640 = 0.107616,$$

$$\Phi((3 - 6)/\sqrt{4.2}) - \Phi((1 - 6)/\sqrt{4.2}) \approx 0.071617 - 0.007349 = 0.064269.$$

17. P164例4.42, (2)中第二个 $\text{Var}(\bar{S}_1)$ 应为 $\text{Var}(\bar{S}_2)$.

18. P169例4.47后的注释, $P(X > n) = P(X_1 + \dots + X_n \leq 1)$ 右端变量 应为 $U_1, \dots, U_n$, 与本题定义一致.

19. P178习题58, 被积函数的“乘积之比” 应为“和之比”, 并补充 $\int_0^1 g(x)dx > 0$:

$$\lim_{n \rightarrow \infty} \int_0^1 \int_0^1 \dots \int_0^1 \frac{\sum_{i=1}^n f(x_i)}{\sum_{i=1}^n g(x_i)} dx_1 \dots dx_n = \frac{\int_0^1 f(x)dx}{\int_0^1 g(x)dx}.$$

20. P179习题60, “有理数展开前 n 位” 改为“十进制展开前 n 位”.

第五章 统计学基本概念

1. P186注(3), 随机向量 $(X_1, \dots, X_n)$ “常称为样本量”改为“ n 称为样本”.
2. P188, “数量指标的信息、抽样方式、认为假设等”中的“认为假设”改为“人为假设”.
3. P188例5.4解, “简单随机样本的联合分布函数为”改为“联合概率密度函数为”.
4. P191, $nF_n(x) \sim b(n, F(x))$ 中的分布记号 b 应为 B , 即 $B(n, F(x))$.
5. P192例5.6, 最大值的分布函数漏掉 $t \geq \theta$ 时取1的部分, 应为

$$F_T(t) = \begin{cases} 0, & t \leq 0, \\ (t/\theta)^n, & 0 < t < \theta, \\ 1, & t \geq \theta. \end{cases}$$

6. P193–194伽马分布的注, “若记 $Ga(\alpha, \lambda)$ 为 ”改为“若记 $Ga(\alpha, \lambda), \alpha > 0, \lambda > 0$ 为”;
“随机变量 ξ 的概率密度函数为 $Ga(n/2, 1/2)$ ”改为“随机变量 ξ 服从 $Ga(n/2, 1/2)$ 分布”.
7. P194伽马分布与 χ^2 分布的关系, 整句应为: 若 $Y \sim Ga(\alpha, \lambda)$, 则 $2\lambda Y \sim \chi_{2\alpha}^2$
8. P194, t 分布首次发表年份“1907年”改为“1908年”.
9. P194–195实验15, “当自由度 n 多大时其密度函数曲线与标准正态分布密度函数曲线重合”改为“当 n 增大时, 观察两条曲线在给定绘图范围和精度下何时近似重合”.
10. P199推论5.3证明, “再利用 χ^2 分布的性质(3)”改为“再利用 χ^2 分布的性质(2)”.
11. P200关于1936年盖洛普, “分层随机抽样”改为“配额抽样”. 配额抽样不等于各层内随机抽取的分层概率抽样. “《文学摘要》随即破产倒闭”亦不准确, 宜改为“这次失败损害其声誉, 并促成其于1938年停刊”.
12. P204习题4, 所问“抽样分布”应改为“这10个观测对应的样本分布”.
13. P204习题8(1), “样本空间和抽样分布”改为“样本空间和样本分布”

第六章 参数点估计

1. P210注及P221例6.20结尾, “可以证明低阶矩优于高阶矩”改为“可以证明低阶矩一般优于高阶矩”
2. P210例6.7, 删除 “这是一个超几何分布,”
3. P214例6.11, 题设 $\theta > 0$ 改为 $\theta \in \mathbb{R}$
4. P215例6.13解, “该方程的解没有显示表达式”中的“显示”应为“显式”.
5. P215, “矩估计不需要总体分布的知识”应改为“不必知道完整分布形式, 但必须知道所用总体矩与待估参数的关系”.
6. P217例6.15证明, “再注意到样本方差等于 σ^2/n ”改为“再注意到样本均值的方差 $\text{Var}(\bar{X}) = \sigma^2/n$ ”.
7. P220定义6.3, “至少存在一个 θ , 使上式不等号成立”应为“使严格不等式成立”.
8. P221例6.20, 泊松分布的四阶中心矩应为 $E(X - \lambda)^4 = 3\lambda^2 + \lambda$, 不是 $2\lambda^2 + \lambda$. 对 $n \geq 2$,

$$\text{Var}(S^2) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right) = \frac{\lambda}{n} + \frac{2\lambda^2}{n-1}.$$
9. P222–223克拉默–拉奥下界推导, 改为令

$$\mathcal{L}_{2,\theta}(\mathbf{X}) = \{h(\mathbf{X}) : E_\theta h(\mathbf{X}) = 0, E_\theta (h^2(\mathbf{X})) < \infty\}.$$

并将几乎必然相等的随机变量视为同一元素. 对任意的 $Z, Y \in \mathcal{L}_{2,\theta}(\mathbf{X})$, 定义内积为 $\langle Z, Y \rangle = \text{Cov}_\theta(Z, Y)$.

10. P225例6.23, $g(\lambda) = \lambda^{-1}$ 的导数应改为 $g'(\lambda) = -\lambda^{-2}$.
11. P228模拟研究, “注意到方差约为 N/k ”应改为“标准差约为 N/k ”.
12. P230习题2, “试求 p 的矩估计”改为“试求 (p_1, p_2) 的矩估计”.
13. P235习题31, 分母 $\sqrt{2\pi}\sigma^2$ 应改为 $\sqrt{2\pi}\sigma^3$, 并明确 $x \in \mathbb{R}, \sigma > 0$.
14. P236习题42, 几何分布 $P(X = k) = p(1-p)^{k-1}$ 的取值应改为 $k = 1, 2, \dots$, 不是从0开始.
15. P237习题44, 按原密度 $x > \theta$ 改为 $x \geq \theta$
16. P238习题52, 密度记号 $f(x, \theta)$ 改为 $f(x; \mu)$, 并补全为 $f(x; \mu) = e^{-(x-\mu)} I_{[\mu, \infty)}(x)$, $\mu \in \mathbb{R}$.
17. P238习题59之后, 删除“上式说明在一定条件下两者是等价的... 变动很大.”
18. P238习题60, “问 λ 的无偏估计是否为MVUE”改为“设 $n > 2$, 考察 $\tilde{\lambda} = (n-1)/\sum X_i$ 是否为 λ 的MVUE”.
19. P239习题61, “5人一组, 分为20组”改为“5人一组, 分为40组”.

第七章 区间估计

1. P241基本概念, “到19世纪80年代末”改为“到20世纪80年代末”; “学术期刊已经要求论文必须给出置信区间”改为“一些医学期刊开始强调报告置信区间”
2. P241可靠度与精度的例子, “但是可靠性比第一个要高”改为“但是可靠性比第二个要高”.
3. P244枢轴变量法第(2)步, “使得它的分布 F 已知”改为“使得 $S(T, U, \theta)$ 的分布 F 已知”; 第(3)步, A, B 只能与 T, U, a, b 及已知常数有关.
4. P247例7.6, “若 $\mu_2 > \mu_1$, 则认为 A 线比 B 线产量高”应改为“ B 线比 A 线产量高”.
5. P253–254自助法的适用条件, “不对总体的分布做任何假设”及“没有理论(如CLT)”改为“不要求指定总体的参数分布族, 但仍须有合适的抽样模型和自助近似的有效性条件”.
6. P256自助 t 算法第(c)步, 内层均值中的下标 b 应改为内层下标 k , 即 $\bar{\theta}^* = R^{-1} \sum_{k=1}^R \tilde{\theta}_k^*$.
7. P257置信限引入与P258例7.13, AQI是无量纲指数, 应删除其单位 $\mu\text{g}/\text{m}^3$, 并将“颗粒物的上限是否小于50”改为“所定义的AQI总体均值的上限是否小于50”.
8. P259死亡率渐近置信区间, 右端点符号改为加号, 即 $[\hat{p} - u_{0.025} \hat{se}(\hat{p}), \hat{p} + u_{0.025} \hat{se}(\hat{p})]$.
9. P262习题9、10, “按流量收费”与前文“按分钟收费”不一致, 应统一为“按分钟收费”

第八章 假设检验

1. P285成对比较的差值方向, “总体 $Z = Y - X$ ”与“ $Z_i = X_i - Y_i$ ”不一致. 改为统一 $Z_i = Y_i - X_i$, 相应均值为 $E(Y) - E(X)$; 或全部反向定义, 同时反转单侧假设. 例8.13已经采用 $Y_k - X_k$, 应保持一致.
2. P292符号检验例8.18, 例8.13已定义 $Z_k = Y_k - X_k$, 正差表示偏好B, 因此“ p 为偏好A啤酒的比例”应改为“在打分差非零的条件下偏好B的概率”.

3. P292–293发票例8.19, 数值式的分母 $\sqrt{10(1-0.1)}$ 改为 $\sqrt{0.1(1-0.1)}$. “不正规发票已超出15%”改为“样本不正规比例为15%, 高于假设值10%”.
4. P295–296正态均值似然比例8.21, 最后的临界分位数 $t_{\alpha/2}(n-1)$ 改为 $t_{n-1}(\alpha/2)$.
5. P296均匀分布例8.22, “此时样本空间 $\Theta = (0, \theta)$, $\Theta_0 = (0, \theta_0)$ ”改为“参数空间 $\Theta = (0, \infty)$, 原假设参数子集 $\Theta_0 = (0, \theta_0]$ ”.
6. P298–299例8.24, 备择写为“ $H_1 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2$ 不完全相同”改为“ $H_1 : \sigma_1^2, \dots, \sigma_m^2$ 不完全相同”.
7. P300双侧 p 值, $2 - 2\Phi(t_{\text{obs}})$ 改为 $2[1 - \Phi(|t_{\text{obs}}|)]$.
8. P300–301例8.25的模拟, p 值0.0025改为0.003.
9. P301表8.2, $n = 10000$ 时原值 7.69×10^{-24} 改为约 7.62×10^{-24} .

10. P303 FDR定义应改为

$$Q = \frac{V}{\max(R, 1)}, \quad \text{FDR} = E(Q).$$

11. P303 BH过程后的不等式, 改为 $p_{(R)} \leq Rq/m$, 从而 $mp_{(R)}/R \leq q$ (仅当 $R > 0$).
12. P303–304 BH的证明, “由构建过程可以看出, 此过程恰好拒绝 R 个原假设, 并且

$$p_{(R)} < \frac{R\alpha}{m} \Rightarrow \frac{mp_{(R)}}{R} < q.$$

记 $m_0 \leq m$ 为真实的原假设个数, $X_i = I(p_i \leq p_{(R)})$ 表示第 i 原假设是否被拒绝. 由于在原假设下 $p_i \sim U(0, 1)$, $X_i \sim B(1, p_{(R)})$. 在各假设之间相互独立的假设下,

$$V = \sum_{i=1}^{m_0} X_i \sim B(m_0, p_{(R)}). \text{ 因此由定义有}$$

$$\text{FDR} = \frac{1}{R}E(V) = \frac{m_0 p_{(R)}}{R} \leq \frac{mp_{(R)}}{R} < q.$$

改为 由构建过程可知, 此过程恰好拒绝 R 个原假设, 且 $p_{(R)} \leq \frac{Rq}{m}$, 第 i 个原假设被拒绝当且仅当 $p_i \leq \frac{Rq}{m}$. 记 $\mathcal{H}_0$ 为真实原假设的指标集, $|\mathcal{H}_0| = m_0 \leq m$, 则错误拒绝个数 $V = \sum_{i \in \mathcal{H}_0} I(p_i \leq \frac{Rq}{m})$. 令 $R^{(i)}$ 为将 p_i 换成 0 后的拒绝个数, 它只依赖于其余 p 值, 且易验证 $\{p_i \leq \frac{Rq}{m}\} = \{p_i \leq \frac{R^{(i)}q}{m}\}$, 在此事件上 $R = R^{(i)}$. 在各假设相互独立的假设下, 当 $i \in \mathcal{H}_0$ 时 $p_i \sim U(0, 1)$ 与 $R^{(i)}$ 独立, 因此

$$\text{FDR} = E\left[\frac{V}{R \vee 1}\right] = \sum_{i \in \mathcal{H}_0} E\left[\frac{I(p_i \leq R^{(i)}q/m)}{R^{(i)}}\right] = \sum_{i \in \mathcal{H}_0} E\left[\frac{R^{(i)}q/m}{R^{(i)}}\right] = \frac{m_0}{m}q \leq q.$$

13. P304“BH仅在... 相互独立时候能够控制FDR”应改为“独立是本节所用的充分条件, 适当的PRDS正依赖条件下也可控制”.
14. P305总结第4点, “完全可能在低水平下拒绝原假设, 而在高水平下接受原假设”方向颠倒. 应改为“完全可能在低水平下不能拒绝原假设, 而在高水平下拒绝原假设”
15. P306习题5, 样本量 n 应改为10.
16. P307习题10, “正态分别”改为“正态分布”.
17. P307习题14, “样本方差 $\sigma^2 = 64$ ”改为“样本方差 $S^2 = 64$ ”.
18. P308习题23, “药物治疗治疗”删去重复的“治疗”

第九章 非参数假设检验

1. P316式(9.3)之前的差异说明,“最后一行的值越小”改为“最后一行各项的绝对值越小”.
2. P318例9.3,“这些这批数据”改为“这批数据”,“符合地太好”改为“符合得太好”.
3. P319–320例9.4,杂合子Aa的理论频数 $100 \times 2 \times \hat{p}^2(1 - \hat{p})$ 改为 $100 \times 2\hat{p}(1 - \hat{p})$.
4. P324例9.7,“解”环境在期望频数表后过早结束,应移到统计量计算和结论之后.
5. P325齐一性检验的注,“另一个属B”改为“另一个属性B”.“即视属性A为随机的”改为“在各行样本量固定、各组独立多项抽样的模型下,可得到同样形式的统计量及渐近自由度”.
6. P329符号检验引入,“数据差不是正态...无法用t检验”应改为“小样本精确成对t检验要求差值正态;在独立同分布且有限正方差等条件下可另用大样本近似”.
7. P330啤酒符号模型,符号0的“都很好”改为“无偏好或评价相同”
8. P330双侧符号p值,“p值定义如下”的公式 改为

$$p = \min\{1, 2P(X \leq x')\}, \quad X \sim B(m, 1/2).$$

9. P330“没有任何分布的假定”改为“不要求差值为正态,但仍须独立性及零假设下非零正负号等概率等条件”.
10. P332单侧两分布假设, $H_0 : F_1(x) \leq F_2(x) \leftrightarrow H_1 : F_1(x) > F_2(x), -\infty < x < \infty$.
改为 $H_0 : F_1(x) \leq F_2(x), -\infty < x < \infty \leftrightarrow H_1 : \exists x, F_1(x) > F_2(x)$.

第十章 相关分析和回归分析

1. P341首段,“analysis of varince”应为“analysis of variance”;学习目标中“误差为不相关的正态随机变量”应为“误差为相互独立、同分布的正态随机变量”.
2. P342数据分类,有序数据应单列为一类,即第2类,第3类为定量数据;“两个相邻间隔数据之差总是相同的”应为“相同数值差表示相同的量的差别”.
3. P343注(4),“X, Y夹角的余弦”应为“ $X - E(X), Y - E(Y)$ 夹角的余弦”
4. P343图10.1的说明中,“抛物型曲线关系”应为“余弦函数关系”;“拟合直线与x轴夹角的余弦”应为“两个中心化样本向量夹角的余弦”.
5. P345例10.1解答, $r = 0.9468$ 应为 $r = 0.9486$;删除“即得到图10.2中虚线的斜率”.
6. P345例10.1解答(2),“其余的10%要用其他的变量变化来解释”应为“其余约10%的样本变异未被该线性拟合解释”.
7. P346例10.2表头的 $\ln x, \ln y$ 应为 $\lg x, \lg y$;本例正文相应改为 $t = \lg x, u = \lg y, u = \lg c + bt$. 末尾的 $y = cx^d$ 应为 $y = cx^b$.
8. P347所谓“修正的肯德尔相关系数”,将分母改为同向对与异向对总数后所得的是Goodman–Kruskal的 γ 系数. 设同向对、异向对数为C, D, 则 $\gamma = (C - D)/(C + D)$;如指Kendall的 τ_b , 应为

$$\tau_b = \frac{C - D}{\sqrt{(C + D + T_x)(C + D + T_y)}},$$

其中 T_x, T_y 分别为仅在 x 、仅在 y 上并列的对数,同时在两者上并列的对不计入 T_x, T_y .

9. P347例10.3表头“不及格率序 $y_i/\%$ ”应为“不及格率序 y_i ”;P348解答中两处对条件取sgn应

改用指示函数, 即

$$\sum_{i < j} I\{(x_i - x_j)(y_i - y_j) > 0\} = 16, \quad \sum_{i < j} I\{(x_i - x_j)(y_i - y_j) < 0\} = 39.$$

10. P350公式(10.11)及P352公式(10.18), 函数 g 的定义中不应含 $\min$, 应为

$$g(a, \beta_1) = \sum_{i=1}^n (y_i - a - \beta_1 x_i)^2 = \|\mathbf{y} - a\mathbf{1} - \beta_1 \mathbf{x}\|^2,$$

再对 a, β_1 求最小值.

11. P350例10.4中截距公式 $\hat{a} = \bar{y} - \hat{\beta}_1 \bar{x}$, 应为 $\hat{a} = \bar{y} - \hat{\beta}_1 \bar{x}$; P356定理10.1证明后的截距公式也应为

$$\hat{a} = \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}.$$

12. P351身高单位换算, “68.25英寸=174厘米”应为“68.25英寸=173.355厘米”. “取父母平均值, 所以方差为子代方差的一半”需补充父母身高等方差且协方差为零等条件; 一般有

$$\text{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{\text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_2)}{4}.$$

13. P352定义 $\hat{\boldsymbol{\eta}}$ 的公式, 二项列向量外多余的转置符号应去掉, 即 $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \hat{\eta}_2)^T$; P353常数方向投影公式末项 $\hat{\beta}_0$ 应为 $\hat{\beta}_0 \mathbf{1}$, 其中 $\hat{\beta}_0 = \bar{y}$.

14. P355正态假定, “不相关且 $e_i \sim N(0, \sigma^2)$ ”应为“相互独立且 $e_i \sim N(0, \sigma^2)$ ”

15. P357例10.5之后, 删除“因为 $\text{Var}(t_n) = n/(n-2)$, 由中心极限定理”, 改为“在原假设下统计量服从自由度为10的 t 分布, 直接计算其双侧尾概率”. “可以作 $\hat{\beta}_1$ 的区间估计”应为“可以作 β_1 的区间估计”.

16. P358公式(10.38)和(10.41), 区间中心 $(a + \beta_1 x)$ 均应为 $(\hat{a} + \hat{\beta}_1 x)$; (10.41)及P359例10.6中针对某个新家庭的区间应称为“预测区间”, 而非均值参数的置信区间.

17. P358公式(10.40)后, 新观测 y 的正态分布写错, 应为

$$y \sim N(a + \beta_1 x, \sigma^2), \quad y - \hat{y} \sim N\left(0, \left[1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right] \sigma^2\right).$$

18. P359–360例10.7, 变量变换 $t = \ln x, u = \ln y$ 应为 $t = \lg x, u = \lg y$, 才与所列数值一致; 区间 $[-0.7849, -0.8677]$ 的端点次序应改为 $[-0.8677, -0.7849]$.

19. P360例10.8数据表及图10.9, t 表示照射次数, 表头及横轴的“min”应删除或改标“照射次数”; y 表示细菌数, 表头及纵轴的“kV”应删除. 每次照射为6分钟, 累计时间为 $6t$ 分钟.

20. P360–361例10.8, 用原始15组数据重新作对数最小二乘拟合, 应为

$$\ln \hat{y} = 5.966499 - 0.217899t \quad (\text{指数模型}),$$

$$\ln \hat{y} = 6.408061 - 1.174631 \ln t \quad (\text{幂函数模型}).$$

将未舍入拟合值反变换到原尺度后, 有

$$Q_1 \approx 4059.30, \quad R_1^2 \approx 0.968316,$$

$$Q_2 \approx 75046.20, \quad R_2^2 \approx 0.414240.$$

原列4015, 77746及相应统计量应更正; 公式(10.42)分子中的 $\hat{y}$ 应为 $\hat{y}_i$.

21. P361注中, 删除因变量的别名“外生变量”及自变量的别名“内生变量”.

22. P365定理10.2(6), 单个系数的分布 $\hat{\beta}_i \sim N_p(\beta_i, d(i)\sigma^2)$ 应为 $\hat{\beta}_i \sim N(\beta_i, d(i)\sigma^2)$.

23. P365–366例10.10, 依例10.9的原始数据, 应为

$$\begin{aligned}\hat{m}(47, 18) &= 2.893951, & \hat{\sigma} &= 0.525439, \\ h &= \frac{1}{10} + (\mathbf{x} - \bar{\mathbf{x}})^T S_{xx}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = 0.564120, \\ \hat{m}(47, 18) &\sim N(m(47, 18), 0.564120\sigma^2).\end{aligned}$$

均值 $m(47, 18)$ 的90%置信区间应为

$$[2.893951 \pm 0.525439 \times 1.894579\sqrt{0.564120}] = [2.146262, 3.641640].$$

原列5.81, 1.4837, 2.01426及[1.16, 10.46]与题给数据不符; $t_{10-3} \times 0.05$ 应为 $t_7(0.05)$.

24. P366关于增加自变量的讨论, 应为: 同一组数据的嵌套最小二乘模型中, 增加自变量不会增大训练残差平方和, 但不保证减小参数估计方差或提高样本外预测精度. 完全共线时 $X^T X$ 不可逆; 高度相关但未完全共线时可能病态, 不等于必然不可逆.

25. P367第一段, 由样本 $R^2 = 0$ 不能推出因变量与自变量独立; 应为“ $R^2 = 0$ 表示该样本中自变量的线性拟合未解释中心化因变量的变异”.

26. P367 “如果原假设 H_0 成立, 那么回归平方和为0”应为“在 H_0 成立时, 回归平方和仍有抽样波动, 且 $RSS/\sigma^2 \sim \chi_p^2$ ”.

27. P367–368变量选择准则, 应为

(1) RMS_k 准则 (残差均方准则). 设所选模型含 k 个自变量, 记 $q = k + 1$ 为参数个数 (含截距项), 令

$$RMS_k = \frac{ESS_k}{n - q},$$

选择 k 及相应的变量组合, 使 RMS_k 最小. 这等价于使修正决定系数 $\bar{R}^2 = 1 - \frac{(n-1)ESS_k}{(n-q)TSS}$ 最大.

(2) C_p 准则 (马洛斯 (C. L. Mallows) 于1964年提出). 从预测的观点出发, 令

$$C_p = \frac{ESS_k}{\hat{\sigma}^2} - (n - 2q), \quad \hat{\sigma}^2 = \frac{ESS_p}{n - p - 1},$$

其中 $\hat{\sigma}^2$ 为全模型 (含全部 p 个自变量) 的残差均方. 选择 C_p 小且接近 q 的模型.

28. P369回归诊断第一段, “那么说明回归是有效的”应为“则未发现明显的均值形式或方差方面的问题, 但不能据此证明模型假定成立”.

29. P370“双盲”的定义, “都不知道实验内容和目的”应为“受试者及相关研究人员不知道各受试者接受的是试验处理还是对照处理”.

30. P371附录(2)(b), “即 A 的维数”应为“即 A 的秩”; P372附录(4)证明中“记 $p \times 1$ 列向量”应为“记 $k \times 1$ 列向量”.

31. P372附录(5)谱分解证明, $Q \text{diag}(\lambda_1, \dots, \lambda_n)Q$ 的末项应为 Q^T ; 附录(6)残差平方和展

开中的各 e_i 应为完整误差向量 $\mathbf{e}$, 即

$$\hat{\mathbf{e}}^T \hat{\mathbf{e}} = \sum_{i=1}^{n-p-1} \mathbf{e}^T \mathbf{p}_i \mathbf{p}_i^T \mathbf{e} = \sum_{i=1}^{n-p-1} (\mathbf{p}_i^T \mathbf{e})^2.$$

32. P374附录(8)末尾, 截距关系中的三个加号应改为减号, 方差式末项漏乘 σ^2 , 应为

$$\begin{aligned} a &= \beta_0 - \bar{\mathbf{x}}^T \beta, & \hat{a} &= \hat{\beta}_0 - \bar{\mathbf{x}}^T \hat{\beta}, \\ E(\hat{a}) &= \beta_0 - \bar{\mathbf{x}}^T \beta = a, \\ \text{Var}(\hat{a}) &= \left(\frac{1}{n} + \bar{\mathbf{x}}^T S_{xx}^{-1} \bar{\mathbf{x}} \right) \sigma^2. \end{aligned}$$

33. P374第一式中的 $\hat{\mathbf{x}}$ 应为 $\bar{\mathbf{x}}$; 本页均值置信区间根号内多余的 σ^2 应删除, 即

$$\hat{m}(\mathbf{x}) \pm \hat{\sigma} t_{n-p-1}(\alpha/2) \sqrt{\frac{1}{n} + (\mathbf{x} - \bar{\mathbf{x}})^T S_{xx}^{-1} (\mathbf{x} - \bar{\mathbf{x}})}.$$

34. P377习题8更正为

8. 根据数据 $(x_i, y_i), i = 1, 2, \dots, n$. 我们可以作 y 关于 x 的线性回归 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, 也可以计算 X, Y 之间的皮尔逊相关系数 r . 记 $Q_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, 试:
- (1) 用几何观点证明 $r = \sqrt{\frac{S_{xx}}{S_{yy}}} \hat{\beta}_1$, $\frac{r}{\sqrt{1-r^2}} = \sqrt{\frac{S_{xx}}{Q_e}} \hat{\beta}_1$;
 - (2) 证明在正态假定下, 关于总体相关系数 ρ 的检验问题 $H_0: \rho = 0 \leftrightarrow H_1: \rho \neq 0$ 中的检验统计量 $t = \frac{\sqrt{n-2} r}{\sqrt{1-r^2}}$ 在原假设成立时服从 t_{n-2} 分布.