2025 应用统计模拟试题（回忆版）与参考答案

1 第一题：描述性统计分析

题目描述：为研究美国科技创新公司的市值分布情况，随机抽取了 10 家公司作为样本。其市值（单位：亿美元）数据如下：`0.2, 0.4, 0.8, 1.2, 1.4, 1.8, 2.2, 2.4, 3.0, 3.1`

问题：

1. 计算该样本数据的四分位数，并根据结果绘制箱形图。
2. 基于箱形图，请描述这组数据的分布特征。
3. 计算样本均值，并与中位数进行比较，解释两者之间产生差异的原因。

解答：

1. 四分位数计算与箱形图

首先，我们将数据从小到大排序（数据已排序）：`0.2, 0.4, 0.8, 1.2, 1.4, 1.8, 2.2, 2.4, 3.0, 3.1`

- **中位数 (Q2):** 由于有 10 个数据点（偶数），中位数是中间两个值的平均值。

$$Q2 = \frac{1.4 + 1.8}{2} = 1.6$$

- **下四分位数 (Q1):** Q1 是数据前半部分的中位数（0.2, 0.4, 0.8, 1.2, 1.4）。

$$Q1 = 0.8$$

- **上四分位数 (Q3):** Q3 是数据后半部分的中位数（1.8, 2.2, 2.4, 3.0, 3.1）。

$$Q3 = 2.4$$

- **四分位距 (IQR):**

$$IQR = Q3 - Q1 = 2.4 - 0.8 = 1.6$$

2. 数据特征描述

- **分布形状:** 数据分布相对**对称**。中位数（1.6）非常接近均值（1.65），箱体（Q1 到 Q2 和 Q2 到 Q3 的距离）和晶须的长度也大致对称，表明数据没有明显的偏斜。
- **离散程度:** 数据的全距为 $3.1 - 0.2 = 2.9$，四分位距为 1.6。数据点分布比较均匀，没有极端值，整体离散程度适中。

3. 均值与中位数的差异分析

- **均值 (Average) 计算:**

$$\bar{x} = \frac{0.2 + 0.4 + 0.8 + 1.2 + 1.4 + 1.8 + 2.2 + 2.4 + 3.0 + 3.1}{10} = \frac{16.5}{10} = 1.65$$

- **差异比较与原因:**
  - **差异:** 样本均值 1.65 与中位数 1.6 非常接近，差异很小。
  - **原因:** 这种微小差异的原因在于数据分布的**高度对称性**和**无异常值**。均值和中位数都是衡量数据中心趋势的指标。当数据分布对称时，这两个指标的值会非常接近。与之前的偏态数据不同，这里没有极端值来" 拉动" 均值，因此均值和中位数都能很好地代表数据的中心点。

(AI 生成部分答案仅供参考，建议参考书本，并且不保证数据与原卷一致:)

---

2 第二题：置信区间与检验方法

题目描述：假设一个随机变量 X 从一个正态分布 $N(\mu, \sigma^2)$ 中抽样，其中总体均值 μ 和方差 σ^2 均未知。设定的置信水平为 90%。

问题：

1. 在 σ^2 未知的情况下，作 T 检验，写出置信区间。请问置信水平 90% 是否意味着真值落入置信区间的概率为 90%?
2. 假设现在总体标准差 σ 已知，作 Z 检验，写出置信区间，并对比 1）中的结果

3. 根据 1），2）说明 T 检验和 Z 检验的应用场景。

解答：

1. T 检验与置信水平的解释

- **检验方法与置信区间：** 当总体方差 σ^2 未知时，我们必须用样本标准差 s 来估计它。此时，应当使用 **T 分布**来构造均值的置信区间。设样本大小为 n，样本均值为 $\bar{x}$，样本标准差为 s。则均值 μ 的 90% 置信区间表达式为：

$$\left(\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \quad \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

  其中：

  - $\alpha = 1 - 0.90 = 0.10$，所以 $\alpha/2 = 0.05$。
  - $t_{0.05,n-1}$ 是自由度为 $n-1$ 的 t 分布上侧 5% 的分位数。

- **置信水平 90% 的含义：** “置信水平为 90%” **并不意味着**总体真值 μ 有 90% 的概率落入某一个具体的、已计算出的置信区间内。正确的解释是：如果我们以相同的方式，从同一个总体中**重复抽取无数次样本**，并为每一次抽样都构造一个 90% 的置信区间，那么在这些无限多的置信区间中，**大约有 90% 的区间会包含总体的真实均值** μ。对于我们计算出的任何一个特定区间，μ 要么在其中，要么不在，不存在概率问题。

2. Z 检验与结果对比

- **检验方法与置信区间：** 当总体标准差 σ 已知时，我们应当使用 **Z 分布（正态分布）**来构造均值的置信区间。其 90% 置信区间的表达式为：

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

  其中，$z_{\alpha/2} = z_{0.05} \approx 1.645$。

- **与 T 区间的对比：**

  1. **分布基础：** Z 区间基于正态分布，而 T 分布的尾部比正态分布更” 厚”，这意味着在相同的置信水平下，T 分布的临界值 $t_{\alpha/2,n-1}$ 通常大于正态分布的临界值 $z_{\alpha/2}$。
  2. **区间宽度：** 因此，假设样本标准差 s 恰好等于总体标准差 σ，T 区间通常会比 Z 区间**更宽**。这种” 增宽” 是为了弥补使用样本 s 估计总体 σ 所带来的不确定性。当样本量 n 足够大时（通常认为 $n > 30$），T 分布会非常接近正态分布，$t_{\alpha/2,n-1}$ 会趋近于 $z_{\alpha/2}$，两个区间的宽度差异会变小。

3. T 检验与 Z 检验的应用场景

- **Z 检验 (Z-test):**
  - **核心前提: 总体方差 σ^2 已知**。
  - **应用场景:** 主要用于理论教学或某些已知总体参数的特定工业流程中。在实际研究中，总体方差通常是未知的，因此 Z 检验的应用相对较少。此外，当样本量 n 非常大时（例如 $n > 100$），可以用 Z 检验作为 T 检验的近似。
- **T 检验 (T-test):**
  - **核心前提: 总体方差 σ^2 未知**，需要用样本方差 s^2 来估计。
  - **应用场景:** 这是现实世界中**绝大多数**关于均值推断的场景。无论是进行假设检验还是构造置信区间，只要总体方差是未知的，就应该使用 T 检验。它是统计推断中应用最广泛的工具之一。

(AI 生成部分答案仅供参考，建议参考书本，并且不保证数据与原卷一致:)

---

3 第三题：假设检验

题目描述: 对一个服从正态分布的随机变量 X 进行假设检验。已知信息如下：

- 总体方差 $\sigma^2 = 25$ (即 $\sigma = 5$)。
- 原假设 $H_0 : \mu = 2$。
- 备择假设 $H_1 : \mu \neq 2$ (双边检验)。
- 显著性水平 $\alpha = 0.05$ (对应置信水平 95%)。
- 从总体中抽取一个样本，样本量 $n = 25$，计算得到样本均值 $\bar{x} = 2.5$。

问题:

1. 计算在原假设 H_0 为真时的拒绝域，并绘制该条件下的零分布图。
2. 计算检验统计量 z-value，并将其位置在 1）的零分布图中标出。

3. 计算本次检验的 p-value，并依据它判断是否应拒绝原假设 H_0。

解答:

1. 拒绝域与零分布

- **零分布 (Null Distribution):** 在原假设 $H_0 : \mu = 2$ 为真的前提下，样本均值 $\bar{x}$ 的抽样分布服从正态分布。其均值为 $\mu_0 = 2$，标准差（即标准误）为 $\frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1$。所以，零分布为 $N(2, 1^2)$。

- **拒绝域 (Rejection Region):** 这是一个双边检验，显著性水平 $\alpha = 0.05$，所以每侧的拒绝区域面积为 $\alpha/2 = 0.025$。我们需要找到标准正态分布两侧的临界值 $\pm z_{\alpha/2}$。

$$z_{0.025} = 1.96$$

拒绝域是检验统计量 Z 的值大于 1.96 或小于-1.96 的区域。即:

$$Z < -1.96 \quad 或 \quad Z > 1.96$$

- **零分布示意图:** 可以绘制一个以 2 为中心，标准差为 1 的正态分布曲线。在横轴上，标出拒绝域的边界点，即 $2 - 1.96 \times 1 = 0.04$ 和 $2 + 1.96 \times 1 = 3.96$。任何落在 $(-\infty, 0.04] \cup [3.96, +\infty)$ 区间的 $\bar{x}$ 值都会导致拒绝 H_0。

2. Z-Value 计算

- **计算检验统计量 (z-value):**

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.5 - 2}{5/\sqrt{25}} = \frac{0.5}{1} = 0.5$$

- **标注:** 在上述的零分布图（$N(2, 1^2)$）上，样本均值 $\bar{x} = 2.5$ 位于中心值 2 的右侧。对应的 z-value 是 0.5，这个值明显位于接受域 $(-1.96, 1.96)$ 之内，远离拒绝域。

3. P-Value 计算

- **计算 p-value:** p-value 是在原假设为真的情况下，观察到当前样本结果（或更极端结果）的概率。对于双边检验，我们需要计算 $|Z| > 0.5$ 的概率。

$$p\text{-value} = P(|Z| > 0.5) = P(Z > 0.5) + P(Z < -0.5)$$

由于正态分布的对称性，这等于:

$$p\text{-value} = 2 \times P(Z > 0.5)$$

查标准正态分布表或使用计算器可得 $P(Z > 0.5) = 1 - P(Z \leq 0.5) = 1 - 0.6915 = 0.3085$。

$$p\text{-value} = 2 \times 0.3085 = 0.617$$

- **决策:** 我们比较 p-value 和显著性水平 α。

$$0.617(\text{p-value}) > 0.05(\alpha)$$

因为 p-value 远大于显著性水平，我们**没有足够的证据拒绝原假设** H_0。**结论:** 统计证据不足以认为总体均值不等于 2。

---

4 第四题：参数估计

题目描述: 一个随机变量 X 的概率密度函数 (PDF) 为:

$$f(x;\lambda) = \lambda^2 x e^{-\lambda x}, \quad \text{for } x > 0$$

$X_1, X_2, \ldots, X_n$ 是从该分布中抽取的一组独立同分布的样本。

问题:

1. 求参数 λ 的矩估计量 (Method of Moments Estimator)。
2. 求参数 λ 的最大似然估计量 (Maximum Likelihood Estimator)。

解答:

1. 矩估计量 (MOM)

矩估计法的思想是用样本矩来估计总体矩。我们使用一阶矩（均值）。

- **第一步：计算总体均值** $E[X]$ 该概率密度函数是伽马分布 $\Gamma(k,\theta)$ 的一个特例，其中参数 $k = 2$，$\theta = 1/\lambda$。伽马分布的均值为 $E[X] = k\theta$。因此，

$$E[X] = 2 \cdot \frac{1}{\lambda} = \frac{2}{\lambda}$$

（也可以通过积分 $\int_0^\infty x \cdot f(x)dx$ 直接计算得到）

- **第二步：令总体均值等于样本均值**样本均值为 $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$。令 $E[X] = \bar{X}$:

$$\frac{2}{\lambda} = \bar{X}$$

- **第三步：解出 λ 的估计量** λ 的矩估计量 $\hat{\lambda}_{MOM}$ 为:

$$\hat{\lambda}_{MOM} = \frac{2}{\bar{X}}$$

2. 最大似然估计量 (MLE)

最大似然估计法的思想是找到一个参数值，使得当前观测到的这组样本出现的概率（似然）最大。

- **第一步：写出似然函数** $L(\lambda)$ 似然函数是所有样本点概率密度的连乘积：

$$L(\lambda; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \lambda) = \prod_{i=1}^{n} (\lambda^2 x_i e^{-\lambda x_i})$$

$$L(\lambda) = (\lambda^2)^n \left(\prod_{i=1}^{n} x_i \right) e^{-\lambda \sum_{i=1}^{n} x_i} = \lambda^{2n} \left(\prod_{i=1}^{n} x_i \right) e^{-\lambda n \bar{X}}$$

- **第二步：取对数，得到对数似然函数** $\ln L(\lambda)$ 取对数可以简化计算：

$$\ln L(\lambda) = \ln(\lambda^{2n}) + \ln \left(\prod_{i=1}^{n} x_i \right) + \ln(e^{-\lambda n \bar{X}})$$

$$\ln L(\lambda) = 2n \ln(\lambda) + \sum_{i=1}^{n} \ln(x_i) - \lambda n \bar{X}$$

- **第三步：对** λ **求导，并令其为 0**

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{2n}{\lambda} - n\bar{X}$$

令导数为 0：

$$\frac{2n}{\lambda} - n\bar{X} = 0$$

- **第四步：解出** λ **的估计量**

$$\frac{2n}{\lambda} = n\bar{X}$$

$$\hat{\lambda}_{MLE} = \frac{2n}{n\bar{X}} = \frac{2}{\bar{X}}$$

在此问题中，λ 的矩估计量与最大似然估计量相同。

---

5 第五题：构造统计分布

题目描述: 我们有两组来自正态分布的独立随机样本：

- 样本 $X_1, \ldots, X_n$ 来自总体 $N(\mu_1, \sigma_1^2)$。
- 样本 $Y_1, \ldots, Y_m$ 来自总体 $N(\mu_2, \sigma_2^2)$。

问题:

1. 假设 $\mu_1 = \mu_2 = \mu$，找到两个统计量 T_1 和 T_2，使它们都服从 **t 分布**且相互**独立**。解释其原理。
2. 假设 $\sigma_1^2 = \sigma_2^2 = \sigma^2$，定义一个统计量 T_3，使其服从 **F 分布**，并说明其自由度。
3. 定义新统计量 $T_4 = 1/T_3$，请问 T_4 服从什么分布?

解答:

1. 构造独立的 t 分布统计量

- **构造统计量:** 我们可以分别基于两个独立的样本来构造两个 t 统计量：
  - $T_1(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \frac{\bar{X} - \mu}{S_X/\sqrt{n}}$，其中 $\bar{X}$ 和 S_X 是 X 样本的均值和标准差。
  - $T_2(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \frac{\bar{Y} - \mu}{S_Y/\sqrt{m}}$，其中 $\bar{Y}$ 和 S_Y 是 Y 样本的均值和标准差。
- **服从 t 分布的原因:** t 分布由一个标准正态分布的随机变量除以一个独立的、除以其自由度的卡方分布随机变量的平方根构成。
  - 对于 T_1：$\frac{\bar{X} - \mu}{\sigma_1/\sqrt{n}}$ 服从标准正态分布 $N(0,1)$。$\frac{(n-1)S_X^2}{\sigma_1^2}$ 服从自由度为 $n-1$ 的卡方分布 χ^2_{n-1}。在正态总体中，样本均值和样本方差是独立的。因此，T_1 的构造符合 t 分布的定义，服从自由度为 $n-1$ 的 t 分布。
  - 同理，T_2 也符合 t 分布的定义，服从自由度为 $m-1$ 的 t 分布。
- **相互独立的原因:** 统计量 T_1 **仅是样本 X 的函数**。统计量 T_2 **仅是样本 Y 的函数**。由于原始样本 X 和样本 Y 是相互独立的，任何只依赖于各自样本的函数（如此处的 T_1 和 T_2）也必然是**相互独立的**。

2. 构造 F 分布统计量

- **定义统计量:** F 分布由两个独立的、各自除以其自由度的卡方分布随机变量的比值构成。在 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 的前提下，样本方差的比值可以构造 F 统计量。

$$T_3(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \frac{S_X^2}{S_Y^2}$$

- **原理与自由度:** 我们知道 $\frac{(n-1)S_X^2}{\sigma^2} \sim \chi_{n-1}^2$ 和 $\frac{(m-1)S_Y^2}{\sigma^2} \sim \chi_{m-1}^2$。根据 F 分布的定义:

$$F = \frac{\chi_{df1}^2/df1}{\chi_{df2}^2/df2} = \frac{\left(\frac{(n-1)S_X^2}{\sigma^2}\right)/(n-1)}{\left(\frac{(m-1)S_Y^2}{\sigma^2}\right)/(m-1)} = \frac{S_X^2/\sigma^2}{S_Y^2/\sigma^2} = \frac{S_X^2}{S_Y^2}$$

  因此，统计量 T_3 服从 **F 分布**。

  - 分子自由度 $df_1 = n - 1$。
  - 分母自由度 $df_2 = m - 1$。

3. T4 的分布

根据 F 分布的一个重要性质，如果一个随机变量 F 服从自由度为 (d_1, d_2) 的 F 分布，那么它的倒数 $1/F$ 将服从自由度为 (d_2, d_1) 的 F 分布。

在问题 2 中，我们已经知道:

$$T_3 = \frac{S_X^2}{S_Y^2} \sim F(n-1, m-1)$$

因此，新的统计量 T_4 为:

$$T_4 = \frac{1}{T_3} = \frac{1}{S_X^2/S_Y^2} = \frac{S_Y^2}{S_X^2}$$

我们可以通过 F 分布的定义来验证这一点:

$$T_4 = \frac{S_Y^2}{S_X^2} = \frac{\left(\frac{(m-1)S_Y^2}{\sigma^2}\right)/(m-1)}{\left(\frac{(n-1)S_X^2}{\sigma^2}\right)/(n-1)}$$

分子的表达式是自由度为 $m-1$ 的卡方变量除以其自由度，分母是自由度为 $n-1$ 的卡方变量除以其自由度。因此，T_4 服从 **F 分布**，其自由度与 T_3 的自由度正好相反:

- 分子自由度 $df_1 = m - 1$。
- 分母自由度 $df_2 = n - 1$。

所以，$T_4 \sim F(m-1, n-1)$。

6 第六题：无截距项的最小二乘法

题目描述: 给定一个数据集 $\{(x_i, y_i)|i = 1, \ldots, n\}$。我们使用一个无截距项的简单线性回归模型来拟合数据：

$$Y_i = \beta x_i + \epsilon_i$$

其中，误差项 ϵ_i 相互独立，且服从均值为 0、方差为 σ^2 的正态分布，即 $\epsilon_i \sim N(0, \sigma^2)$。

问题:

1. 求解斜率参数 β 的最小二乘估计量 $\hat{\beta}$。
2. 说明 $\hat{\beta}$ 服从什么分布。
3. 说明残差平方和 (RSS) 服从什么分布。

解答:

1. 求解最小二乘估计量 $\hat{\beta}$

最小二乘法的目标是最小化残差平方和 (RSS)。

$$RSS(\beta) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \beta x_i)^2$$

为了找到使 RSS 最小的 β，我们对 β 求导并令其等于 0。

$$\frac{d(RSS)}{d\beta} = \sum_{i=1}^{n} 2(y_i - \beta x_i)(-x_i) = -2\sum_{i=1}^{n}(x_i y_i - \beta x_i^2)$$

令导数为 0：

$$-2\left(\sum x_i y_i - \beta \sum x_i^2\right) = 0$$

$$\sum x_i y_i = \beta \sum x_i^2$$

解出 β 的最小二乘估计量 $\hat{\beta}$：

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

2. $\hat{\beta}$ 的分布

$\hat{\beta}$ 是观测值 y_i 的线性组合，其中 y_i 是正态随机变量。正态变量的线性组合仍然是正态变量，因此 $\hat{\beta}$ 服从**正态分布**。我们可以进一步确定其均值和方差：

- **均值 $E[\hat{\beta}]$:**

$$E[\hat{\beta}] = E\left[\frac{\sum x_i y_i}{\sum x_i^2}\right] = \frac{\sum x_i E[y_i]}{\sum x_i^2}$$

因为 $E[y_i] = E[\beta x_i + \epsilon_i] = \beta x_i$，所以：

$$E[\hat{\beta}] = \frac{\sum x_i(\beta x_i)}{\sum x_i^2} = \frac{\beta \sum x_i^2}{\sum x_i^2} = \beta$$

这是一个无偏估计。

- **方差 $\mathbf{Var}(\hat{\beta})$:**

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}\left(\frac{\sum x_i y_i}{\sum x_i^2}\right) = \frac{1}{(\sum x_i^2)^2}\mathrm{Var}\left(\sum x_i y_i\right)$$

因为 y_i 相互独立，所以 $\mathrm{Var}(\sum a_i Y_i) = \sum a_i^2 \mathrm{Var}(Y_i)$。

$$\mathrm{Var}(\hat{\beta}) = \frac{1}{(\sum x_i^2)^2} \sum x_i^2 \mathrm{Var}(y_i)$$

因为 $\mathrm{Var}(y_i) = \mathrm{Var}(\beta x_i + \epsilon_i) = \mathrm{Var}(\epsilon_i) = \sigma^2$，所以：

$$\mathrm{Var}(\hat{\beta}) = \frac{\sum x_i^2 \sigma^2}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}$$

综上所述，$\hat{\beta}$ 服从正态分布：

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right)$$

3. 残差平方和 (RSS) 的分布

我们可以利用第二问中关于 $\hat{\beta}$ 分布的结果，通过对总误差平方和进行分解来推导 RSS 的分布。

- **第一步：分解总误差平方和**

我们从真实误差项 $\epsilon_i = y_i - \beta x_i$ 的平方和入手。这是一个服从 $\sigma^2 \chi_n^2$ 分布的随机变量。

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta x_i)^2$$

通过在括号内加减 $\hat{\beta} x_i$，我们可以将其分解：

$$\sum (y_i - \hat{\beta} x_i + \hat{\beta} x_i - \beta x_i)^2 = \sum \left((y_i - \hat{\beta} x_i) + (\hat{\beta} - \beta) x_i\right)^2$$

展开平方项：

$$= \sum (y_i - \hat{\beta} x_i)^2 + 2 \sum (y_i - \hat{\beta} x_i)(\hat{\beta} - \beta) x_i + \sum ((\hat{\beta} - \beta) x_i)^2$$

中间的交叉项为 $2(\hat{\beta}-\beta)\sum x_i(y_i-\hat{\beta}x_i)$。根据最小二乘法的正规方程 (Normal Equation)，我们知道 $\sum x_i(y_i-\hat{\beta}x_i)=0$。因此交叉项为 0。

这样，我们就得到了一个关键的分解式：

$$\sum_{i=1}^{n}\epsilon_i^2=\sum_{i=1}^{n}(y_i-\hat{\beta}x_i)^2+(\hat{\beta}-\beta)^2\sum_{i=1}^{n}x_i^2$$

写成更简洁的形式：

$$\sum\epsilon_i^2=\mathrm{RSS}+(\hat{\beta}-\beta)^2\sum x_i^2$$

- **第二步：标准化并利用第二问的结果**

  将上式两边同除以 σ^2:

  $$\sum_{i=1}^{n}\left(\frac{\epsilon_i}{\sigma}\right)^2=\frac{\mathrm{RSS}}{\sigma^2}+\frac{(\hat{\beta}-\beta)^2\sum x_i^2}{\sigma^2}$$

  我们逐一分析这三项的分布：

  - **左侧项**: 因为 $\epsilon_i\sim N(0,\sigma^2)$，所以 $\epsilon_i/\sigma\sim N(0,1)$。左侧是 n 个独立标准正态变量的平方和，因此 $\sum(\epsilon_i/\sigma)^2\sim\chi_n^2$。
  - **右侧第二项**: 从第二问我们知道，$\hat{\beta}\sim N(\beta,\frac{\sigma^2}{\sum x_i^2})$。这意味着 $\frac{\hat{\beta}-\beta}{\sqrt{\sigma^2/\sum x_i^2}}\sim N(0,1)$。那么它的平方，即 $\frac{(\hat{\beta}-\beta)^2}{\sigma^2/\sum x_i^2}=\frac{(\hat{\beta}-\beta)^2\sum x_i^2}{\sigma^2}$，就服从自由度为 1 的卡方分布, χ_1^2。

- **第三步：利用矩生成函数 (MGF) 确定 RSS 的分布**

  我们的分解式现在可以写成：

  $$Q_n=Q_{rss}+Q_1$$

  其中 $Q_n\sim\chi_n^2$, $Q_1\sim\chi_1^2$, 且 $Q_{rss}=\mathrm{RSS}/\sigma^2$。根据科克伦定理，在线性模型中，RSS 与参数估计量 $\hat{\beta}$ 是独立的。因此，Q_{rss} 和 Q_1 也是独立的。

  对于独立的随机变量，其和的矩生成函数等于各自矩生成函数的乘积：

  $$M_{Q_n}(t)=M_{Q_{rss}}(t)\cdot M_{Q_1}(t)$$

  我们知道卡方分布 χ_k^2 的 MGF 是 $(1-2t)^{-k/2}$。代入可得：

  $$(1-2t)^{-n/2}=M_{Q_{rss}}(t)\cdot(1-2t)^{-1/2}$$

  求解 $M_{Q_{rss}}(t)$：

  $$M_{Q_{rss}}(t)=\frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}}=(1-2t)^{-(n-1)/2}$$

  这个结果正是自由度为 $n-1$ 的卡方分布的矩生成函数。

- **结论**

  因此，我们证明了：

$$\frac{\text{RSS}}{\sigma^2} \sim \chi^2_{n-1}$$

声明：本文档中的部分解答由 AI 生成，仅供学习和参考使用。