**产品质量管理项目报告**

**项目概述与目标**

本项目旨在应用概率论与数理统计知识，特别是统计过程控制 (SPC) 的方法，对一个（模拟的）生产过程进行质量分析与评估 .

主要目标包括：

评估产品生产的工艺水平

判断生产过程是否处于统计受控状态

项目将遵循以下步骤进行：数据收集/生成、描述性统计分析、正态性与均值检验、工序能力指数计算与评估、控制图绘制与分析，最终得出关于工艺水平和过程稳定性的结论

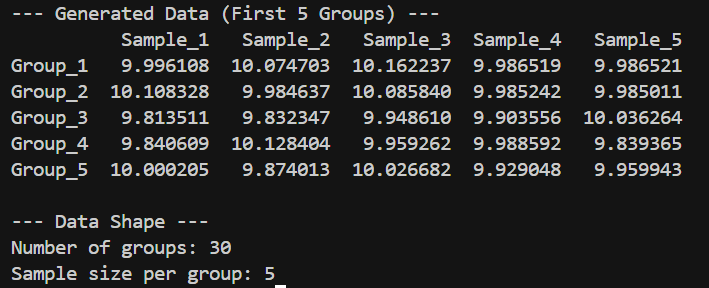
数据收集/生成

**要求：** 收集实际数据或生成模拟数据，至少 25 组，每组至少 5 个样本

**说明：** 我们生成模拟数据。假设我们监控某零件的直径，其目标值为 ，公差范围为 。我们模拟一个生产过程，生成 30 组数据，每组 5 个样本。

**Python 代码 (数据生成):**

|  |
| --- |
| Python import numpy as np import pandas as pd # --- Parameters --- num\_groups = 30 # 组数 (>= 25) sample\_size = 5 # 每组样本量 (>= 5) target\_mean = 10.0 # 假设目标均值 process\_std\_dev = 0.1 # 假设过程标准差 specification\_lower\_limit = 9.8 # 规格下限 TL specification\_upper\_limit = 10.2 # 规格上限 TU# --- Data Generation ---# 设置随机种子以保证结果可复现 np.random.seed(42) data = [] for i in range(num\_groups):# 模拟服从正态分布的数据# 可以在这里引入一些轻微的偏移或变异来模拟真实过程  group\_mean = target\_mean + np.random.normal(0, 0.02) # 模拟组间均值的轻微波动  group\_data = np.random.normal(loc=group\_mean, scale=process\_std\_dev, size=sample\_size)  data.append(group\_data)  # 将数据转换为 DataFrame 方便处理 df\_data = pd.DataFrame(data, index=[f'Group\_{i+1}' for i in range(num\_groups)]) df\_data.columns = [f'Sample\_{j+1}' for j in range(sample\_size)] print("--- Generated Data (First 5 Groups) ---") print(df\_data.head()) print("\n--- Data Shape ---") print(f"Number of groups: {df\_data.shape[0]}") print(f"Sample size per group: {df\_data.shape[1]}") |



**数据描述性统计分析**

**要求：** 计算均值、方差、极差，绘制直方图等

**说明：** 描述性统计帮助我们初步了解数据的中心趋势、离散程度和分布形状。

**总体均值 (Grand Mean, ):所有数据的平均值，估计过程的中心。**

**组内均值( ):每组数据的平均值。**

**组内标准差( ):每组数据的标准差，衡量组内波动**

**平均标准差( ):所有组内标准差的平均值，估计过程的平均波动**

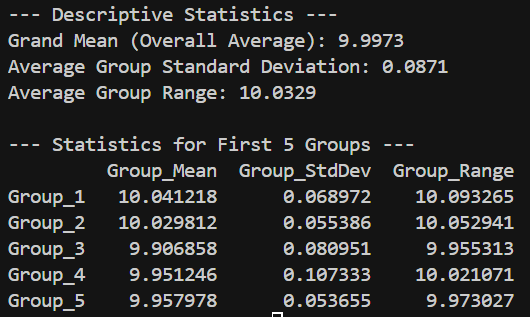
**组内极差( R\_i ):每组数据的最大值与最小值之差**

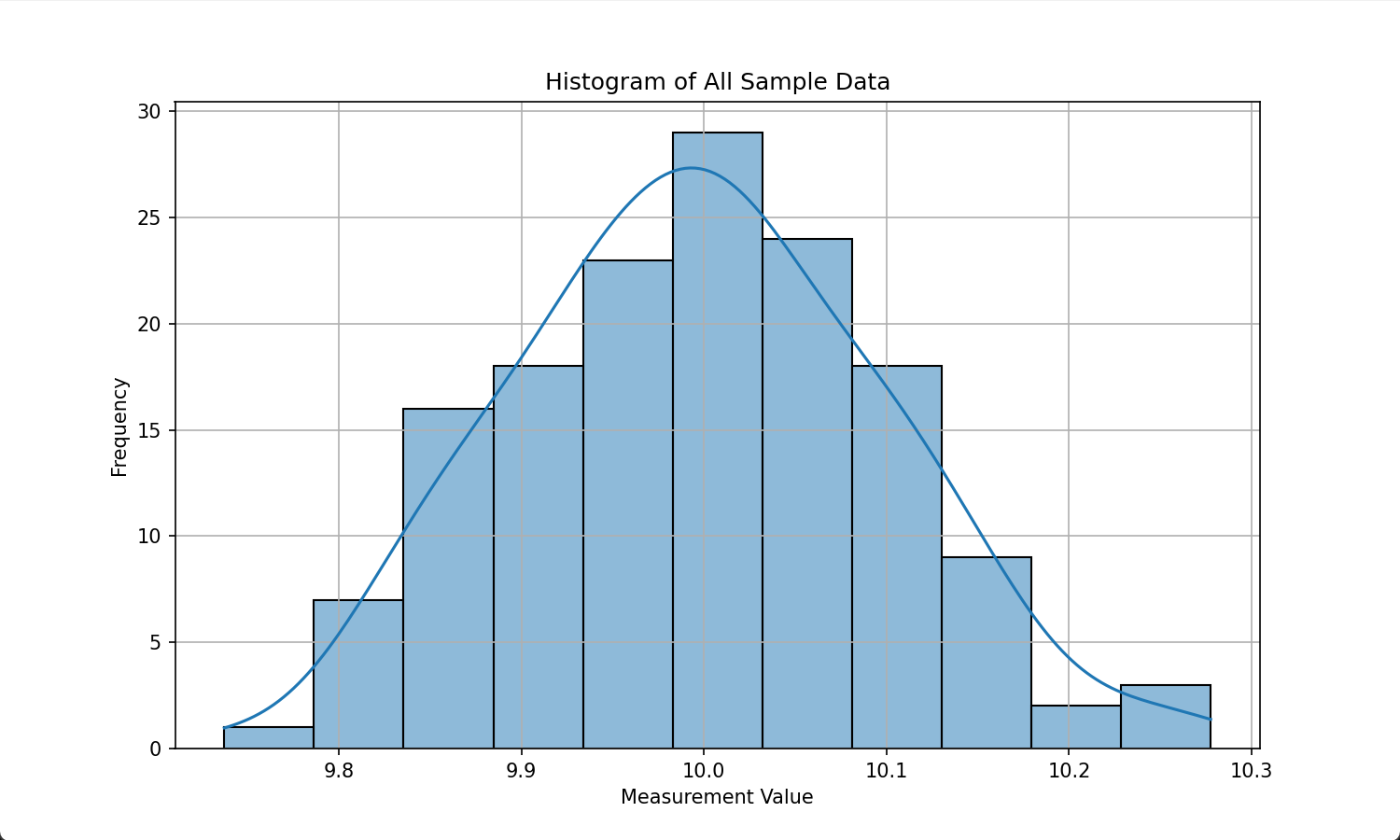
**平均极差( ):** 所有组内极差的平均值

**直方图:** 显示数据的频率分布，帮助判断是否接近正态分布。

**Python 代码 (描述性统计):**

|  |
| --- |
| Python import matplotlib.pyplot as plt import seaborn as sns  # --- Calculate Descriptive Statistics ---# Group means, standard deviations, ranges df\_data['Group\_Mean'] = df\_data.mean(axis=1) df\_data['Group\_StdDev'] = df\_data.std(axis=1, ddof=1) # 使用样本标准差 (ddof=1) df\_data['Group\_Range'] = df\_data.max(axis=1) - df\_data.min(axis=1)  # Overall statistics grand\_mean = df\_data['Group\_Mean'].mean() avg\_std\_dev = df\_data['Group\_StdDev'].mean() avg\_range = df\_data['Group\_Range'].mean() print("\n--- Descriptive Statistics ---") print(f"Grand Mean (Overall Average): {grand\_mean:.4f}") print(f"Average Group Standard Deviation: {avg\_std\_dev:.4f}") print(f"Average Group Range: {avg\_range:.4f}") print("\n--- Statistics for First 5 Groups ---") print(df\_data[['Group\_Mean', 'Group\_StdDev', 'Group\_Range']].head())  # --- Histogram of all data points --- all\_data\_flat = df\_data.iloc[:, :sample\_size].values.flatten() plt.figure(figsize=(10, 6)) sns.histplot(all\_data\_flat, kde=True) # kde=True 添加核密度估计曲线 plt.title('Histogram of All Sample Data') plt.xlabel('Measurement Value') plt.ylabel('Frequency') plt.grid(True) plt.show() |





**数据正态性检验与总体均值检验**

**要求：** 进行数据正态性检验和总体均值检验

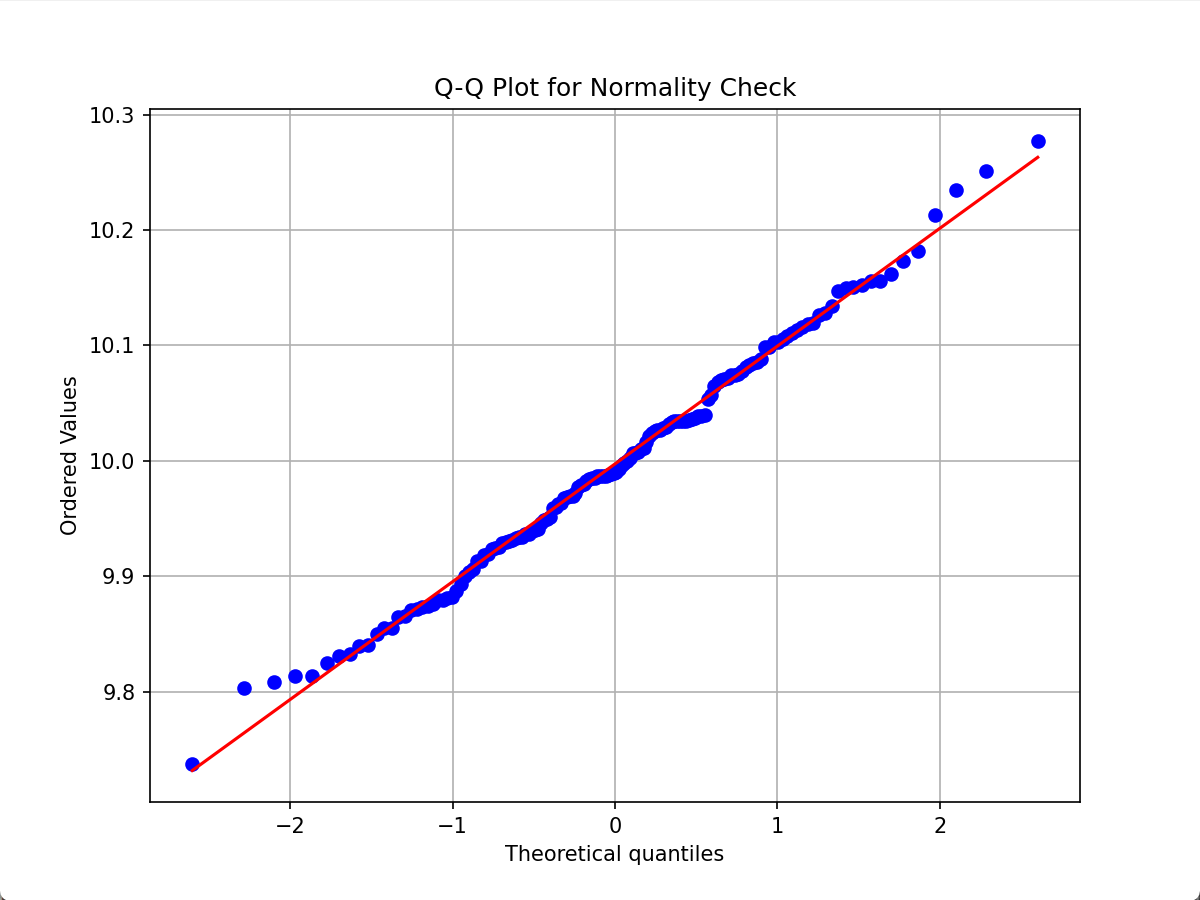
**说明：**

**正态性检验:** 许多 SPC 工具（如 Cpk 计算和基于 的控制图）假设数据服从正态分布 。我们需要检验这个假设。常用的方法有 Shapiro-Wilk 检验或 Kolmogorov-Smirnov 检验。Q-Q 图也是一种直观的图形检验方法。

**总体均值检验:** (PPT 中未明确要求此项检验，但提到了总体均值的概念 ())。 如果有目标均值 ，可以进行单样本 t 检验，判断过程的实际均值是否显著偏离目标值。

**Python 代码 (正态性检验 & 均值检验):**

|  |
| --- |
| Python from scipy import stats # --- Normality Test (Shapiro-Wilk) ---# 对所有数据点进行检验 shapiro\_test\_stat, shapiro\_p\_value = stats.shapiro(all\_data\_flat) print("\n--- Normality Test (Shapiro-Wilk on all data) ---") print(f"Test Statistic: {shapiro\_test\_stat:.4f}") print(f"P-value: {shapiro\_p\_value:.4f}") alpha\_normality = 0.05if shapiro\_p\_value > alpha\_normality:  print("P-value > alpha: Fail to reject the null hypothesis (Data appears normally distributed).") else:  print("P-value <= alpha: Reject the null hypothesis (Data does not appear normally distributed).")  # --- Q-Q Plot --- plt.figure(figsize=(8, 6)) stats.probplot(all\_data\_flat, dist="norm", plot=plt) plt.title('Q-Q Plot for Normality Check') plt.grid(True) plt.show()  # --- One-Sample T-test (Example: Test if mean is equal to target\_mean) ---# H0: population mean = target\_mean# Ha: population mean != target\_mean t\_test\_stat, t\_p\_value = stats.ttest\_1samp(all\_data\_flat, target\_mean) print("\n--- One-Sample T-test (vs Target Mean) ---") print(f"T-statistic: {t\_test\_stat:.4f}") print(f"P-value: {t\_p\_value:.4f}") alpha\_ttest = 0.05if t\_p\_value > alpha\_ttest:  print(f"P-value > alpha: Fail to reject H0 (No significant difference from target mean {target\_mean}).") else:  print(f"P-value <= alpha: Reject H0 (Significant difference from target mean {target\_mean}).") |



**结果解释:**

**正态性检验:** 如果 p 值大于显著性水平 (通常为 0.05)，则不能拒绝数据来自正态分布的原假设。Q-Q 图上的点应大致落在直线上。若不符合正态性，可能需要数据变换或使用非参数方法

**均值检验:** 如果 p 值大于显著性水平，则没有足够证据表明总体均值与目标值有显著差异。

**工序能力指数 (Cpk) 计算与评估**

**要求：** 计算并评估数据工序能力指数

**说明:** 工序能力指数衡量生产过程满足规格要求的能力。

**潜在工序能力指数 ( ):** 衡量过程固有变异相对于规格宽度的能力，不考虑过程均值的偏移

**实际工序能力指数 ($ C\_{pk} $):** 考虑了过程均值 相对于规格中心 的偏移 。

或者更常用的定义是： (这等价于考虑单侧能力的 和 )

**估算 和 :** 通常用总体均值 估计 。 可以用多种方法估计，一种常用方法是通过平均标准差 除以一个与样本量相关的常数 ( )，或者使用所有数据的合并标准差。 我们这里使用所有数据的标准差作为 的估计

**评估标准 (常见工业标准):**

: 工序能力不足，需要改进。

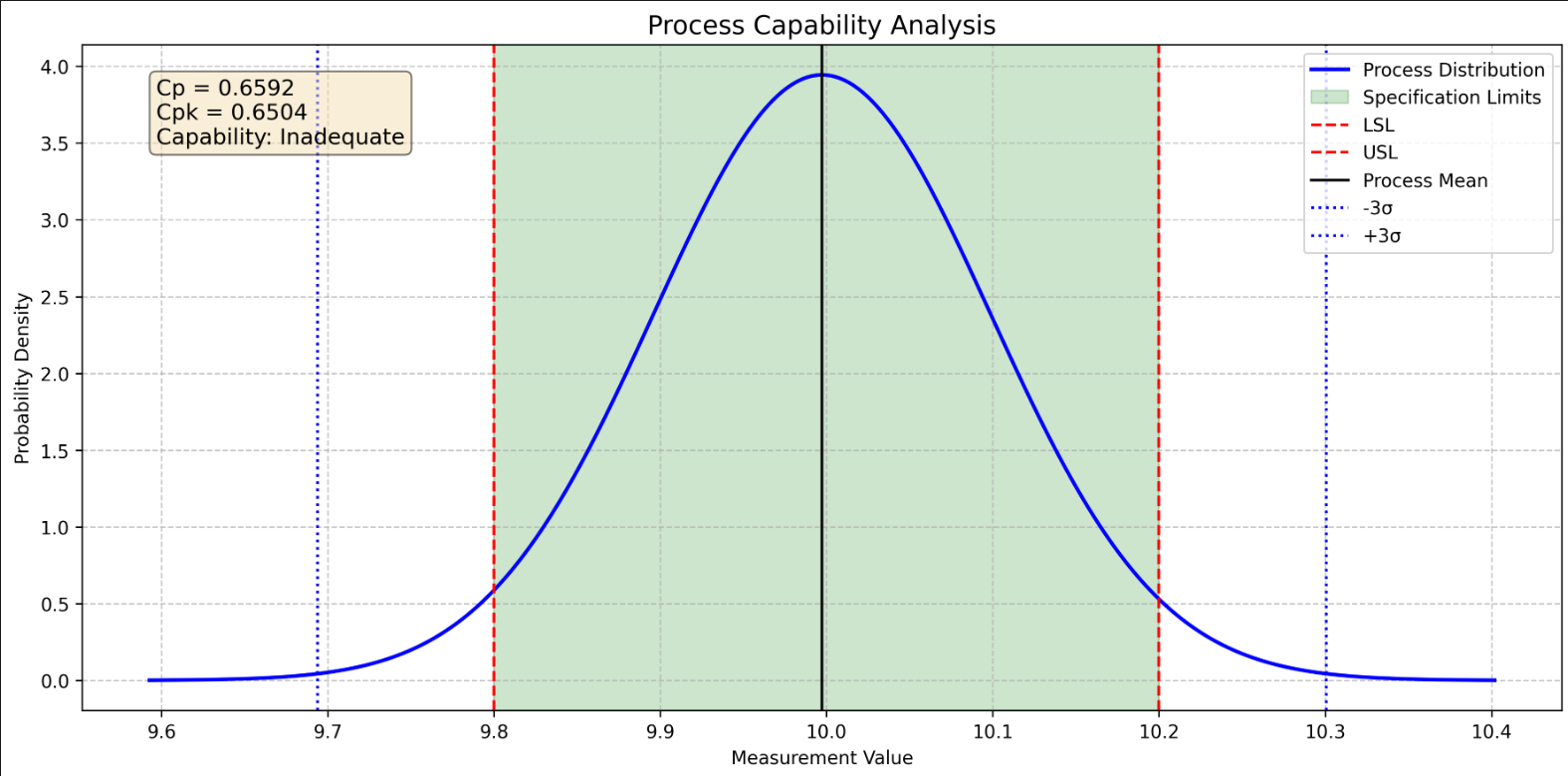
: 工序能力尚可，可能需要控制。

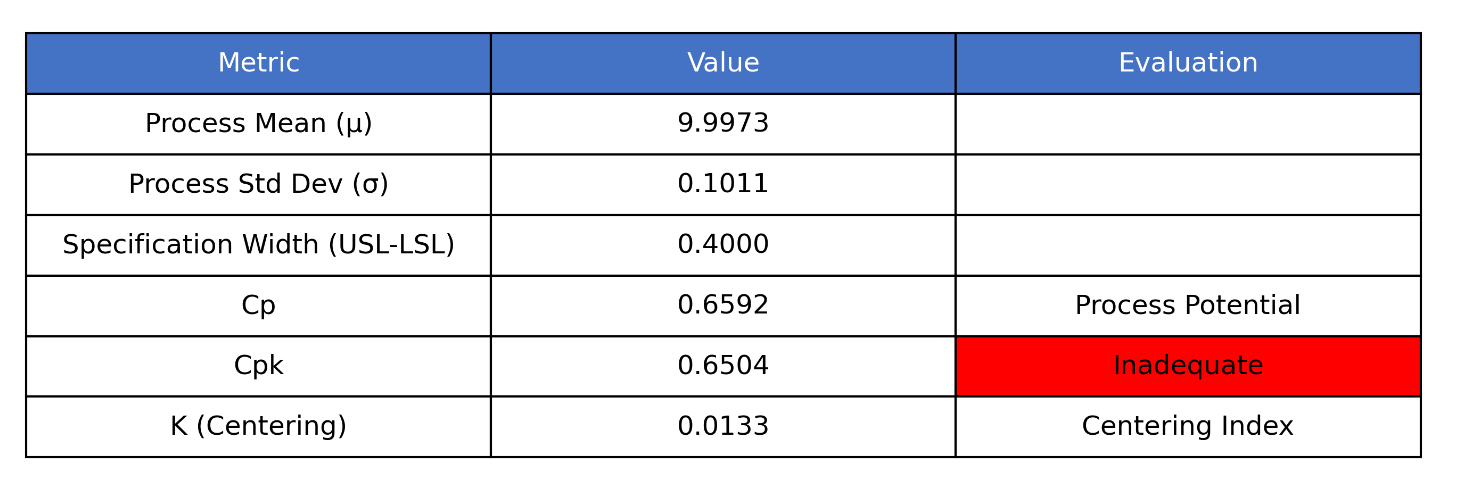
: 工序能力充足。

: 工序能力过高 (可能成本也高)。

**Python 代码 (Cpk 计算):**

|  |
| --- |
| Plain Text --- Estimate Process Parameters ---# Estimate mu using the grand mean mu\_hat = grand\_mean  # Estimate sigma using the standard deviation of all data points sigma\_hat = np.std(all\_data\_flat, ddof=1) # Use sample standard deviation of all points print("\n--- Process Parameter Estimates ---") print(f"Estimated Process Mean (mu\_hat): {mu\_hat:.4f}") print(f"Estimated Process Std Dev (sigma\_hat): {sigma\_hat:.4f}")  # --- Calculate Cp ---if sigma\_hat > 0:  Cp = (specification\_upper\_limit - specification\_lower\_limit) / (6 \* sigma\_hat)  print(f"\nPotential Process Capability (Cp): {Cp:.4f}") else:  Cp = float('inf')  print("\nPotential Process Capability (Cp): Infinite (sigma\_hat is zero)")  # --- Calculate Cpk ---if sigma\_hat > 0:  Cpu = (specification\_upper\_limit - mu\_hat) / (3 \* sigma\_hat)  Cpl = (mu\_hat - specification\_lower\_limit) / (3 \* sigma\_hat)  Cpk = min(Cpu, Cpl)# Alternative Cpk using K (from PPT)  T0 = (specification\_upper\_limit + specification\_lower\_limit) / 2if (specification\_upper\_limit - specification\_lower\_limit) > 0:  K = abs(mu\_hat - T0) / ((specification\_upper\_limit - specification\_lower\_limit) / 2)  Cpk\_alt = Cp \* (1 - K) if K <= 1 else 0 # [cite: 7]  print(f"Process Capability Index (Cpk) based on min(Cpu, Cpl): {Cpk:.4f}")#print(f"Process Capability Index (Cpk\_alt) based on K: {Cpk\_alt:.4f}") # Usually Cpk = min(Cpu, Cpl) is preferredelse:  Cpk = float('inf') # Handle case where TU=TL  print("Cannot calculate Cpk (TU=TL)")  else:  Cpk = float('inf')  print("\nProcess Capability Index (Cpk): Infinite (sigma\_hat is zero)")  # --- Evaluate Cpk --- print("\n--- Cpk Evaluation ---") if Cpk >= 1.67:  print(f"Cpk = {Cpk:.2f}: Process capability is excellent.") elif Cpk >= 1.33:  print(f"Cpk = {Cpk:.2f}: Process capability is adequate/good.") elif Cpk >= 1.0:  print(f"Cpk = {Cpk:.2f}: Process capability is marginal. Improvement may be needed.") else:  print(f"Cpk = {Cpk:.2f}: Process capability is inadequate. Significant improvement needed.")  # Check if mean is outside specification limitsif K > 1:  print("Warning: Estimated process mean is outside the specification limits (K > 1) |





Process Capability Analysis Ability Table

**均值 ( )和标准差(S)控制图描绘**

**要求：** 描绘均值控制图和标准差控制图 。 **说明：** 控制图用于监控过程随时间的变化，判断过程是否处于统计受控状态

这意味着过程的变异主要是由随机因素引起的，而非可识别的异常因素 。

控制图 (均值控制图): 监控过程均值的稳定性 ()。

中心线 (CL): (总体均值) () 控制限 (LCL): (使用 时) 或 是与样本量 相关的常数，需要查表获得

**S 控制图 (标准差控制图):** 监控过程变异的稳定性 ()。

中心线 (CL):

控制限 (UCL/LCL): (UCL), (LCL) , 是与样本量 n 相关的常数。(PPT中 对应 )。

**控制图常数 (部分示例，需根据实际样本量 n 查完整表):** 对于 n=5:

(用于 图, 基于 S)

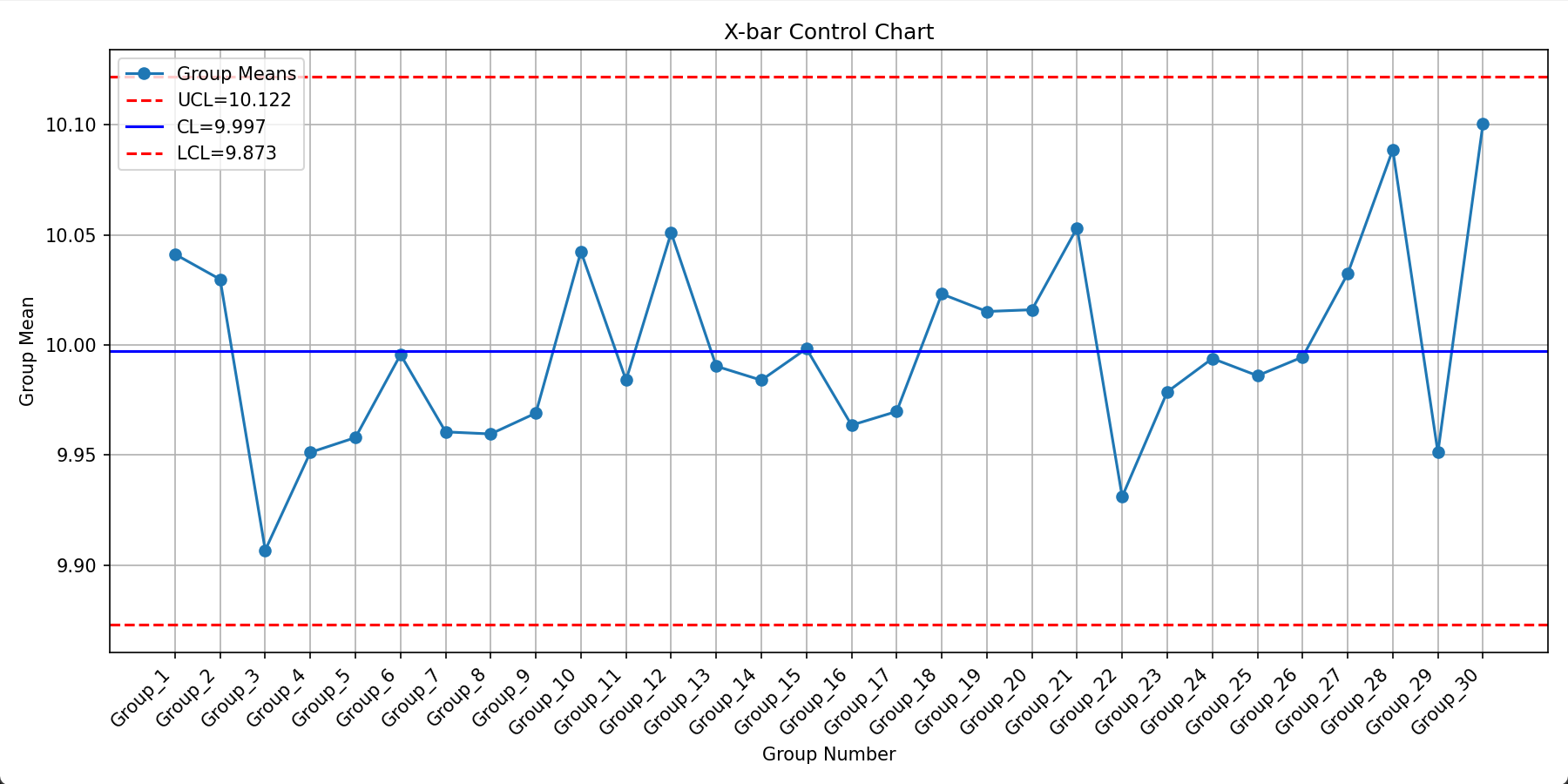
(用于 S 图)

(用于 S 图)

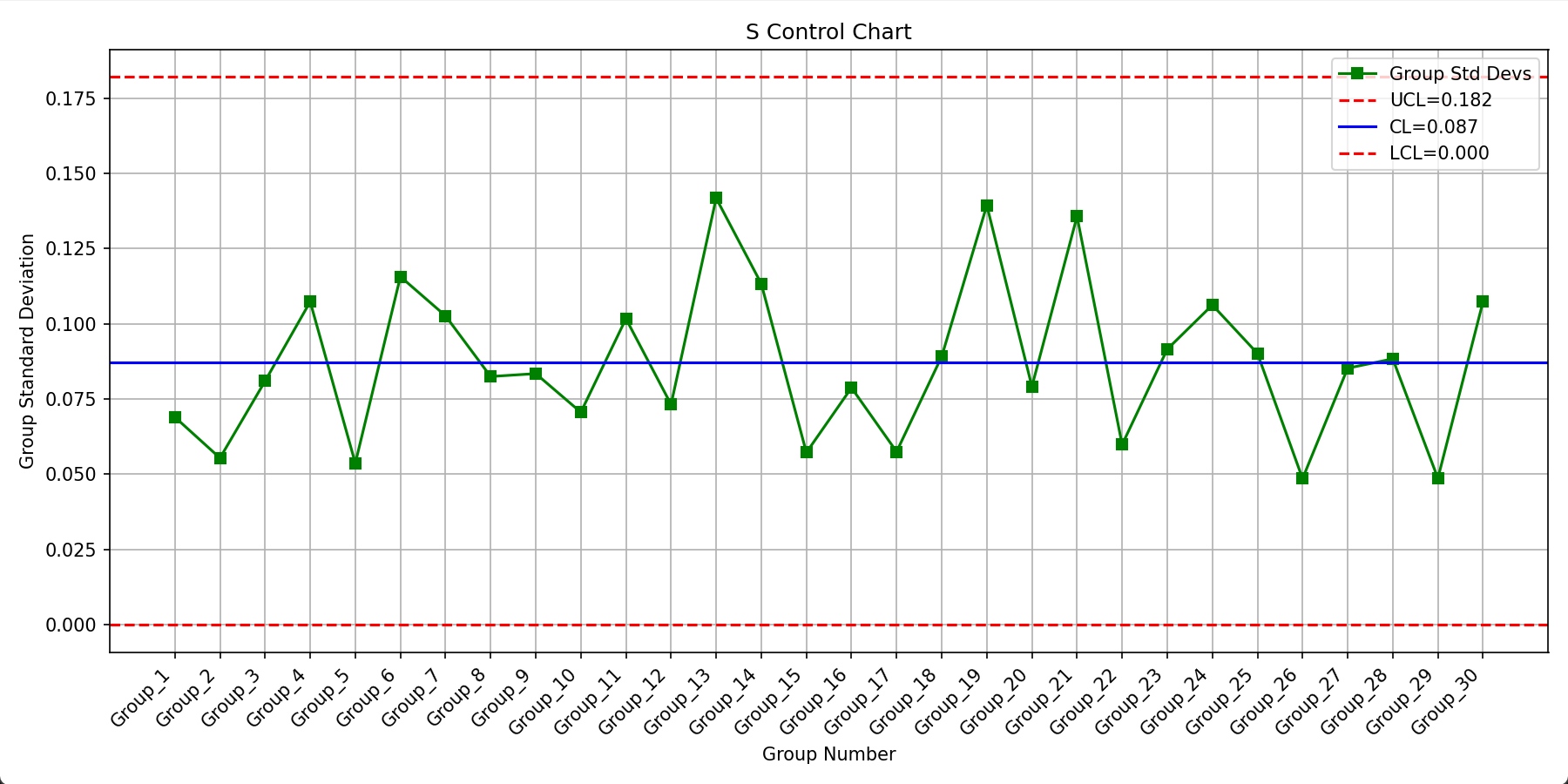
(用于估计 $ \sigma $)

**Python 代码 (控制图):**

|  |
| --- |
| Plain Text --- Control Chart Constants (for n=5) ---# These constants depend on the sample size 'n'# You need to look these up in standard SPC tables A3 = 1.427 # For X-bar chart with S B3 = 0 # For S chart (Lower limit factor) B4 = 2.089 # For S chart (Upper limit factor)# --- X-bar Chart Calculations --- xbar\_cl = grand\_mean xbar\_ucl = grand\_mean + A3 \* avg\_std\_dev xbar\_lcl = grand\_mean - A3 \* avg\_std\_dev  # --- S Chart Calculations --- s\_cl = avg\_std\_dev s\_ucl = B4 \* avg\_std\_dev s\_lcl = B3 \* avg\_std\_dev  # --- Plot X-bar Chart --- plt.figure(figsize=(12, 6)) plt.plot(df\_data.index, df\_data['Group\_Mean'], marker='o', linestyle='-', label='Group Means') plt.axhline(xbar\_ucl, color='r', linestyle='--', label=f'UCL={xbar\_ucl:.3f}') plt.axhline(xbar\_cl, color='b', linestyle='-', label=f'CL={xbar\_cl:.3f}') plt.axhline(xbar\_lcl, color='r', linestyle='--', label=f'LCL={xbar\_lcl:.3f}') plt.title('X-bar Control Chart') plt.xlabel('Group Number') plt.ylabel('Group Mean') plt.xticks(rotation=45, ha='right') plt.legend() plt.grid(True) plt.tight\_layout() plt.show()  # --- Plot S Chart --- plt.figure(figsize=(12, 6)) plt.plot(df\_data.index, df\_data['Group\_StdDev'], marker='s', linestyle='-', color='g', label='Group Std Devs') plt.axhline(s\_ucl, color='r', linestyle='--', label=f'UCL={s\_ucl:.3f}') plt.axhline(s\_cl, color='b', linestyle='-', label=f'CL={s\_cl:.3f}') plt.axhline(s\_lcl, color='r', linestyle='--', label=f'LCL={s\_lcl:.3f}') plt.title('S Control Chart') plt.xlabel('Group Number') plt.ylabel('Group Standard Deviation') plt.xticks(rotation=45, ha='right') plt.legend() plt.grid(True) plt.tight\_layout() plt.show() |



**此S控制图表明，尽管过程的Cpk值较低(0.65)表示过程能力不足，但过程的变异性是处于统计控制状态的。这意味着改进应该集中在减小整体变异或调整过程均值，而不是消除特殊原因导致的异常变异。**



**分析与判断:**

**工艺水平 (基于 Cpk):**

查看计算出的 值。

根据评估标准 (如 为能力充足) 判断工艺水平是否满足要求 。如果 低，说明过程固有变异相对于规格来说太大，或者过程均值偏离规格中心太远 。需要分析是 不足 (变异大) 还是 K 值过大 (均值偏移)。

**过程受控状态 (基于控制图):**

检查 图和 S 图。

**寻找失控信号:** 是否有点超出控制限 (UCL/LCL)? 是否存在违反其他运行规则的情况 (如连续多点在中心线一侧、连续递增/递减等)?

**结论:**

如果两个图上的点都在控制限内，并且没有明显的非随机模式 (如趋势、周期性等)，则可以认为过程处于统计受控状态。这意味着过程变异主要是随机的。

如果存在失控点或明显的非随机模式，则表明过程受到异常因素影响，处于失控状态。需要调查并消除异常原因，使过程恢复受控

**综合结论示例 (基于模拟数据可能的结果):**

**通过对模拟数据的分析**：

**工艺水平:** 计算得到的 值为 [在此处插入计算出的 Cpk 值]。根据 [在此处插入评估标准，如 Cpk >= 1.33]，该工艺水平被评估为 [在此处插入评估结果，如：充足、尚可、不足]。

**过程控制:** 均值控制图和标准差控制图显示 [在此处描述控制图情况，如：所有点均在控制限内且未观察到明显非随机模式 / 观察到第 X 组数据点超出了 Y 图的 Z 控制限]。因此，该生产过程 [在此处插入结论，如：处于统计受控状态 / 未处于统计受控状态]。

**建议:** 如果过程失控，需要识别并排除异常因素。如果过程受控但 Cpk 不足，则需要进行过程改进以减少变异或调整过程均值。