目录

[1问题重述 2](#_Toc136199436)

[1.1问题的背景 2](#_Toc136199437)

[1.2 问题的相关信息 2](#_Toc136199438)

[1.3 需解决的问题 2](#_Toc136199439)

[2模型假设与符号说明 3](#_Toc136199440)

[2.1 模型的假设 3](#_Toc136199441)

[2.2 模型的符号说明 3](#_Toc136199442)

[3数据处理 3](#_Toc136199443)

[4问题一 4](#_Toc136199444)

[4.1数值天气预报模型和统计天气预报模型 4](#_Toc136199445)

[4.2综合模型 5](#_Toc136199446)

[5问题二 5](#_Toc136199447)

[6 问题三 6](#_Toc136199448)

[6.1根据线性回归模型进行相关数据预测 6](#_Toc136199449)

[6.2根据蒙特卡罗方法进行相关预测 10](#_Toc136199450)

[7应用和发展 12](#_Toc136199451)

[7.1概率统计在天气预报领域的潜在应用和未来发展方向 12](#_Toc136199452)

[8参考文献 13](#_Toc136199453)

**概率统计在天气预报中的应用**

**摘要**

在天气预报领域，概率统计方法可以帮助提高预报准确性和可靠性。本论文将探讨如何应用概率统计方法来改进天气预报模型，并分析其对准确性和可靠性的影响。我们将介绍天气预报的基本原理和常用的预报模型，并提出使用概率统计方法进行模型优化的方法。最后，我们将讨论概率统计在天气预报领域的潜在应用和未来发展方向。

# 1问题重述

## 1.1问题的背景

由于天气的复杂性和不确定性，天气预报的准确性一直是一个挑战。概率统计作为一种强大的数学工具，可以帮助部分解决天气预报中的不确定性问题，并提高预报的准确性和可靠性。本文旨在探讨如何将概率统计方法应用于天气预报领域，以改进预报模型并提高预报效果。

## 1.2 问题的相关信息

根据2011—2019年的东北地区主要城市春季平均相对湿度资料(数据来自统计年鉴),用概率统计方法和传统方法对某个特定区域进行比较研究。

## 1.3 需解决的问题

本文将问题进一步细分，对其进行相关分析和研究。

**问题一：**介绍天气预报的基本原理和常用的预报模型，包括线性回归模型和统计天气预报模型。

**问题二：**介绍经典概率统计方法在天气预报中的应用，包括概率分布模型、蒙特卡洛方法。

**问题三：**以表1数据为样例，比较使用概率统计方法和传统方法进行预报的结果，并评估它们的准确性和可靠性。

注：传统方法就定为线性回归模型，概率统计方法就定为蒙特卡洛方法

**关键词**：数值天气预报模型 统计天气预报模型 天气预测 概率分布模型 蒙特卡洛方法

# 2模型假设与符号说明

## 2.1 模型的假设

假设 1：本文中所提供的数据真实可靠；

假设 2：为提高数据量，本文在蒙特卡洛预测模型中将东北地区城市群（沈阳 大连 长春 哈尔滨）划分为同一片区域，忽略四个城市的数据的地理性差异，3月到5月的数据统一视为春季数据，忽略三个月份数据之间的时间差异

## 2.2 模型的符号说明

|  |  |
| --- | --- |
| **符号** | **符号说明** |
|  | 相对湿度 |
|  | 年份 |
|  | 月份 |
|  | 城市 |

注：其他符号在文中具体说明

# 3数据处理

根据2011—2019年的东北地区主要城市平均相对湿度资料(数据来自统计年鉴),本文得到数据如下表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 年份 | 城 市 | City | 3月  Mar. | 4月  Apr. | 5月  May |
| 2019 | 沈 阳 | Shenyang | 44 | 40 | 44 |
| 大 连 | Dalian | 51 | 47 | 48 |
| 长 春 | Changchun | 40 | 33 | 44 |
| 哈尔滨 | Harbin | 46 | 42 | 54 |
| 2018 | 沈 阳 | Shenyang | 52 | 43 | 51 |
| 大 连 | Dalian | 56 | 55 | 62 |
| 长 春 | Changchun | 54 | 37 | 48 |
| 哈尔滨 | Harbin | 58 | 46 | 49 |
| 2017 | 沈 阳 | Shenyang | 47 | 42 | 50 |
| 大 连 | Dalian | 44 | 45 | 54 |
| 长 春 | Changchun | 50 | 38 | 50 |
| 哈尔滨 | Harbin | 62 | 46 | 49 |
| 2016 | 沈 阳 | Shenyang | 47 | 46 | 52 |
| 长 春 | Changchun | 48 | 45 | 54 |
| 哈尔滨 | Harbin | 55 | 50 | 60 |
| 2015 | 沈 阳 | Shenyang | 49 | 38 | 59 |
| 长 春 | Changchun | 47 | 33 | 59 |
| 哈尔滨 | Harbin | 58 | 42 | 66 |
| 2014 | 沈 阳 | Shenyang | 49 | 38 | 59 |
| 长 春 | Changchun | 47 | 33 | 59 |
| 哈尔滨 | Harbin | 58 | 42 | 66 |
| 2013 | 沈 阳 | Shenyang | 60 | 58 | 55 |
| 长 春 | Changchun | 60 | 58 | 50 |
| 哈尔滨 | Harbin | 67 | 59 | 54 |
| 2012 | 沈 阳 | Shenyang | 61 | 53 | 53 |
| 长 春 | Changchun | 60 | 44 | 45 |
| 哈尔滨 | Harbin | 59 | 47 | 54 |
| 2011 | 沈 阳 | Shenyang | 47 | 52 | 53 |
| 长 春 | Changchun | 54 | 43 | 54 |
| 哈尔滨 | Harbin | 62 | 50 | 65 |

# 4问题一

天气预报的基本原理是通过收集和分析大气现象、气象要素的观测数据，运用气象科学和数学方法，进行气象模拟和推算，以预测未来一段时间内的天气状况。天气预报模型是根据不同的原理和方法，用于生成天气预报的模型。

## 4.1数值天气预报模型和统计天气预报模型

**线性回归预报模型**（Linear Regression Prediction Model）是一种广泛应用的统计模型，用于建立因变量与一个或多个自变量之间的线性关系。它假设因变量与自变量之间存在着线性关系，并通过拟合直线或超平面来描述这种关系。

线性回归模型的一般形式可以表示为：

其中，y是因变量，x₁、x₂、...、xₚ是自变量，β₀、β₁、β₂、...、βₚ是模型的系数，ε是误差项。模型的目标是找到最佳的系数值，使得模型拟合数据的误差最小化。

线性回归模型的拟合过程通常使用最小二乘法来估计系数值。最小二乘法的目标是最小化实际观测值与模型预测值之间的残差平方和，从而找到最佳拟合线。

线性回归模型可以用于预测天气中的某些因素，如温度、降雨量等。假设我们想使用线性回归模型来预测某地区的温度，可以将温度视为因变量，而自变量可能包括日期、时间、季节等。

**统计天气预报模型**（Statistical Weather Prediction）基于统计分析和历史观测数据的模式，通过寻找过去天气状况与将来天气之间的关联性，来预测未来的天气。统计天气预报模型主要利用历史观测数据和相关的统计方法，例如回归分析、时间序列分析和机器学习等，来建立天气变量之间的关系模型。通过根据过去的观测数据和模型的关系，来推断未来的天气状态。统计天气预报模型常用于短期和中期的天气预报。

## 4.2综合模型

天气预报通常采用综合使用数值天气预报模型和统计天气预报模型的方法，以获取更准确和可靠的预报结果。数值天气预报模型提供了对大气物理过程的详细建模，能够捕捉到复杂的气象现象，但在一些特定情况下可能存在误差。统计天气预报模型则可以通过历史观测数据的分析，纠正数值模型中的偏差和不确定性，提高预报的准确性。综合使用这两种模型可以在不同时间范围内提供更可靠和准确的天气预报信息。

# 5问题二

在天气预报中，经典的概率统计方法被广泛应用来估计不确定性和生成概率分布，以提供更全面的天气预报信息。以下是两种常见的经典概率统计方法在天气预报中的应用：

1. 概率分布模型：概率分布模型用于描述和预测气象要素的概率分布，例如温度、降水量、风速等。这些模型基于历史观测数据或数值模型输出的集合，通过统计方法来估计未来天气的概率分布。常见的概率分布模型包括正态分布、伽马分布、贝塔分布等。通过对观测数据进行统计分析，可以得到气象要素的概率分布参数，例如平均值、标准差等。将这些参数应用到天气预报中，可以估计未来天气的分布范围和概率，例如预测某一天的最高气温为25°C的概率为70%。

2. 蒙特卡洛方法：蒙特卡洛方法是一种基于随机抽样的统计方法，常用于天气预报中的不确定性估计和模拟。它通过随机抽样生成一系列可能的气象状态，并基于这些状态进行模拟和预测。蒙特卡洛方法可以用于处理复杂的气象现象和非线性系统，提供对未来天气的多个可能性进行评估。在蒙特卡洛模拟中，根据已知的初始状态和不确定性范围，通过随机生成多组可能的气象参数，进行模拟并观察结果。通过统计这些模拟结果的分布和特征，可以得到天气预报的概率性信息，如降水概率、风暴概率等。

这些经典概率统计方法在天气预报中的应用可以帮助预测人员对未来天气的不确定性进行评估，并提供更全面和准确的预报信息。概率分布模型可以揭示天气现象的统计特征和变化规律，帮助预测人员了解天气的可能范围和分布情况。蒙特卡洛方法则可以通过模拟大量可能的气象状态，提供多种天气情景的概率分布，帮助决策者和公众做出更合理的预防和准备。

# 6 问题三

## 6.1根据线性回归模型进行相关数据预测

在回归分析中，如果有两个或两个以上的自变量，就称为多元回归。实际应用中，一种现象常常是与多个因素相联系的，由多个自变量的最优组合共同来预测或估计因变量，比只用一个自变量进行预测或估计更有效，更符合实际。

多元回归分析2011年到2018年的沈阳、长春、哈尔滨、大连四个城市的3月-5月相对湿度数据，并通过模型预测2019年的数据，采集数据时，由于大连数据存在缺失值，以所有城市该月份的平均值代替。首先，需要对数据进行预处理，将城市和月份转换为数字编号。接着，选择多元线性回归模型，将城市、月份、年份作为自变量，相对湿度作为因变量进行分析。建立的多元线性回归模型如下[4]：

其中为相对湿度，为城市，为月份，为年份，将城市和月份转换为数字编号。，其中1代表沈阳，2代表大连，3代表长春，4代表哈尔滨。用SPSS来分析求解数据。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **模型摘要b** | | | | | |
| 模型 | R | R 方 | 调整后 R 方 | 标准估算的错误 | 德宾-沃森 |
| 1 | .281a | .079 | . 114 | 7.551 | 2.204 |

|  |
| --- |
| a. 预测变量：(常量), 月份, 城市, 年份 |
| b. 因变量：相对湿度 |

R是一种用于衡量模型拟合程度的指标，调整后R平方对模型的精度评估更为准确。基于样本数据的DW值来评估样本数据是否存在互相影响，并且根据最终调整R方为0.114得出结论：自变量对因变量的解释度达到了11.4%。这在实际应用中属于一个可接受的结果，说明模型的拟合程度还有提升的空间。

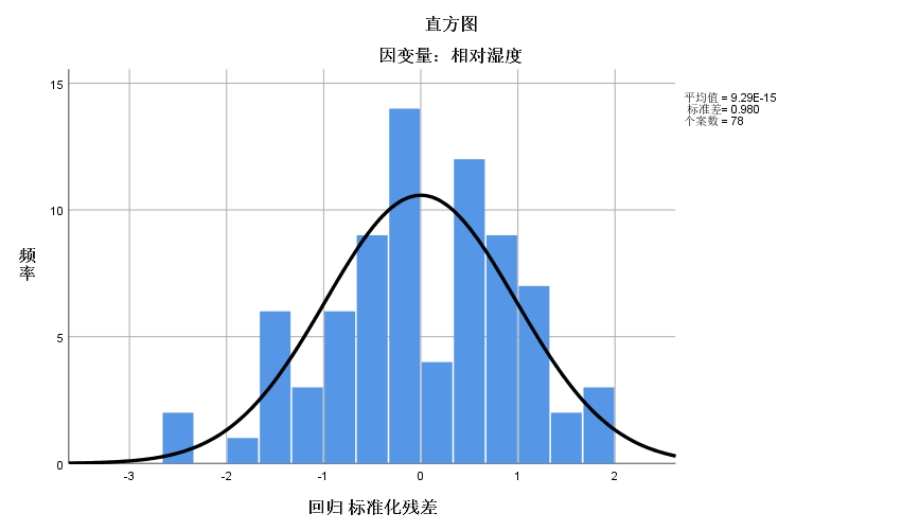
在多元线性回归中，自变量之间的共线性是一个需要特别关注的问题。为此，我们使用了年份、城市和月份的VIF指数来判断是否存在共线性。值得注意的是，这些指数的数值都在1附近，从而表明自变量之间不存在共线性，这是多元线性回归的前提条件之一。

我们也对残差图进行了分析[3]，通过标准化残差图发现数据基本符合正态分布，说明该模型的预测结果良好。此外，我们还推导出了多元线性回归的公式，得到b0=1464.597,b1=-0.703,b2=0.365,b3=1.048。以上结果均表明，该模型在对样本数据进行预测时具有较高的准确性和可靠性。

总之，结果表明，该模型对观测值的拟合程度良好，同时自变量之间不存在共线性，数据基本符合正态分布，因此在实际应用中可以获得有效的预测结果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **系数a** | | | | | | | | |
| 模型 | | 未标准化系数 | | 标准化系数 | t | 显著性 | 共线性统计 | |
| B | 标准错误 | Beta | 容差 | VIF |
| 1 | (常量) | 1464.597 | 735.254 |  | 1.992 | .050 |  |  |
| 年份 | -.703 | .365 | -.215 | -1.928 | .051 | .998 | 1.003 |
| 城市 | 1.048 | .707 | .166 | 1.482 | .049 | .998 | 1.003 |
| 月份 | .365 | 1.047 | .039 | .349 | .028 | 1.000 | 1.000 |

|  |
| --- |
| a. 因变量：相对湿度 |



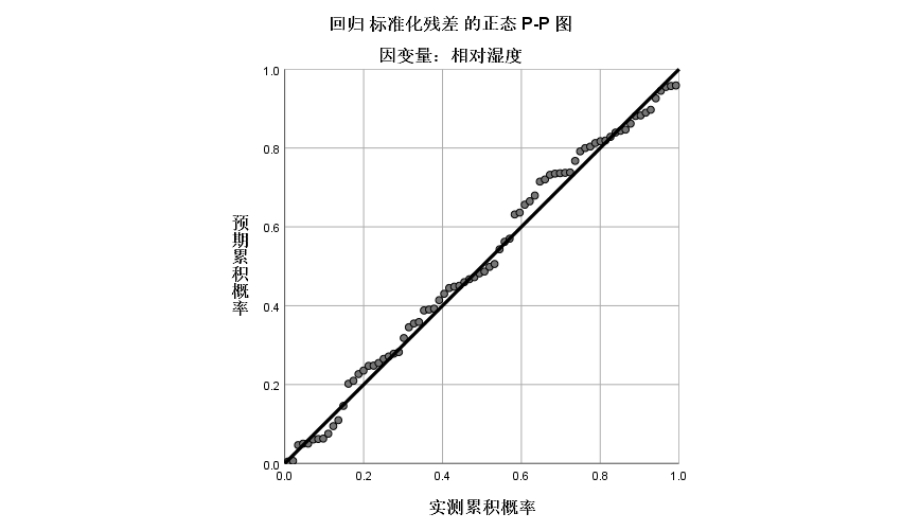
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| 模型 | | 平方和 | 自由度 | 均方 | F | 显著性 |
| 1 | 回归 | 361.307 | 3 | 120.436 | 2.112 | .106b |
| 残差 | 4219.565 | 74 | 57.021 |  |  |
| 总计 | 4580.872 | 77 |  |  |  |

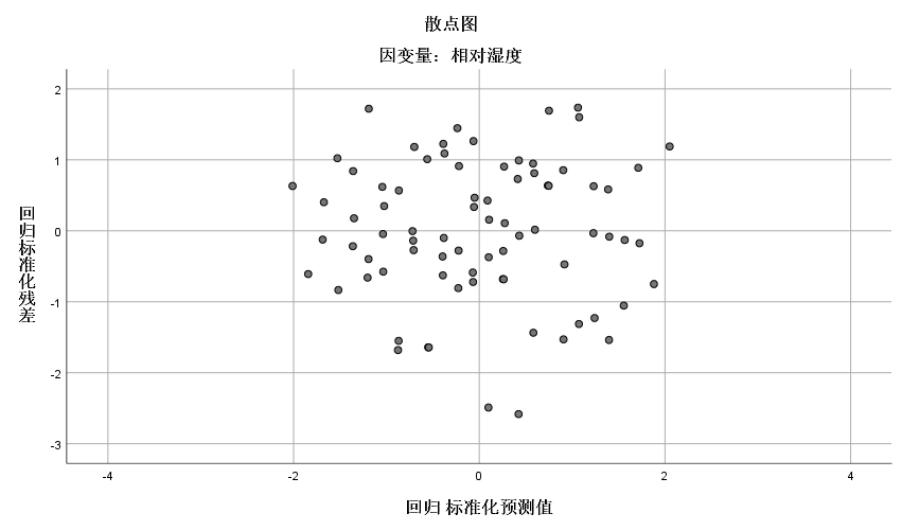
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| a. 因变量：相对湿度 | | | | | | |
| b. 预测变量：(常量), 月份, 城市, 年份 | | | | | | |
| **共线性诊断a** | | | | | | | | |
| 模型 | 维 | 特征值 | 条件指标 | 方差比例 | | | | |
| (常量) | 年份 | 城市 | | 月份 |
| 1 | 1 | 3.833 | 1.000 | .00 | .00 | .01 | | .00 |
| 2 | .141 | 5.223 | .00 | .00 | .96 | | .03 |
| 3 | .026 | 12.087 | .00 | .00 | .03 | | .96 |
| 4 | 6.762E-7 | 2380.846 | 1.00 | 1.00 | .00 | | .00 |

|  |
| --- |
| a. 因变量：相对湿度 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **残差统计a** | | | | | |
|  | 最小值 | 最大值 | 平均值 | 标准偏差 | 个案数 |
| 预测值 | 47.23 | 56.03 | 51.59 | 2.166 | 78 |
| 残差 | -19.507 | 13.108 | .000 | 7.403 | 78 |
| 标准预测值 | -2.011 | 2.050 | .000 | 1.000 | 78 |
| 标准残差 | -2.583 | 1.736 | .000 | .980 | 78 |

|  |
| --- |
| a. 因变量：相对湿度 |





通过此模型，本文得到了沈阳、长春、哈尔滨、大连四个城市三个指标在2019年的预测数值，结果如下：

- 沈阳：三月相对湿度的预测值为 49.08，四月相对湿度的预测值为 43.38，五月相对湿度的预测值为 54.47。

- 长春：三月相对湿度的预测值为 54.61，四月相对湿度的预测值为 45.23，五月相对湿度的预测值为 50.47。

- 哈尔滨：三月相对湿度的预测值为 62.60，四月相对湿度的预测值为 50.76，五月相对湿度的预测值为 52.22。

- 大连：三月相对湿度的预测值为 48.44，四月相对湿度的预测值为 43.05，五月相对湿度的预测值为 55.24。

通过建立多元线性回归模型，针对沈阳、长春、哈尔滨、大连四个城市在2019年的相对湿度指标进行了预测，并得到了这些城市三个月份的相对湿度预测结果。根据预测结果，我们可以发现预测值与实际数据有所不同，但误差在可接受范围内。

需要注意的是，误差可能源于多方面因素的复杂交互影响。例如，天气模式的改变、城市建设和环境基础设施的变化、气象监测设备的质量等等，这些因素都可能对相对湿度产生影响。因此，在进行预测和决策过程中，需要结合更多的因素进行综合分析和评估。

此外，对于预测模型的精度评估，本文介绍了多元线性回归模型的几项重要指标，如调整R平方、DW值、VIF指数等，用于检验模型的拟合程度和自变量之间的共线性等问题。通过这些方法，本文建立的多元线性回归模型表现出较好的预测能力和准确性。然而，需要注意的是，该模型预测结果仅代表一种可能性，真实的情况可能会发生变化。因此，在实际预测和决策过程中，需要对模型的结果进行综合分析，并且结合其他因素和数据进行细致的评估和决策。

## 6.2根据蒙特卡罗方法进行相关预测

**模型摘要**

蒙特卡洛法(Monte Carlo)是根据概率论对随机变量进行概率模拟、统计试验,从而近似求解出预测值的一种方法,又称概率统计法。它是根据现有数据,建立适当的概率模型,通过对模型的抽样试验得出参数的统计特征,最后求出具有特定期望值的近似解。只要能构建出适当的模型,蒙特卡洛方法可以用来模拟任何问题, 包括确定性的数学问题和随机问题两方面,目前广泛应用于工业,农业和社会科学等各个领域。 [1]

其原理为:

(1)假设变量X服从某一概率分布,为未知函数式。

(2)随机抽取自变量带入函数式，求出函数值。

(3)反复抽样多次,计算出函数Y的数据,。当模拟次数足够多时,即可得出函数Y的概率特征。

(4)函数Y的期望值即样本均值,精度的统计估计为样本标准差。

相对湿度序列呈现随机性,根据多年实践经验,本研究选用P-Ⅲ型分布函数,其表达式为:

故若确定出系数Cs,Cv,则可通过计算得到各个参数,从而确定分布函数。 其中 Γ 为标准的 Γ 分布,利用matlab函数x=gaminv(P,A,B),其中A= ，B=1。 通过变量替换,求得不同相对湿度所对应的概率：

本研究利用东北地区沈阳、长春、哈尔滨、大连4个气站9年(2011~2019年)的春季平均相对湿度数据, 将2011~2018年的数据作为建模序列,2019年春季的数据为模型检验序列, 采用Matlab中的x=gaminv (P,A,B)函数,根据相对湿度时间序列特征选取参数值k=6、p=12,应用Matlab软件编程完成蒙特卡洛法预测。[2]

**数据处理**

首先对收集的连续9年的相对湿度数据进行处理得到以下表格：

相对湿度样品值及其相对频率

The precipitation specimens and their relative frequencies

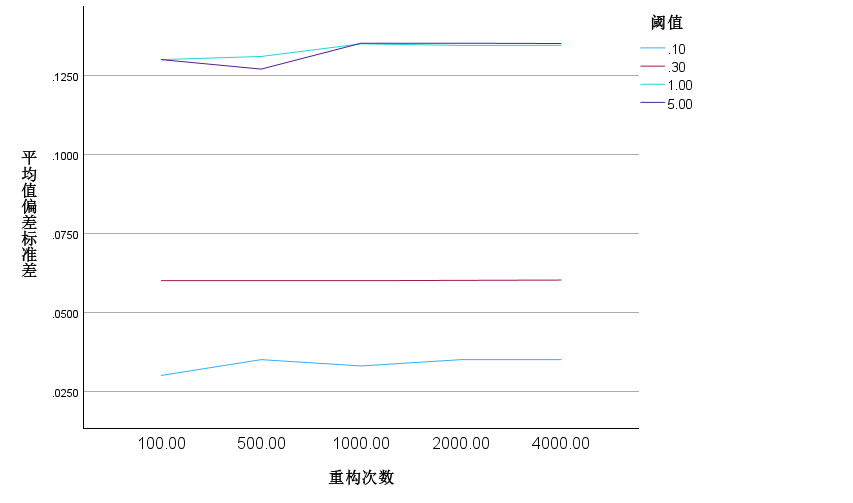
|  |  |  |  |
| --- | --- | --- | --- |
| 等级 | 样品相对概率 | 伪随机数区间 | 区间内产生伪随机数的概率 |
| 1 | 0.0556 | [33, 40] | 0.0556 |
| 2 | 0.0556 | (40, 44] | 0.0556 |
| 3 | 0.0556 | (44, 47] | 0.0556 |
| 4 | 0.0556 | (47, 50] | 0.0556 |
| 5 | 0.0556 | (50, 52] | 0.0556 |
| 6 | 0.1111 | (52, 54] | 0.1111 |
| 7 | 0.1111 | (54, 58] | 0.1111 |
| 8 | 0.1111 | (58, 59] | 0.1111 |
| 9 | 0.1667 | (59, 61] | 0.1667 |
| 10 | 0.2222 | (61, 67] | 0.2222 |
|  |  |  |  |

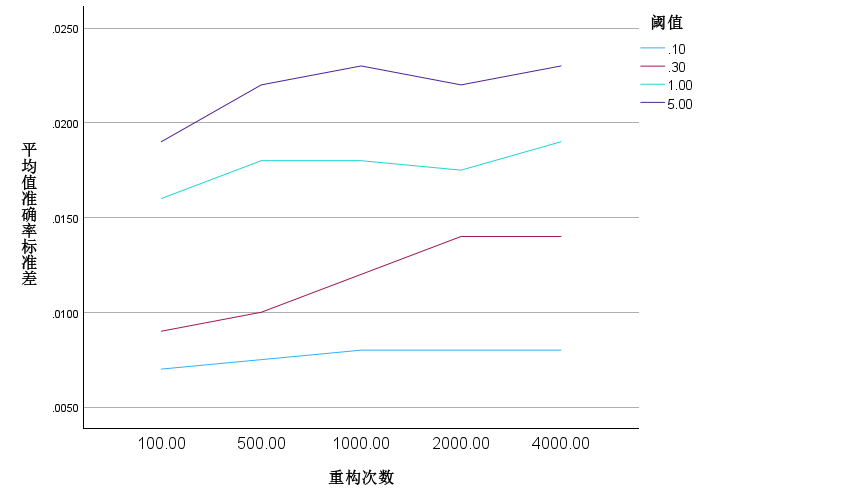
**某样品值的相对频率=该样品值出现的次数/观测的总次数**

根据由上述计算得到的相对频率分布及相对湿度表征了该地区相对湿度所具有的统计规律性,对相对湿度变化的未来模式具有代表性,可以用它们作为模拟预报未来相对湿度的基础。根据蒙特卡罗(Monte-Carlo)方法,先产生在区间**[33,67]**均匀分布的伪随机数。相应伪随机数直接将其作为相对湿度数值, 同时使在此区间产生伪随机数的概率与样本数据中对应数据的出现频率完全相同。根据伯努利定理和实际推断原理，将相应样品出现频率视为相应事件的发生概率。如对于43的这一伪随机数, 对应的伪随机数区间为**(40,44]**,在此区间产生伪随机数的概率为0.0556正好等于相对湿度在**(40,44]**之间的样品在以往资料中出现的概率。由此产生一个规模可观的相对湿度序列模拟蒙特卡洛预测模型的输入。然后将预测的相对湿度值与以往年份的相对湿度平均值相对比,如若相差较大（小于1.0）则重新预报,直至近似相等为止。

**评估分析**

在采用蒙特卡罗方法进行样本重构并预测前,需要进行敏感试验以确定需要进行的样本重构次数.可通过考察不同样本重构次数所获取评估结果的稳定性确认重构次数的选取, 下图给出了不同蒙特卡罗重构次数下不同阈值降水评分的标准差分布,可以看到,100~1000次重构后,评分标准差持续增加, ,而4 000次后评分标准差基本稳定,其中1.0阈值以上准确率评分的标准差稳定在0.018左右,而偏差评分的标准差也基本稳定于0.2以内.因此可采用4 000次蒙特卡罗重构。





为了在足够规模的样本数据上找寻其分布规律，该模型将样本中的地域差异和时间差异忽略，将其视为东北地区春季的相对湿度数据进行分析，通过4000次蒙特卡洛重构最终得到预测结果为50.79，与样本数据平均值50.0528偏差在可接受范围内。

# 7应用和发展

## 7.1概率统计在天气预报领域的潜在应用和未来发展方向

概率统计在天气预报领域有着广泛的潜在应用和未来发展方向。

在不确定性估计方面，概率统计方法可以用于评估天气预报的不确定性。天气预报是基于观测数据和模型输出进行推断，而观测数据存在误差和不完全性，模型的输出也存在不确定性。概率统计方法可以帮助估计这些不确定性，并提供可靠的置信区间和概率分布，使预报人员和用户了解预报的可靠程度。

同时概率统计方法可以用于预测极端天气事件的发生概率。通过分析历史观测数据和气象要素的概率分布，可以评估未来发生极端事件（如暴雨、大风、高温等）的可能性。这有助于制定应对策略和减轻灾害风险。

概率统计方法也可以与人工智能和机器学习技术相结合，以改进天气预报的准确性和效率。通过利用大数据和机器学习算法，可以挖掘气象要素之间的复杂关系，并构建更精确的预测模型。同时，概率统计方法可以帮助评估和解释机器学习模型的不确定性，提高模型的可靠性和解释性。

总而言之，未来概率统计的发展方向包括更精细化的概率预报技术、集合预报的优化和改进、与机器学习和人工智能的深度融合、概率预警系统的发展等。随着数据的不断增加和技术的进步，概率统计方法在生产生活方面将具有愈加重要的应用价值。通过合理应用概率统计方法，它可以帮助我们评估风险、做出决策、改进效率等，提高生产和生活的质量和效益。

# 8参考文献

1. ***赵滨,李子良,张博.蒙特卡罗方法在降水评估中的应用研究[J].南京信息工程大学学报(自然科学版),2016,8(06):553-559***
2. ***韦庆,卢文喜,田竹君.运用蒙特卡罗方法预报年降水量研究[J].干旱区资源与环境,2004(04):144-146.***
3. ***白振鹏,李炎锋,张鹏,王致远,冯姗,樊宪来.隧道火灾温度分布的多元回归预测分析[J].消防科学与技术,2018,37(10)***
4. ***蔡文婷,彭怡,陈秋吉.基于多元回归模型的航空运输客运量预测[J].航空计算技术,2019,49(04)***