“概率论与数理统计”课程大作业

姓名：xxx，学号：1145141919810，序号：44，指导老师：林珍华

题目：关于西安电子科技大学本科生2020年全国数学竞赛非数学专业成绩的统计分析

正文：

（1）采用抽样调查的方式，自定义总体和样本，明确样本值和样本容量

本次统计的总体为西安电子科技大学大二至大四年级参与全国大学生数学竞赛的全体非数学专业组学生的初赛成绩，本次统计采用调查问卷的方式进行抽样调查，通过微信小程序“问卷星”设计问卷，调查问卷如图1所示。



图1 调查问卷样式 1

本次统计的样本为调查问卷所获得的竞赛成绩，西安电子科技大学参加2020全国大学生数学竞赛非数学专业组学生的总人数为2328人，参与调查问卷人数82人，有效结果82个，样本容量n为82。样本值如表1所示。

表1 样本值

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 样本值 | 序号 | 样本值 | 序号 | 样本值 | 序号 | 样本值 |
| 1 | 33 | 22 | 13 | 43 | 23 | 64 | 39 |
| 2 | 30 | 23 | 35 | 44 | 39 | 65 | 52 |
| 3 | 61 | 24 | 65 | 45 | 50 | 66 | 46 |
| 4 | 50 | 25 | 7 | 46 | 24 | 67 | 30 |
| 5 | 26 | 26 | 20 | 47 | 35 | 68 | 31 |
| 6 | 51 | 27 | 19 | 48 | 38 | 69 | 31 |
| 7 | 52 | 28 | 56 | 49 | 42 | 70 | 32 |
| 8 | 33 | 29 | 50 | 50 | 19 | 71 | 17 |
| 9 | 38 | 30 | 39 | 51 | 36 | 72 | 24 |
| 10 | 24 | 31 | 38 | 52 | 41 | 73 | 35 |
| 11 | 32 | 32 | 19 | 53 | 44 | 74 | 20 |
| 12 | 33 | 33 | 51 | 54 | 29 | 75 | 53 |
| 13 | 38 | 34 | 41 | 55 | 7 | 76 | 42 |
| 14 | 50 | 35 | 49 | 56 | 50 | 77 | 31 |
| 15 | 26 | 36 | 38 | 57 | 31 | 78 | 25 |
| 16 | 24 | 37 | 47 | 58 | 19 | 79 | 60 |
| 17 | 39 | 38 | 59 | 59 | 37 | 80 | 15 |
| 18 | 42 | 39 | 46 | 60 | 16 | 81 | 44 |
| 19 | 40 | 40 | 46 | 61 | 23 | 82 | 24 |
| 20 | 1 | 41 | 15 | 62 | 25 |  |  |
| 21 | 24 | 42 | 33 | 63 | 44 |  |  |

每隔十分建立样本数值区间，对该区间样本频数进行统计，并计算样本频率。经分析，样本频数与频率如表2所示。

表2 样本频数与频率

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 样本数值区间 | 1-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 | 61-70 | 样本总体 |
| 频数 | 3 | 11 | 15 | 25 | 18 | 8 | 2 | 82 |
| 频率 | 0.0365 | 0.1375 | 0.1829 | 0.3049 | 0.2195 | 0.0976 | 0.0244 | 1 |

（2）计算样本均值和样本方差

经分析得，样本最小值为1，样本最大值为65，样本中位数为35，样本重数（频数最大的数）为24，样本均值 ，样本方差，样本均方差。

（3）对样本分布参数进行点估计（矩估计或最大似然估计）和区间估计

假设竞赛成绩为随机变量X，且X服从正态分布，那么参数有和。

若采用矩估计，参数和的矩估计分别为

若采用最大似然估计，参数和的最大似然估计分别为

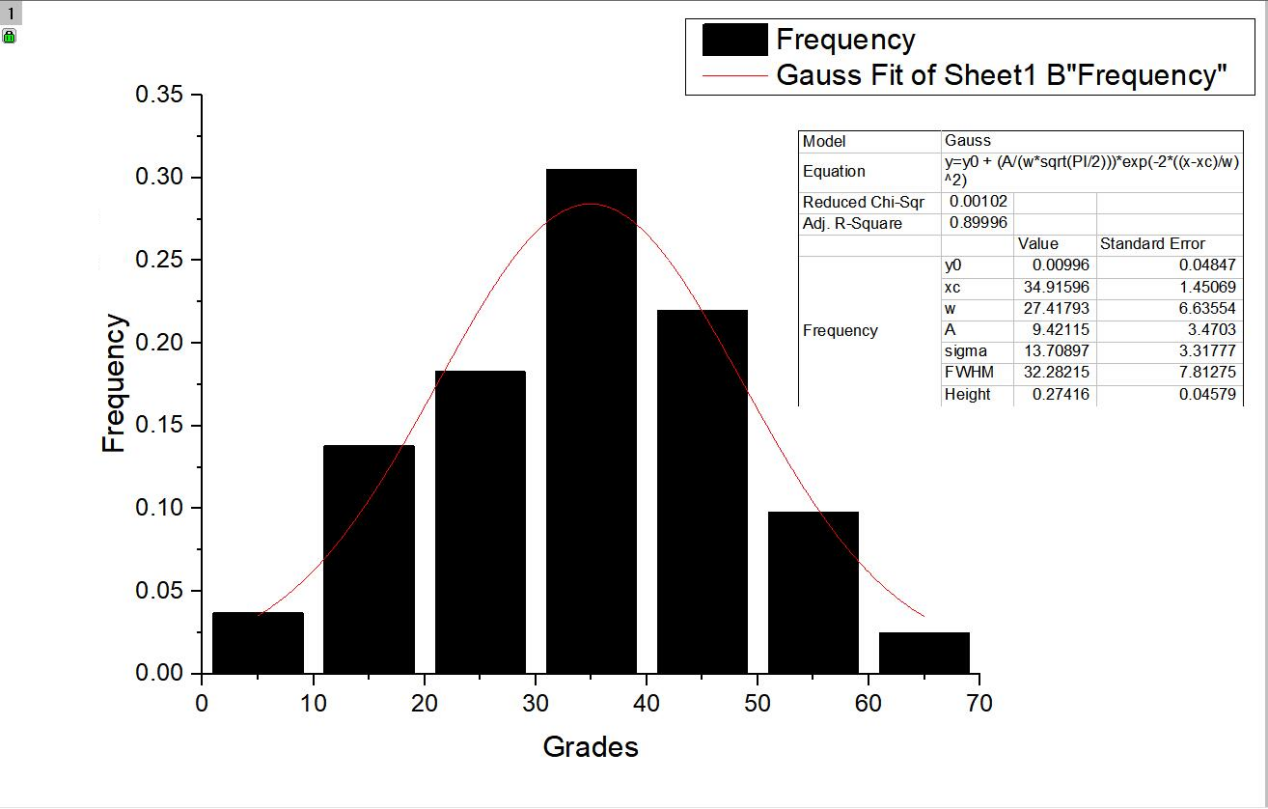


图2 拟合正态分布曲线 1

图2 采用Origin软件进行Guass拟合，因数据取值过多，样本量较少，故取样本区间中点作为取样点，样本区间频率作为取样点频率进行曲线拟合，横轴为样本值，纵轴为频率。

若采用Origin软件采用最小二乘法进行Gauss拟合（如图2），拟合方程为y=y0 + (A/(w\*sqrt(PI/2)))\*exp(-2\*((x-xc)/w)^2)，则拟合得到的参数和分别为

若采用区间估计，则参数和的置信水平为0.99的置信区间分别为(30.9175, 38.7411)和(124.2803, 281.0116)，计算过程如下：

（4）自定义假设，并进行假设检验

若全校大二至大四年级参与全国大学生数学竞赛的全体非数学专业组学生的初赛成绩服从正态分布，试问在显著性水平α=0.05下，可否认为此次抽样统计的样本均值和样本方差是合理的？

答：是合理的。

对于样本均值：

（I）需要检验

（II）选择统计量

（III）由于α=0.05，因此临界点为，从而接受域为(-1.96, 1.96)；

（IV）由于n=82，σ=180，，因此检验统计量U的样本值为

（V）由于u=-0.0086∈(-1.96, 1.96)，因此接受H0，即认为此次抽样统计的样本均值合理。

对于样本方差：

（I）需要检验

（II）选择统计量

（III）由于α=0.05，因此临界点为和，从而接受域为(58.0025, 107.7797)；

（IV）由于n=82，μ=35，，因此检验统计量U的样本值为

（V）由于u=82.1556∈(58.0025, 107.7797)，因此接受H0，即认为此次抽样统计的样本方差合理。

（5）小结

本次统计研究了西安电子科技大学本科生2020年全国数学竞赛非数学专业成绩的分布情况。采用抽样调查的方式，获得82个样本值，计算了样本均值和样本方差，并假设样本值服从正态分布，对期望和方差进行参数估计和区间估计，在进行曲线拟合过程中，采用样本取值区间的中点以及区间概率值进行拟合，还自定义了西安电子科技大学本科生2020年全国数学竞赛非数学专业成绩的分布，并对本次抽样调查进行样本均值和样本方差合理性的检验。