

C 程序设计成绩与编程时间关系分析研究

罗 强, 铁治欣

(浙江理工大学信息学院, 杭州 310018)

摘 要: 普遍认为编程练习时间的长短直接影响着 C 语言课程的学习效果。笔者收集了一个教学班中所有学生的编程练习时使用计算机的时间记录及他们的期末与省二级考试的成绩。首先通过 χ^2 拟合度检验法验证这个教学班的这两次考试成绩服从正态分布。然后, 按照 C 语言教学大纲中规定的最低要求课外上机时间, 将这个教学班的学生分为两组, 第一组为课外上机时间超过或等于教学大纲规定的最低课外上机时间的学生, 第二组为课外上机时间少于教学大纲规定的最低课外上机时间的学生。用 T 检验对这两组学生的期末与省二级考试的成绩进行分析, 结果表明: 第一组学生的期末考试及省二级考试的平均成绩分别比第二组至少高 9.0 分与 7.0 分; 第一组学生的期末考试及省二级考试的通过率分别比第二组高 18.75% 及 19.54%。因此, 认为上机练习对学生学习 C 语言具有较大的促进作用。

关键词: C 语言; 学生成绩; 假设检验; 上机时间; 作业

中图分类号: G642 **文献标志码:** A

0 引 言

C 语言是在工业界应用最广泛的程序设计语言之一^[1-2], 也是目前最普及的程序设计教学语言之一^[3-7], 文献[7]中列出了学习 C 语言的 10 个理由。为了学习 C 语言, 学生需要有一定时间的上机编程练习。普遍认为上机编程时间的长短直接影响着 C 语言课程的学习效果。究竟上机编程时间对 C 语言课程学习有多大的促进作用, 目前相关的文献报道较少。文献[8-9]提出了一些 C 语言课程教学实验系统 and 教学改革方法, 文献[10-12]对教学中的数据挖掘方法进行了综述。

笔者收集了一个教学班中所有学生的上机编程时间记录及他们的期末与省二级考试的成绩。首先通过 χ^2 检验法验证这个教学班的这两次考试成绩服从正态分布, 这也是通常我们所期望的结果。然后, 按照 C 语言教学大纲中规定的最低课外上机编程时间, 将这个教学班的学生分为两组, 第一组为课

外上机编程时间超过或等于教学大纲规定的最低上机编程时间的学生, 第二组为课外上机编程时间少于教学大纲规定的最低上机编程时间的学生。用 T 检验对这两组学生的期末与省二级考试的成绩进行分析。

1 假设检验方法概述

1.1 χ^2 拟合度检验

χ^2 拟合度检验方法用来检验一组采样数据是否服从一个特定的分布^[13]。给出一组采样样本数据 X_1, X_2, \dots, X_n , χ^2 拟合度检验就是要测试如下假设检验:

H_0 : 这组采样数据服从某个特定分布。

H_a : 这组采样数据不服从这个特定分布。

检验统计量: 为了计算 χ^2 拟合度检验的统计量, 需要将样本分成 k 个区间, 统计量的定义为:

$$\chi^2 = \sum_{i=1}^k (O_i - E_i)^2 / E_i \quad (1)$$

式(1)中 O_i 是样本数据落在第 i ($1 < i < k$) 个区间的频数; E_i 是样本数据落在第 i ($1 < i < k$) 个区间的期望频数, 它可以用下式计算:

$$E_i = N(F(Y_u) - F(Y_l)) \quad (2)$$

式(2)中 F 是要检验的样本数据所服从的分布的累积分布函数; Y_u 是区间 i ($1 < i < k$) 的上限, Y_l 是区间 i ($1 < i < k$) 的下限; N 是样本数据的个数。

检验统计量 χ^2 近似地服从自由度为 $k-r-1$ 的 χ^2 分布, 其中 k 是区间的个数, r 为被估计参数的个数。

因此, 当下列条件成立时, 拒绝原假设 H_0 , 即采样样本数据不服从检验的特定分布; 否则就接受原假设 H_0 , 即认为采样样本数据服从检验的特定分布。

$$\chi^2 > \chi_{(\alpha, k-r-1)}^2 \quad (3)$$

式(3)中 $\chi_{(\alpha, k-r-1)}^2$ 为自由度为 $k-r-1$ 的 χ^2 分布在显著水平为 α 时的临界值。

1.2 双样本 T 检验^[13-14]

给定两组随机采样样本数据, Y_1, Y_2, \dots, Y_{n_1} 和 Z_1, Z_2, \dots, Z_{n_2} , 它们的样本数量分别为 n_1 和 n_2 , 假定它们分别抽取自均值为 μ_1 和 μ_2 , 方差为 σ_1^2 和 σ_2^2 两个总体。

可以用下面的假设检验来验证这两个总体是否具有方差齐性。

H_0 : 两个总体的方差 σ_1^2 与 σ_2^2 具有方差齐性。

H_a : 两个总体的方差 σ_1^2 与 σ_2^2 不具有方差齐性。

检验统计量为:

$$F = S_1^2 / S_2^2 \quad (4)$$

式(4)中:

$$S_1 = \sqrt{\frac{1}{n_1} \sum_{i=1}^{n_1} (Y_i - \bar{Y})^2},$$

$$S_2 = \sqrt{\frac{1}{n_2} \sum_{i=1}^{n_2} (Z_i - \bar{Z})^2}$$

检验统计量 F 服从自由度为 $v_1 = n_1 - 1$ 与 $v_2 = n_2 - 1$ 的 F 分布 $F(v_1, v_2)$ 。因此, 当下述条件成立时, 就拒绝原假设 H_0 , 即认为两个总体的方差 σ_1^2 与 σ_2^2 不具有方差齐性; 否则就接受原假设 H_0 , 即认为两个总体的方差 σ_1^2 与 σ_2^2 具有方差齐性。

$$F \geq F_{\alpha/2, v_1, v_2} \text{ 或 } F \leq F_{1-\alpha/2, v_1, v_2}$$

这里 $F_{\alpha/2, v_1, v_2}, F_{1-\alpha/2, v_1, v_2}$ 是自由度为 $v_1 = n_1 - 1$ 和 $v_2 = n_2 - 1$ 的 F 分布 $F(v_1, v_2)$ 在显著水平为 α 时的临界值。

如果两个总体的方差 σ_1^2 与 σ_2^2 具有方差齐性, 则可以利用下述的双样本 T 检验来判定这两个总体的均值 μ_1 和 μ_2 的关系。

$$H_0: \mu_1 - \mu_2 \geq \delta$$

$$H_a: \mu_1 - \mu_2 < \delta$$

检验统计量为:

$$t = \frac{(\bar{Y} - \bar{Z}) - \delta}{S \sqrt{1/n_1 + 1/n_2}} \quad (5)$$

式(5)中:

$$S = \sqrt{\frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}},$$

$$\bar{Y} = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i, \bar{Z} = \frac{1}{n_2} \sum_{i=1}^{n_2} Z_i$$

检验统计量 t 服从自由度为 $v = n_1 + n_2 - 2$ 的 t 分布。因此, 当下列条件成立时, 拒绝原假设 H_0 , 否则接受原假设 H_0 。

$$t \leq -t_{\alpha, v}$$

这里 $t_{\alpha, v}$ 是自由度为 $v = n_1 + n_2 - 2$ 的 t 分布在显著水平为 α 时的临界值。

2 数据收集

按照 2008 年修订的 C 语言课程的教学大纲的要求, C 语言课程的教学活动包括 16 次理论讲解课与 16 次实验课。每次理论讲解课为 2 学时; 每次实验课为 3 学时。每次实验课后老师会布置一些课后作业, 要求学生用不少于 3 学时的时间来完成每次的课后作业。由于作业系统安装在计算中心, 一般情况下, 学生会在计算中心的公共机房中完成这些作业。

为了分析学生成绩与上机编程时间之间的关系, 选择了一个教学班来进行本次分析研究。这个教学班有 84 名学生。这个教学班的学生的期末成绩与省二级考试成绩分别列在表 1。将学生成绩 X 分为 5 个区间: 区间 A ($X \geq 90$), 区间 B ($80 \leq X < 90$), 区间 C ($70 \leq X < 80$), 区间 D ($60 \leq X < 70$), 区间 F ($X < 60$) 进行统计, 得到学生期末成绩与省二级考试成绩的分区间统计结果如图 1 和图 2 所示。

表 1 该教学班的学生成绩与上机编程时间

学生编号	期末成绩	省二级 考试成绩	课后上机 编程时间/min
1	88	62	1 170
2	87	80	2 566
...
84

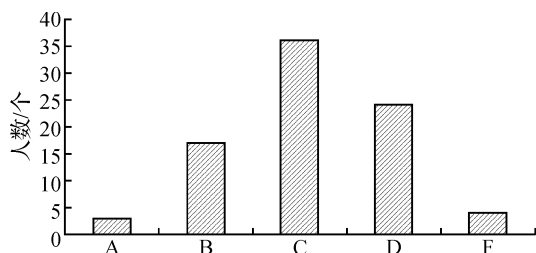


图1 学生期末成绩分区间统计

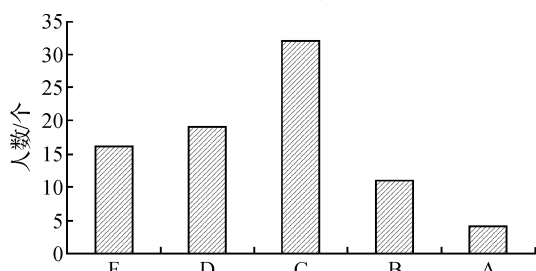


图2 学生的省二级考试成绩分区间统计

学生上机编程时间可以分为两个部分:第一部分为上实验课时的上机编程时间;第二部分为课后学生做课后作业时的上机编程时间,将这部分时间称作课后上机编程时间。前者对所有学生都一样的,所以在分析时只用了课后上机编程时间。

一般情况下,本校学生在本校的计算机中心的公共机房中完成他们的课后作业。因此某个学生的课后编程时间可以从我们的自己开发的作业系统及机房管理系统中抽取出来。作业系统记录了每个学生的包括交作业时间在内的交作业记录。机房管理系统记录了每台计算机的使用记录,包括使用者、使用开始时间与结束时间系统信息。每个学生的某次课后编程时间是从某次实验课结束到他(她)交这次作业为止这段时间内,他(她)在公共机房中使用计算机的时间的总和。某个学生的总的课后编程时间是他(她)在公共机房中完成其16次课后作业时使用计算机的时间的总和。每个学生的总的课后编程时间如表1所示。

3 学生成绩与课后编程时间关系的分析

由于学生在学习C语言课程时的努力程度各不相同,所以他们得到的分数也各不相同,所以可以认为学生期末成绩与省二级考试成绩为随机变量,因此可以利用上述讨论过的假设检验方法来分析学生成绩与课后编程时间关系。首先,利用 χ^2 拟合度检验来检验学生期末成绩与省二级考试成绩是否服从正态分布。然后,利用双样本T检验来检验花更多课后上机编程时间的学生比花较少课后上机编程时间的学生得到更好的成绩。

3.1 学生成绩分布的假设检验

3.1.1 学生期末成绩服从分布的假设检验

要检验的假设为:

H_0 : 学生期末成绩 $X \sim N(\mu, \sigma^2)$ 。

H_a : 学生期末成绩 X 不服从 $N(\mu, \sigma^2)$ 。

这里 μ, σ^2 是未知数,分别表示学生成绩的均值与方差。

将学生成绩 X 分为5个区间:区间A($X \geq 90$), B($80 \leq X < 90$), C($70 \leq X < 80$), D($60 \leq X < 70$), F($X < 60$)。检验统计量的计算如表2所示。在本次检验中, $n=84, k=4, r=2$, 显著水平取 $\alpha=0.01$, 因此,若原假设 H_0 为真,检验统计量近似地服从自由度为 $(k-r-1)=1$ 的 χ^2 分布。

如表2所示,检验统计量

$$\chi^2 = 1.630 < \chi_{0.01}^2(1) = 6.63$$

因此,接受原假设,即学生期末成绩 X 在显著水平 $\alpha=0.01$ 下服从正态分布 $N(73.47, 9.70^2)$ 。

3.1.2 省二级考试成绩服从分布的假设检验

应用前述同样的方法,可以对省二级考试成绩服从分布进行假设检验。检验统计量的计算如表3所示。

$$\chi^2 = 3.689 < \chi_{0.01}^2(1) = 6.63$$

因此,接受原假设,即学生的省二级考试成绩 X 在显著水平 $\alpha=0.01$ 下服从正态分布 $N(69.62, 11.32^2)$ 。

表2 学生期末成绩的检验统计量的计算

序号	区间	O_i	$F(Y_u) - F(Y_l)$	E_i	$(O_i - E_i)^2 / E_i$
1	A or B	20	0.250 4	21.034 2	0.051
2	C	36	0.389 4	32.709 1	0.331
3	D	24	0.277 9	23.342 0	0.019
4	F	4	0.082 3	6.914 7	1.229
合计	/	84	/	/	1.630

表3 学生省二级考试成绩的检验统计量的计算

序号	区间	O_i	$F(Y_u) - F(Y_l)$	E_i	$(O_i - E_i)^2 / E_i$
1	A or B	15	0.179 6	14.726 1	0.005
2	C	32	0.307 1	25.178 3	1.848
3	D	19	0.315 7	25.888 1	1.833
4	F	16	0.197 7	16.207 5	0.003
合计	/	82	/	/	3.689

3.2 课后编程时间对学生成绩的影响

一般认为,在学习C语言课程时,如果学生花费更多的时间进行上机编程练习,他将会得到更好的成绩。这里我们用双样本T检验来验证这个命题的真实性。

按照 2008 年修订的 C 语言课程的教学大纲的意见,要求学生在做每次课后作业时应该花费不少于 3 学时的上机编程时间,因此每个学生的总的课后上机编程时间应该不少于 48 学时。但实际情况是有一些学生的课后上机编程时间远少于这个指导课后上机编程时间,有一些学生的课后上机编程时间远多于这个指导课后上机编程时间。根据这个指导课后上机编程时间与学生实际上机编程时间的长短将学生分为两个组:第 I 组为课外上机编程时间超过或等于教学大纲规定的最低时间的学生,第 II 组为课外上机时间少于教学大纲规定的最低时间的学生。基于这个分组方法,本文所检验的这个教学班被分入第 I 组与第 II 组的人数分别为 68 人和 16 人。下面利用这个教学班中学生的期末考试成绩与省二级考试成绩来验证是不是每 I 组学生会比第 II 组学生得到更好的成绩。

3.2.1 两组学生考试通过率的比较

某组学生的考试通过率,就是在某次考试中该组学生的考试成绩为及格及以上的学生人数占该组总人数的百分比。两组学生的期末考试与省二级考试的考试通过率如表 4 所示。从表 4 中可以知道,第 I 组学生的期末考试与省二级考试的考试通过率分别比第 II 组学生高 18.75 与 19.54 个百分点。

表 4 两组学生期末考与省二级考的通过率

组别	期末考试通过率/%	省二级考试通过率/%
I	100.00	83.82
II	81.25	64.29

3.2.2 两组学生期末成绩均值间关系的假设检验

为了用双样本 T 检验来检验两组学生期末成绩均值间关系,首先必须检验这两组学生期末成绩的方差具有方差齐性。这里分别用 μ_I 和 μ_{II} , σ_I^2 和 σ_{II}^2 来表示两组学生成绩的均值与方差。首先来检验两组学生期末成绩的方差具有方差齐性,假设为:

H_0 : 两组学生期末成绩的方差 σ_I^2 和 σ_{II}^2 具有方差齐性。

H_a : 两组学生期末成绩的方差 σ_I^2 和 σ_{II}^2 不具有方差齐性。

检验统计量的计算如表 5 所示。检验统计量 F 为:

$$F=0.324$$

由于 $F_{0.01/2,67,15}=3.26$ 和 $F_{1-0.01/2,67,15}=0.307$, 因此,在显著水平 $\alpha=0.01$ 下,接受原假设,即认为两组学生期末成绩的方差 σ_I^2 和 σ_{II}^2 具有方差齐性。

由于两组学生期末成绩的方差 σ_I^2 和 σ_{II}^2 具有方

差齐性,所以可以用双样本 T 检验来检验两组学生期末成绩均值间关系,做如下假设:

$$H_0: \mu_I - \mu_{II} \geq 9.0$$

$$H_a: \mu_I - \mu_{II} < 9.0$$

检验统计量的计算如表 5 所示,检验统计量 t 为:

$$t=0.288 > -t_{0.01,82} = -2.375$$

因此,在显著水平 $\alpha=0.01$ 下,接受原假设,即认为第 I 组学生的期末成绩的均值比第 II 组大于或等于 9.0 分。

表 5 两组学生期末成绩均值间关系的假设检验相关检验统计量计算

属性	第 I 组	第 II 组
均值	75.412	65.688
标准差	7.710	13.549
样本数量	68	16
自由度	67	15
检验统计量 F	0.324	/
检验统计量 t	0.288	/

3.2.3 两组学生省二级考试成绩均值间关系的假设检验

由于本教学班中有两位同学没有参加省二级考试,所以在本次检验中,样本总数为 82。

运用与前一小节同样的方法,可以对两组学生省二级考试成绩均值间关系进行假设检验。相关的检验统计量计算如表 6 所示。

从表 6 可以知道,检验统计量 F 为:

$$F=1.215$$

由于 $F_{0.01/2,67,13}=3.65$, $F_{1-0.01/2,67,13}=0.274$, 在显著水平 $\alpha=0.01$ 下,接受原假设 H_0 ,即认为两组学生期末成绩的方差 σ_I^2 和 σ_{II}^2 具有方差齐性。

可以用以下假设检验来检验两组学生省二级考试成绩均值间关系。

$$H_0: \mu_I - \mu_{II} \geq 7.0$$

$$H_a: \mu_I - \mu_{II} < 7.0$$

表 6 两组学生省二级考试成绩均值间关系的假设检验相关检验统计量计算

属性	第 I 组	第 II 组
均值	70.853	63.643
标准差	11.215	10.172
样本数量	68	14
自由度	67	13
检验统计量 F	1.215	/
检验统计量 t	0.065	/

从表 6 中可以得到,检验统计量 t 为:

$$t=0.201 > -t_{0.01,80} = -2.373$$

因此,在显著水平 $\alpha=0.01$ 下,接受原假设 H_0 ,即认为第I组学生的省二级考试成绩的均值比第II组大于或等于7.0分。

4 结束语

根据一个C语言课程教学班中学生做课后作业时使用计算机的编程时间多少将学生分成两组:第一组由做课外作业时使用计算机的编程时间多于或等于C语言教学大纲规定的最低时间的同学组成,第二组由做课外作业时使用计算机的编程时间少于C语言教学大纲规定的最低时间的同学组成。使用假设检验的方法对这两组学生的期末考试与省二级考试的成绩进行分析,发现第一组学生的期末考试及省二级考试的平均成绩分别比第二组至少高9.0分与7.0分;第一组学生的期末考试及省二级考试的通过率分别比第二组高18.75及19.54个百分点。所以认为上机编程时间对学生学习C语言具有很大的促进作用。因此,在C语言教学中,除了课堂上的理论与实验教学外,应当适当引导学生通过上机编程练习来完成适当的课外作业。

参考文献:

- [1] DedaSys LLC. Programming Language Popularity[EB/OL]. [2012-10-03]. <http://www.langpop.com/>.
- [2] TIOBE Software. TIOBE Programming Community Index for September 2012[EB/OL]. [2012-10-03]. <http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html>
- [3] Cheng H H. C for the course[J]. Mechanical Engineering Magazine, 2009, (9): 50-52.
- [4] 浙江大学《C语言程序设计》课程组. 国家级精品课程: C语言程序设计[EB/OL]. [2012-10-03]. <http://jpkc.zju.edu.cn/k/409/ml15.htm>.
- [5] 哈尔滨工业大学《C语言程序设计》课程组. 国家级精品课程、教育部-微软精品课程、黑龙江省精品课程、国家教育部双语教学示范课程: C语言程序设计[EB/OL]. [2012-10-03]. <http://cms.hit.edu.cn/elite/>
- [6] 浙江理工大学《C语言程序设计》课程组. 国家双语教学示范课程建设项目: C语言程序设计[EB/OL]. [2012-10-03]. <http://ecs.zstu.edu.cn/ce/KCGK/kcgk.htm>
- [7] Cheng H H. Ten Reasons to Teach and Learn Computer Programming in C[EB/OL]. [2012-10-03]. <http://iel.ucdavis.edu/publication/WhyC.html>
- [8] 丁海燕, 邹疆, 邱莎. C语言实验教学体系及改革措施[J]. 实验技术与管理, 2010, 27(11): 179-181.
- [9] 陈婷. C语言程序设计实验教学改革探究[J]. 实验技术与管理, 2010, 27(10): 182-184.
- [10] Romero C, Ventura S. Educational data mining: a survey from 1995 to 2005[J]. Expert Systems with Applications, 2007, 33(1): 135-146.
- [11] Baker R, Yacef K. The state of educational data mining in 2009: a review and future visions[J]. J Educ Data Mining, 2009, 1(1): 3-17.
- [12] Crist'obal Romero, Sebastian Ventura. Educational data mining: a review of the state of the art[J]. IEEE Transactions on Systems, Man, and Cybernetics—part c: Applications and Reviews, 2010, 40(6): 601-618.
- [13] NIST/SEMATECH. Engineering Statistics Handbook[EB/OL]. [2012-10-03]. <http://www.itl.nist.gov/div898/handbook>
- [14] Walpd R E, Myers R H, Sharon L M, et al. Probability & Statistics for Engineers & Scientists[M]. Upper Saddle River: Prentice Hall, 2002.

Analysis on the Relationship between Students' Scores and Programming Time in Learning the C Programming Language

LUO Qiang, TIE Zhi-xin

(School of Informatics, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Generally speaking, the computer programming time that a student spent on his homework directly affects his score of the course of the C Programming Language. This thesis collects the computer programming time on homework, final scores and scores of the Zhejiang Computer Score Test Band II (ZC-STBII) of students from a teaching class. Firstly, it uses the Chi-Square Goodness-of-Fit Test to verify those final scores and ZCGTBII scores of this teaching class subject to the normal distribution. Then, it divides students of this teaching class into two groups based on students' computer programming time on homework and minimum computer programming time on homework that the C programming language syl-

labus prescribed for a student to learn this course. Then, the authors employ the Two-Sample Pooled T-Test to analyze the final scores and the ZCSTBII scores of the two groups of students. Analysis results show that the mean of the final scores and the mean of ZCSTBII scores of Group I are ZCSTBII those of Group II by 9.0 and 7.0 points, respectively, and the passing rates of final test and ZCGTBII of Group I are greater than those of Group II by 18.75 and 19.54 percentage points, respectively. Therefore, the authors conclude that the amount of time students spend practicing programming on computers plays a very important role in mastering the C programming language.

Key words: C programming language; student score; hypothesis test; computer programming time; homework

(责任编辑: 陈和榜)

(上接第 75 页)

Experimental Investigation on Optimizing the Effect of Particles' Acceleration Driven by Shock Waves

ZHANG Ping, QI Hong-xun, ZHANG Li-te, SHI Hong-hui, HUANG Bao-qian

(School of Machinery and Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: This research designs three different types of accelerating nozzles and machines these nozzles behind the horizontal shock tube. The aim behind the concept is to accelerate subsonic, transonic and supersonic flow respectively so that they can optimize the effect of particles' acceleration. Thus, this research uses high-speed photography to capture the image of solid particles group moving in tube. Meanwhile, it utilizes dynamic pressure measurement to measure and calculate experimental shock wave velocity and Mach number along the axial position. Based on the experimental results, this thesis gives the conclusion. Under the same condition, the smaller the particle's diameter is, the more unobvious the wake phenomenon of particles group movement will be. For different types of nozzle structures, it is necessary to choose appropriate nozzles according to the Mach number of the gas flow in order that they can optimize the effect of particles' acceleration.

Key words: shock wave; particle; acceleration; nozzle

(责任编辑: 朱松英)