

# C 语言教程讲义

Potato

2024 年 6 月 23 日

## 0 计算机基础知识与计算机语言概述

### 1 标准输出和变量

#### 1.1 printf 函数

打开 visual studio，在“源文件”上右键，点击“添加”->“新建项”，创建一个叫 main.c 的文件，在其中输入

```
1  #include <stdio.h>
2
3  int main() {
4      printf("Hello World!");
5
6      return 0;
7  }
```

点击“本地 Windows 调试器”，将会出现类似于这样的界面



恭喜你写出了你的第一段代码。计算机领域有一个传统：刚入门编程的人的第一段代码一定要是 Hello World，这是我们对无垠的计算机世界打的招呼。这个黑乎乎的界面叫做终端（这里不做关于终端、shell 等的详细区分），我们以后会常常和它打交道。

我们发现，我们写在 printf 的引号内的内容会被打印出来（顾名思义，printf 就是 print function 嘛），因此我们可以多写几个 printf（不要忘了每行最后的英文分号），如

```
1  #include <stdio.h>
2
3  int main() {
4      printf("Hello World!");
5      printf("Potato");
6
7      return 0;
8  }
```

运行代码，我们观察到 Hello World 和 Potato 输出在了同一行。我们想要的是这二者各自占一行，那我们就需要在 Hello World!后加一个换行符，也即 \n，代码就变成了

```
1  #include <stdio.h>
2
3  int main() {
4      printf("Hello World!\n");
5      printf("Potato\n");
6
7      return 0;
8  }
```

这次的结果就是我们想要的了。为什么 Potato 也要加换行符？这是因为也许我们在其后还会添加新的 printf，要是到时候忘了给 Potato 加换行符，不就没有实现我们想要的效果吗？为了避免以后出现不必要的问题，我们推荐每个 printf 后都加换行符。

由上面的代码，我们总结出了这些要点：首先，我们要想输出信息，就要将其写在 printf 的引号内；其次，要想换行，就需要在输出的文本最后加一个换行符 \n；另外我们还注意到，我们先输出了 Hello World，再输出了 Potato，也就是说，程序是按照从上到下的顺序执行的。

这是很重要的结论，所以我们使用 printf 输出一遍加强印象吧。于是我们写下了：

```
1 printf("首先,我们要想输出信息,就要将其写在printf的引号内;其次,要想换行,就需要在输出的文本最后加一个换行符\n;\n");
```

我们注意到有个问题，就是第一个换行符是我们想要按照文本形式输出的，但是如果按照上面的写法，它会被用于换行而不是直接输出。为了解决这个问题，我们把代码改成这样即可：

```
1 printf("首先,我们要想输出信息,就要将其写在printf的引号内;其次,要想换行,就需要在输出的文本最后加一个换行符\\n;\n");
```

运行结果是：

A screenshot of the Microsoft Visual Studio debugger window. The output window shows the following text: "Hello World!" followed by "Potato" on a new line. Then, the full line of code from the previous block is printed: "首先, 我们要想输出信息, 就要将其写在printf的引号内; 其次, 要想换行, 就需要在输出的文本最后加一个换行符\n;". Below this, a message indicates the program has exited: "D:\csample\x64\Debug\csample.exe (进程 10500)已退出, 代码为 0." and instructions about the console window.

为什么会这样呢？实际上，我们输入的 \n 和 \\ 都叫做转义符号，用来实现一些特殊的输出效果。它的语法就是反斜杠（\）加一个字母。我们最常用的转义符号就是换行符 \n，实际上还有 \t、\a 等等。因为反斜杠会和下一个字母结合变成转义符号，所以要想输出反斜杠，就需要按照上面的例子那样使用两个反斜杠，这个转义符号的结果就是输出一个反斜杠。

## 1.2 变量

变量顾名思义，就是可以变的量。一个变量肯定要有名字，然后要有一个值。这就像是我们的考试成绩，有一个叫做“数学成绩”的变量，这次可以是 120，下次可以是 130。在 c 语言中，我们这样声明变量：*type name*，啥意思呢，比如我们要声明数学成绩（math\_grade），那么就可以这样写

```
1 int math_grade;
```

这里的 int 是整数的意思，还有其它类型（如小数）我们之后会讲到。所以这行代码的意思就是，我们声明了一个叫做 math\_grade 的变量，它是一个整数。接下来就要给它赋值了，在大部分计算机语言中，赋值号都是=，比如数学成绩是 120，那么就可以写

```
1 math_grade = 120;
```

这段代码的意思就是给 math\_grade 赋予 120 这个值。也即，我们将赋值号右边的值给了左边的变量。我们强调是将右边的值赋给左边，说明了这不是数学中的等号，它不意味着左右两边相等，也不能交换左右两边的位置。

现在我们的 main.c 应该是这样的

```
1 #include <stdio.h>
2
3 int main() {
4     int math_grade;
5     math_grade = 120;
6
7     return 0;
8 }
```

第四行是声明变量，第五行是给变量赋值。如果我们在声明变量时就知道它要赋什么值，那么就可以把声明和赋值写在一起

```
1 int math_grade = 120;
```

如果要同时声明多个类型一样的变量，那么可以写在一行，比如同时声明数学、语文和英语成绩：

```
1 int math_grade, chinese_grade, english_grade;
```

不过笔者推荐每个变量单独成行，因为在最开始写代码时，我们可能由于考虑不周会选错变量类型（比如英语成绩会出现 0.5 分的情况，那么选择整数就不合适了），如果写在同一行会增大修改的难度，因此推荐每个变量分开声明：

```
1 int math_grade;
2 int chinese_grade;
3 int english_grade;
```

那么变量的命名遵循怎样的原则呢？我们这里的命名是 math\_grade，其实也可以命名成 MathGrade，也可以是 mathGrade，可以是拼音 shuxuechengji，甚至可以是汉语的数学成绩，如果你比较懒，直接命名

l 这样的单字母都可以。c 语言要求，变量名开头不是数字、不包含空格、标点和%、@等符号，不是 c 语言关键字（后面会介绍这个概念）即可。那么我们为什么使用了 `math_grade` 呢？

首先我们说明为什么不用汉语或者汉语拼音。一方面，由于编码不同，在你的电脑上正常显示的中文可能在别人电脑上就变成了乱码（这就是第零章要求大家切换 UTF8 编码的原因），这样严重影响了正常的开发和交流工作。另一方面，大家有没有发现，当你在声明好变量后，你敲下“m”时就会出现一个框，可以直接补全 `math_grade`，这是 IDE 的自动补全功能，如果使用中文就无法自动补全了。zhi yu wei shen mo bu yong pin yin, ge wei jue de zhe yang de ke du xing gao ma?

计算机语言的核心在于其逻辑而不是用了什么自然语言作为支撑，哪怕世界上第一门计算机语言是中国人发明的，计算机语言也会愈发趋向于使用字母来构成，因为这样效率更高。这样大家就可以明白那些被大肆宣扬的“汉语编程语言”的荒谬之处了，不要陷入了虚伪的爱国主义的陷阱。

我们命名的 `math_grade` 有个特点：单词全部小写，单词之间使用下划线\_连接。这种命名方式叫做蛇形命名法 (`snake_case`)。至于我们上面提到的 `mathGrade` 也是一种命名法，叫做小驼峰命名法 (`camelCase`)。命名法和一些其它规则构成了代码风格 (`code style`)，本讲义遵循 Google 代码规范，也推荐各位使用这套代码规范。各位在阅读此讲义时应该注意示例代码的代码风格，自己写代码时要模仿示例代码的风格。不过，如果各位是参与到他人的项目，自然是要按照他人已有的标准来写代码。

代码风格提示：变量命名使用蛇形命名，即 `connect_every_lowercase_word_by_underscore`

我们可以这样打印整数变量

```
1  #include <stdio.h>
2
3  int main() {
4      int math_grade;
5      math_grade = 120;
6
7      printf("%d\n", math_grade);
8
9      return 0;
10 }
```

运行程序，输出了 120。我们目前只需要这样理解：引号内的 `%d` 是一个占位符，引号后面的 `math_grade` 就是用来替换占位符的。因此我们就可以实现一些更复杂的输出了，比如

```
1  printf("数学成绩为:%d\n", math_grade);
```

现在拓展一下，如果我们定义了数学、语文和英语三门课的成绩，也即

```
1  int math_grade = 120;
2  int chinese_grade = 100;
3  int english_grade = 140;
```

如何输出呢？我们只需要写三个占位符即可，也就是

```
1  printf("数学成绩为:%d, 语文成绩为:%d, 英语成绩为:%d\n", math_grade, chinese_grade, english_grade);
```

想象一下，假如我们定义了很多变量，有数学期中成绩，数学期末成绩，语文小测成绩，语文期末成绩……我们在代码里看这么多东西就头大，所以我们希望能用自然语言给它们加上注释，便于我们理解代码。在 c 语言中，有两种形式的注释：单行注释和多行注释。

单行注释是两个斜杠 (//)，这两个斜杠之后就可以随便写字了，比如

```
1  #include <stdio.h>
2
3  int main() {
4      int math_grade; // 声明数学成绩
5      math_grade = 120; // 给数学成绩赋值120
6
7      printf("%d\n", math_grade); // 打印数学成绩
8      // 单行注释也可以单独成行
9
10     return 0;
11 }
```

多行注释的语法是/\* \*/，在这之间可以随便写字，比如

```
1  #include <stdio.h>
2
3  int main() {
4      /*
5       该程序的展示了变量的声明与赋值语法，
6       以及使用printf打印整型变量的方法
7      */
8      int math_grade; // 声明数学成绩
9      math_grade = 120; // 给数学成绩赋值120
10
11     printf("%d\n", math_grade); // 打印数学成绩
12
13     return 0;
14 }
```

需要注意的是，多行注释不能这样写

```
1  int math_grade; /* 声明数学成绩
2  math_grade = 120; 给数学成绩赋值 */
```

这样的话，第二段代码就被包括在了注释内，从而无法执行。

那么什么时候用单行注释，什么时候用多行注释呢？如果写注释的目的是对一行或几行代码做简单的说明，那么就用单行注释；假如注释的目的是要阐释一个复杂的原理（比如一段算法的数学原理或工程背景）或者解释一个函数（接下来会讲函数是什么）的功能，那么就应该使用多行注释。

有了变量，肯定也要有常量。在 c 语言中，我们使用#define 来声明常量。比如，圆周率  $\pi$  就是一个常量，我们可以这样在 c 语言中定义圆周率

```
1  #define PI 3
```

也即，定义常量的语法是 `#define 常量名 常量的值`。

我们定义圆周率是 3，是一个整数，因此可以按照上文的方法将其打印出来

```
1 #include <stdio.h>
2
3 #define PI 3
4
5 int main() {
6     printf("圆周率的近似值是%d\n", PI);
7
8     return 0;
9 }
```

你可能觉得奇怪，为什么定义常量的语法和变量区别这么大呢？既不需要写赋值号，也不需要标明类型。实际上，`#define` 是一个宏，在 c 语言中，程序被编译运行前，会有一个叫做预处理器的程序将所有宏替换为对应的值，也就是说（在只考虑 `#define` 宏的情况下），上面的代码经过预处理器会变成这样

```
1 #include <stdio.h>
2
3 int main() {
4     printf("圆周率的近似值是%d\n", 3);
5
6     return 0;
7 }
```

发现了吗？我们定义的常量 `PI` 被直接替换成了它的值 3，既然是直接替换，就相当于写死在了程序里，没法修改，所以就是常量了。实际上，我们行首的 `#include` 也是一个宏，它的效果和 `#define` 类似。关于宏和预处理的知识我们在之后会学习到。

代码风格提示：命名常量，一般采用大驼峰命名法（`CamelCase`），也即所有单词都连在一起，每个单词的首字母大写。如 `CapitalizeTheFirstLetterOfEachWord`

### 1.3 scanf 函数

`scanf` 函数是 `scan function`，它用来从终端获取输入的文本。这个东西涉及到一些我们还没有学到的知识，但是我们后续的学习很难离开它，因此需要在这里对其简单介绍，我们目前不需要了解它的原理，会用就可以。

首先，打开 `visual stdio`，在 `main.c` 中输入一段这样的代码

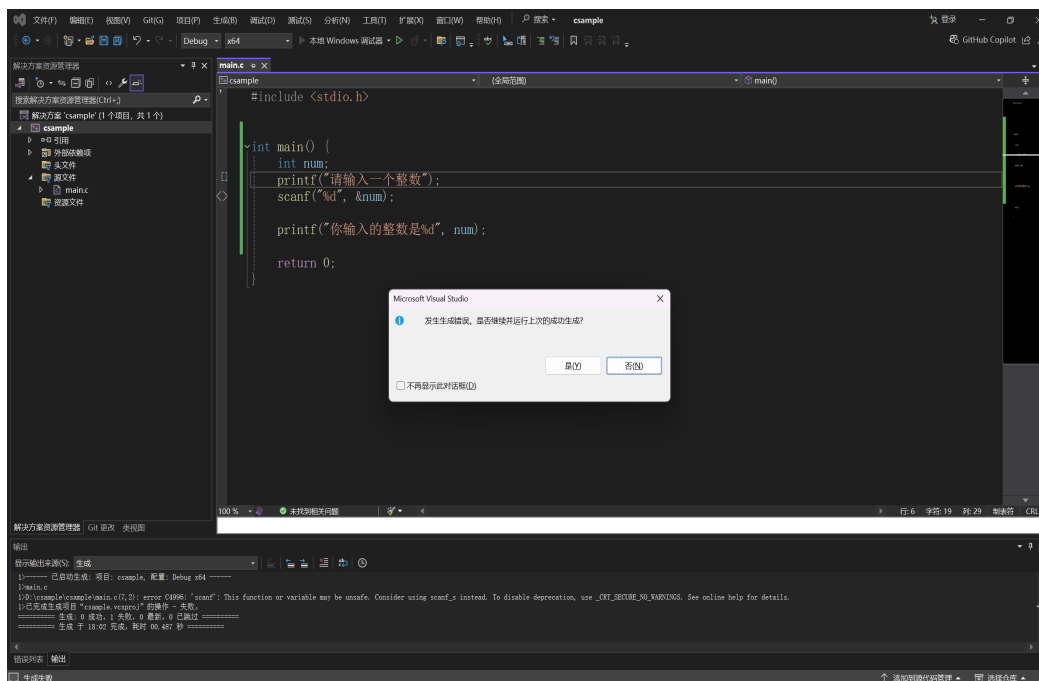
```
1 #include <stdio.h>
2
3 int main() {
4     int num; // 用于储存用户输入的数字
5     printf("请输入一个整数"); // 提示用户进行输入
6     scanf("%d", &num); // 将输入的数字保存在变量num中,注意不要忘了num前的&
7 }
```

```

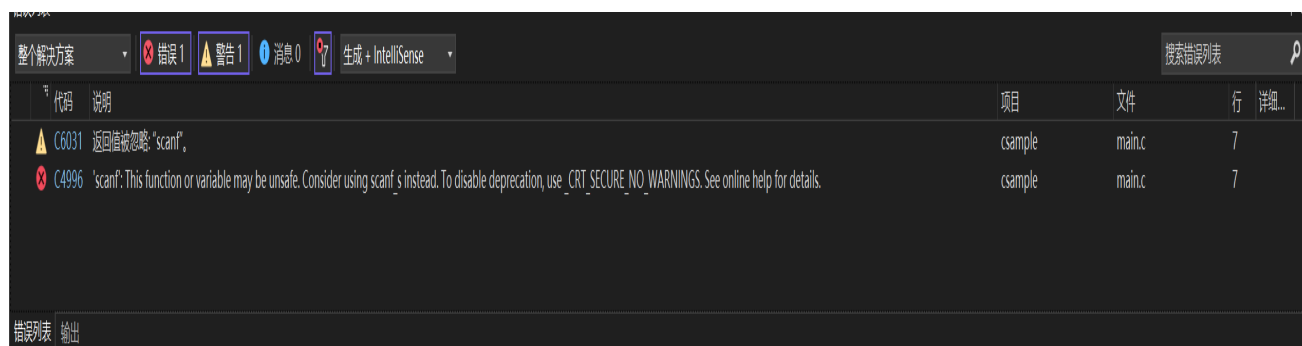
8     printf("你输入的整数是%d", num);
9
10    return 0;
11 }

```

点击运行，出现了意料之外的情况



这个窗口的意思是我们的代码有问题，过不了编译。在实际编程开发时，由于我们很难做到一次就尽善尽美，所以常常会遇到这个窗口。还好代码的试错成本是很低的，我们只需要着手去修补自己的代码即可。我们点击否，注意到 visual studio 最下面出现了这样的文本



带有黄色警示标志的叫做警告（warning），它的意义是代码可能有一些不符合规范的地方，需要修改，但是不改也能运行。而红色标志的叫做错误（error），它的意义是这里的代码完全有问题，过不了编译，程序运行不了。我们必须消除代码中所有的 error，也要尽可能消除所有的 warning。

我们开始着手消灭这个 error，它的提示信息说 “This function or variable may be unsafe. Consider



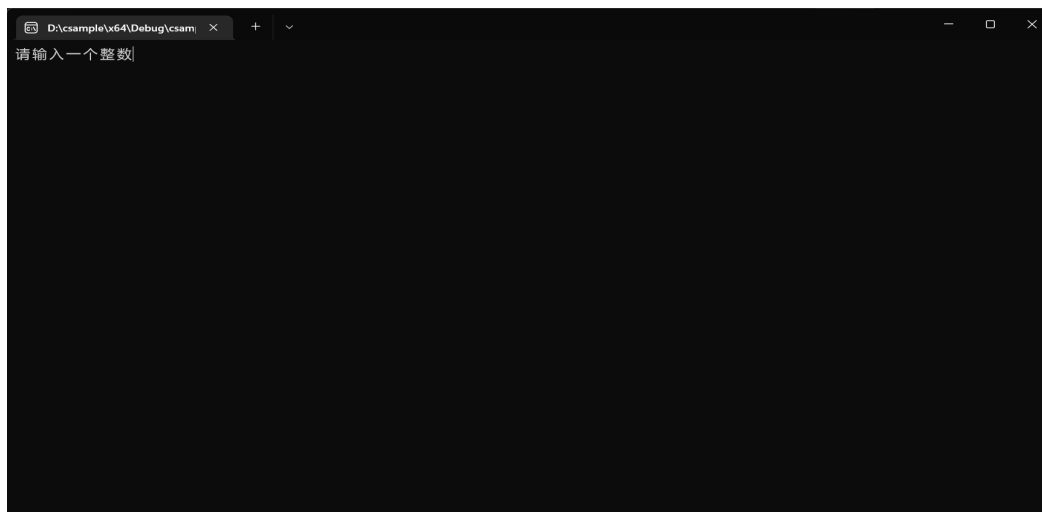
using `scanf_s` instead. To disable deprecation, use `_CRT_SECURE_NO_WARNINGS`. See online help for details.”

翻译成中文，就是“这个函数或者变量可能不安全。考虑使用 `scanf_s` 替代。若要取消废弃（检查），使用 `_CRT_SECURE_NO_WARNINGS`。详细信息见在线帮助。”。它的意思就是说，我们用的这个 `scanf` 函数不安全，已经废弃了，要么我们换用 `scanf_s`，要么我们加上 `_CRT_SECURE_NO_WARNINGS` 来强行使用废弃的函数。

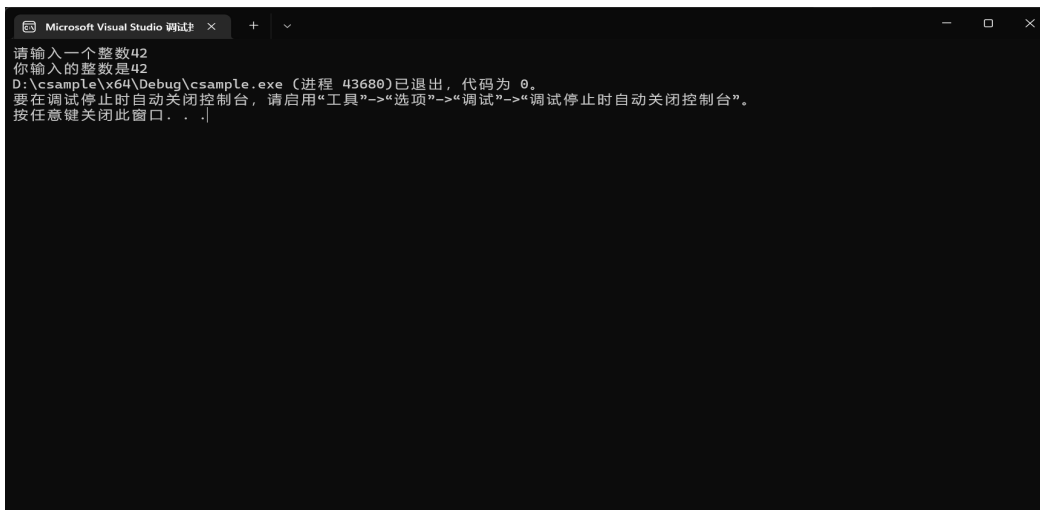
实际上，`scanf` 由于设计不当，会有缓冲区溢出（buffer overflow）的风险，所以我们确实不应该使用它。但是呢，只有我们使用的 visual studio 会报这个错误，其它 ide 不会强制要求你换，而且大部分的 c 语言试题和 c 语言教程仍在使用 `scanf`，所以我们采取强制使用废弃函数的方法来消除这个 error。

在 visual studio 的项目右键，点击属性，点击“C/C++”，点击“预处理器”，点击“预处理器定义”，随后点击右边出现的小箭头，点击“<编辑>”，单独一行输入“`_CRT_SECURE_NO_WARNINGS`”，点击“确定”即可。

此时我们再运行代码，终端首先变成这样



此时我们输入一个数字，点击回车，出现了这样的效果



如果我们要进行的是更复杂的输入呢？比如，我们需要数学、语文和英语三科成绩，那么可以这样写

```
1  #include <stdio.h>
2
3  int main() {
4      int math_grade;
5      int chinese_grade;
6      int english_grade;
7
8      printf("请输入数学、语文、英语成绩");
9      scanf("%d %d %d", &math_grade, &chinese_grade, &english_grade);
10
11     printf("三科成绩分别是:%d, %d, %d", math_grade, chinese_grade, english_grade);
12
13     return 0;
14 }
```

我们只需要按照这样输入：120 130 140，就能输出正确的成绩了。如果我们希望输入是 120,130,140，那么就把 scanf 的一行改成

```
1  scanf("%d,%d,%d", &math_grade, &chinese_grade, &english_grade);
```

这样我们就找到了规律，scanf 的引号内的部分就像是一个模板，只有输入与模板对应，函数才正常运行。不信的话，我们可以试试在修改后的代码输入 120 130 140，会得到这样的结果



非常抽象是不是？你知道更抽象的是什么呢？假如我们使用其它编译器，那么会得到不同的结果<sup>1</sup>。这是因为我们的输入没有符合模板的需求，模板需要我们在每两个成绩间使用英语逗号分割，我们用的却是空格，这种情况被称作未定义行为（undefined behavior），这是c语言中比较难以预测的一种的错误，我们后续会详细讲解。不知道各位记不记得，我们在解决那个error时，留了个warning尚未处理？它说我们没有使用scanf的返回值，这个“返回值”就是用来解决这个问题的，我们学习到函数时会专门提到。

## 2 基本数据类型和数值运算

### 2.1 基本数据类型

我们在上一章介绍变量的时候就已经使用了代表整数的int，但是只有整数显然是远远不够用的，c语言还提供了不同取值范围的整数、小数、字符和字符串。这些统称为基本数据类型。

在计算机科学中，整数一般被称作整型。根据取值范围的不同，总共有四种整型：short（短整型），int（整型），long（长整型）和long long，可以这样声明整型变量

```
1 short a = 1;
2 int b = 2;
3 long c = 3;
4 long long d = 4;
```

short 的取值范围是  $[-32768, 32767]$ <sup>2</sup>，int 的取值范围是  $[-327682147483648, 327672147483647]$ ，long 和 long long 的就更大了。实际上还有一类无符号整型，就是只有非负数的整型。在整型前加一个 unsigned 就是对应的无符号整型

```
1 // 无符号类型
2 unsigned short a = 1;
```

<sup>1</sup>类似的一段代码，以 120 130 140 作为输入，64 位电脑，处理器 intel ultra9，在 win11，msvc 编译下运行结果为 120, -858993460, -858993460；ubuntu，gcc 编译下结果为 120, 4096, 0；ubuntu，clang 编译下结果为 120, 0, 4096

<sup>2</sup>以笔者电脑（64 位 win11，MSVC STL）的 limits.h 作为标准

```
3 unsigned int b = 2;
4 unsigned long c = 3;
5 unsigned long long d = 4;
```

short 的范围是  $[-32768, 32767]$ ，而 unsigned short 的范围是  $[0, 65535]$ ，相当于把负数部分都转移给正数了。其它类型对应的无符号类型也是一样。

介绍了这么多，笔者建议大家：除非你有明确的设计要求，对计算机了解比较清晰，非常清楚自己在写什么，否则就一律使用 int。哪怕是明确了需要处理的数据都是非负也不要无符号，哪怕是确定 short 就够包含自己需要的数据也要用 int。

为什么我们这么说？首先，为什么 c 语言设计了这么多种整型和无符号这些东西呢？我们知道，在 c 语言诞生的年代，内存是十分紧张的资源，所以设计多种占用不同内存大小的整型也是便于程序员精准控制内存。但是现代计算机内存大多都是 8GB，16GB 甚至更高，已经没有必要过于珍惜内存了。

如果我们使用推荐外的类型，一方面会增大开发难度，另一方面容易诞生意料以外的问题。为什么呢？首先请考虑一个问题，以 short 为例，它的最大值是 32767，那假如我们给一个 short 变量赋值 32768 会怎么样？我们写一段代码试一下

```
1 #include <stdio.h>
2 #include <limits.h>
3
4 int main() {
5     short a = SHRT_MAX + 1; // SHRT_MAX就是short的最大值,即32767
6     printf("%hd\n", a);
7
8     return 0;
9 }
```

运行结果是-32768，也就是说，如果我们赋的值超过了数据类型取值范围的最大值，它就会从最小值开始赋。一样的，若我们赋的小于范围最小值，就会从最大值开始赋。这就像是把取值范围这根数轴首尾连接在了一起一样。类似的结果也会出现在其它类型，这是 c 语言的特殊设计。

想象一下，假如你在设计一个学生管理软件，为了节省内存，使用 short 来记录学生的编号（因为你觉得学校小，不可能超过三万人），但是使用你开发的系统的学校今年突然人数暴增，现在学校内有了 32768 人，那么最后一个同学的编号就是-32768，这位同学得有多无奈？因此，一般不要使用比 int 小的类型的。

既然使用 short 有这样的危害，为什么不都使用 long 呢？首先，实际场景中很少存在数据大到连 int 都装不下的情况（谁家学校有三百多亿人啊），因此没必要选这么大的；另一方面，如果需要处理的数据都是普遍极大的，一般都是由开发者自行实现适用的记录长数字的类型，使用 long 不划算。所以我们也一般不考虑比 int 大的类型。

那么为什么无论使用场景都不用无符号呢？首先，无符号的最大值和最小值也是连接在一起的，这种情况更刺激：假如一个学校有 65536 名学生，我们使用 unsigned short 来编号，那么最后一位和第一位的编号是一样的，就没办法区分这二人了。其次，我们有时设计的一些复杂算法可能出错，若算法应只输出正值却输出了负值，那么说明算法一定出了问题，假如我们使用无符号类型，就无法这样判断了。

小数在计算机科学一般称作浮点型，常用的是 double（双精度浮点型）。我们可以这样定义一个浮点变量

1      `double a = 1.23; //` 小数点是英文句号