

我是8位的

I am 8 bits, what about you?

随笔 - 205, 文章 - 0, 评论 - 103, 阅读 - 101万

导航

博客园

首页

新随笔

联系

订阅

管理

<

2022年3月

>

日	一	二	三	四	五	六
27	28	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

公告

你的支持是我的动力

欢迎关注微信公众号“我是8位的”



昵称：我是8位的

园龄：4年7个月

粉丝：288

关注：5

+加关注

盖楼抽奖

#她的梦想在发光#

HWD科技女性故事有奖征集

分享最打动你的科技女性故事

活动时间：2022年3月8日-3月18日

马上参与



搜索

找找看

谷歌搜索

常用链接

我的随笔

我的评论

我的参与

最新评论

我的标签

积分与排名

积分 - 457097

排名 - 1198

概率统计22——假设检验理论（1）

我们可以根据经验或统计量对一些事情做出断言，问题是，如何判断这个断言的合理性？假设检验为我们提供了一种利用样本检验断言是否可靠的方法，能够让我们通过已有的证据验证断言是经过缜密的运算，还是毫无根据的瞎猜。

假设检验的背景

某个机器元件的质量标准是功率，功率越大越好，这个元件影响到公司的核心竞争力。技术组在攻克了重重难题后宣称有了重大突破。总经理非常高兴，宣布批量试制。然而改进后的数据却并不那么漂亮，功率均值从原来的600mW提升到603mW，仅仅提升了0.5%，这让总经理大为光火，立即召集技术组开会。

改进前，600mW

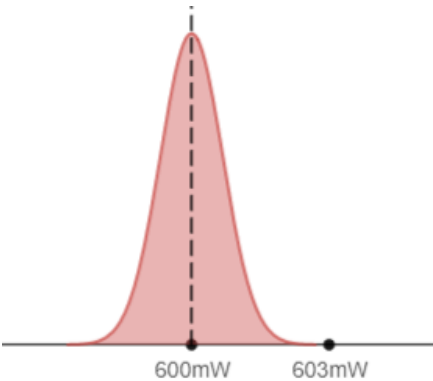
↑ 0.5%

改进后，603mW

总经理：
你们是在逗我吧？

我们取得了
重大改进！

0.5%看起来确实没什么太明显的提升。面对面色铁青的总经理，技术部拿出了改进前的功率分布：



随笔分类 (211)

- ★★资源下载★★(1)
- Java并发编程(1)
- 程序员的数学(24)
- 单变量微积分(31)
- 多变量微积分(24)
- 概率(24)
- 机器学习(27)
- 软件设计(1)
- 数据分析(6)
- 数据结构与算法(27)
- 随笔(5)
- 线性代数(34)
- 项目管理(2)
- 转载(4)

随笔档案 (205)

- 2021年2月(1)
- 2020年3月(2)
- 2020年2月(6)
- 2020年1月(4)
- 2019年12月(7)
- 2019年11月(15)
- 2019年9月(3)
- 2019年8月(6)
- 2019年7月(1)
- 2019年6月(8)
- 2019年5月(3)
- 2019年4月(5)
- 2019年3月(7)
- 2019年2月(3)
- 2019年1月(7)
- 更多

阅读排行榜

- 1. 使用Apriori进行关联分析（一）(29768)
- 2. 线性代数笔记12——列空间和零空间(28772)
- 3. FP-growth算法发现频繁项集（一）——构建FP树(24430)
- 4. 寻找“最好”（2）——欧拉-拉格朗日方程(23099)
- 5. 多变量微积分笔记3——二元函数的极值(22772)

评论排行榜

- 1. 隐马尔可夫模型（一）(8)
- 2. 线性代数笔记12——列空间和零空间(7)
- 3. 线性代数笔记3——向量2（点积）(7)
- 4. FP-growth算法发现频繁项集（一）——构建FP树(5)
- 5. 寻找“最好”（2）——欧拉-拉格朗日方程(4)

推荐排行榜

- 1. 寻找“最好”（2）——欧拉-拉格朗日方程(7)
- 2. FP-growth算法发现频繁项集（一）——构建FP树(7)
- 3. 线性代数笔记3——向量2（点积）(6)
- 4. FP-growth算法发现频繁项集（二）——发现频繁项集(5)
- 5. 隐马尔可夫模型（一）(5)

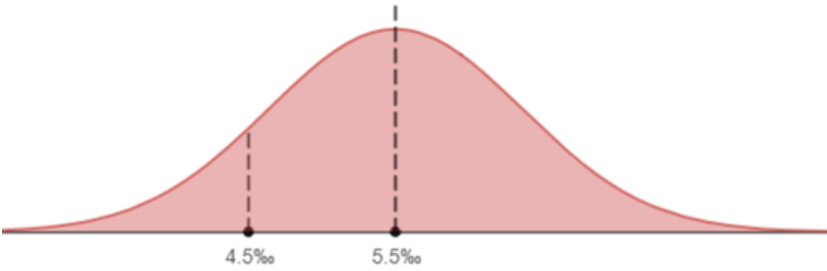
最新评论

- 1. Re:线性代数笔记3——向量2（点积）
如果点积小于0，即夹角小于90°，这个写错了吧。应该是夹角大于90°
--猫猫猫猫猫大人

这是个很陡峭的分布曲线，暗示着这个元件的功率控制得十分精细，或者说功率的波动很小，3mW的波动已经相当于“基因突变”了。这样看来，0.5%的提升确实算得上是重大改进。

某个软件公司用每千行的bug数量衡量软件的质量（这里不讨论这种方法是否合理），多年来也做过了很多项目，平均bug率是5.5‰，即每千行代码5.5个bug，勉强达到CMM2的标准。

最近公司全面实施了新的开发方法，发现新项目的bug率降低到4.5‰，这可是降低了18%！于是QA部门的老大在年会上对着分布图激动地宣布，公司软件质量从此迈上了新台阶。



程序员们瞟了一眼分布图，露出了谜之微笑。



上面两个故事告诉我们，在质量改善的过程中，质量提升的百分比并不能作为有效性的唯一依据，0.5%可能是重大提升，18%也可能根本说明不了问题。检验的依据是，改善后是否在改善前总体正常波动的范围内。如果在正常波动范围内，则可以认为并没有什么改善，否则认为有显著改善。这个显著性具体如何判断呢？

显著性水平与置信水平

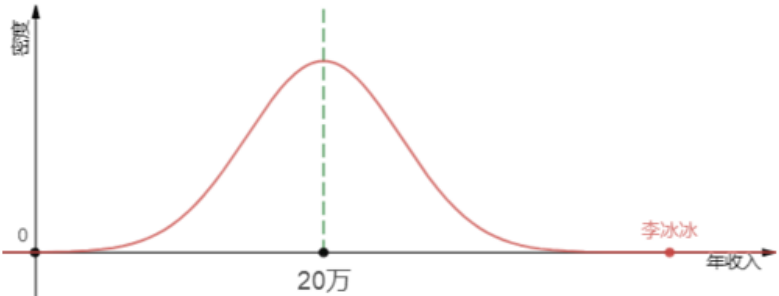
假设下图是哈尔滨市民年收入的概率分布：

2. Re:线性代数笔记10——矩阵的LU分解写的很好，不过LU分解的前提是错的，LU分解只需要第三个条件，如果允许行置换就是下面写到的PLU，可以分解所有矩阵
- wiki3D
3. Re:单变量微积分笔记20——三角替换1 (sin和cos)很nice
- 尹保棕
4. Re:线性代数笔记24——微分方程和exp(At)有些图片挂了呢
- ccchendada
5. Re:寻找“最好”（2）——欧拉-拉格朗日方程提个issue，最速降线中 $v = \{2gh\}^{1/2}$ 与配图不一致，建议以起点为原点，向右伸出x轴，向下伸出y轴建立坐标系
- trustInU



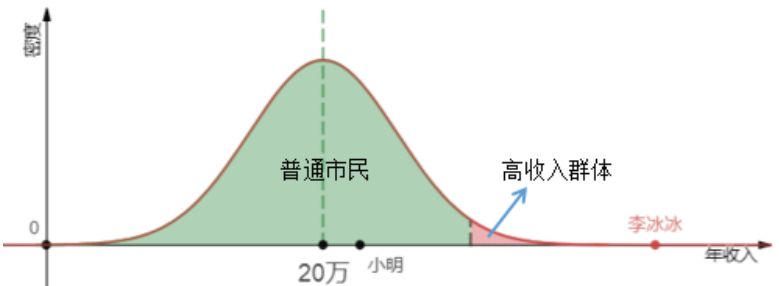
大多数人都集中在20万附近徘徊。

恰好李冰冰出生于哈尔滨市，不小心把她也统计进去了。大明星的收入肯定比普通市民多得多。



李冰冰明显和普通市民不是一伙的，她的年收入显著高于绝大多数市民，早已跨过了高收入的门坎。某个人的收入落在高收入群体的概率就是显著性水平（Significant Level）。

小明的年收入是25万，和大多数人差不多，我们相信小明是一名普通市民的概率还是比较高的，这个概率就是置信水平（Confidenct Level）。



高收入群体占了总体的多大比例呢？也就是显著性水平应该如何取值？

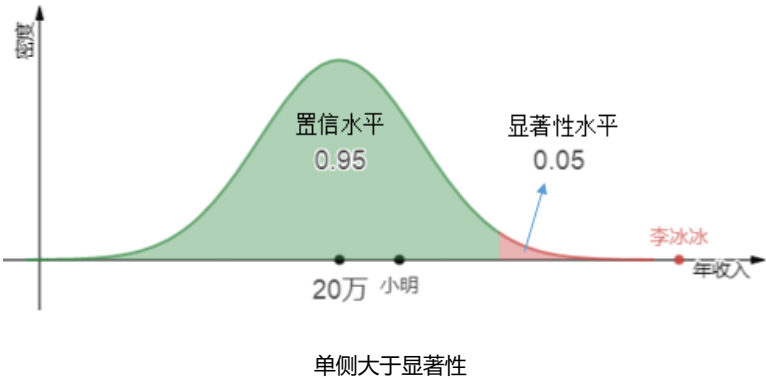
第一个回答这个问题的人是罗纳德·艾尔默·费舍尔 (RonaldAylmerFisher) 。



费舍尔（1890 - 1962）是英国统计学家、生物进化学家、数学家、遗传学家和优生学家。是现代统计科学的奠基人之一。著名的F分布就是费舍尔提出的，并以其姓氏的第一个字母命名。

费舍尔提出0.05显著性水平，至于为何要选择0.05，老头在《科研工作者的统计方法》（Statistical Methods for Research Workers）中做过长篇大论，网上有一篇文章叫“Why P=0.05?” (<http://www.jerrydallal.com/lhsp/p05.htm>)，很值得看一下。

现在我们用显著性水平和置信水平去看待哈尔滨市民年收入的问题：



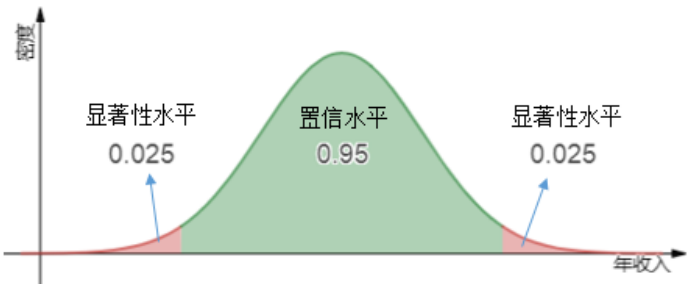
高收入群体在曲线右侧，是“单侧大于”显著性，表示一个人属于高收入群体的概率是0.05，李冰冰明显属于这一侧。相对的，一个人不属于高收入群体的概率是0.95。

现在要判断一个人是否属于低收入群体，此时应当使用“单侧小于”显著性。小明隔壁正好住着一个叫王二的游手好闲的邻居，每个月都等着领政府的救济金混日子，年收入自然很低，绝大多数人都不愿意和他为伍。这类人游手好闲群体在曲线左侧的0.05部分：



单侧小于显著性

除此之外，还有双侧显著性。按照费舍尔的0.05显著性水平理论，双侧检验时要在总体分布的两端各设定一个临界点，临界点以外，两端阴影部分的面积比率各为0.025，表示超低收入群体和高收入群体一样稀少：



有人说我显著性水平不用0.05行不行？当然没问题，具体怎么设定完全取决于实际问题，0.05只是一个常用的取值。如果想检验得严一点，显著性水平要相应放低，比如0.01；如果宽松一点，显著性水平可以稍高一点，比如0.1，此时你也要同时承担置信水平下降的后果。

有了显著性水平后和置信水平后，就可以从直观上回应总经理的愤怒并理解程序员们的谜之微笑，但我们仍然缺少一些理论知识。检验理论究竟蕴含着怎样的逻辑呢？

假设检验的逻辑

我一直很喜欢打乒乓球，一个好友和我从小学一直打到大学毕业，基本上旗鼓相当。工作之后，我天天苦练球技，自认为如果再次交手一定比他更强。年假回家，我俩又一次站在球台边，并以15局为限，看看时隔多日之后谁更胜一筹。

结果是我5胜10败。

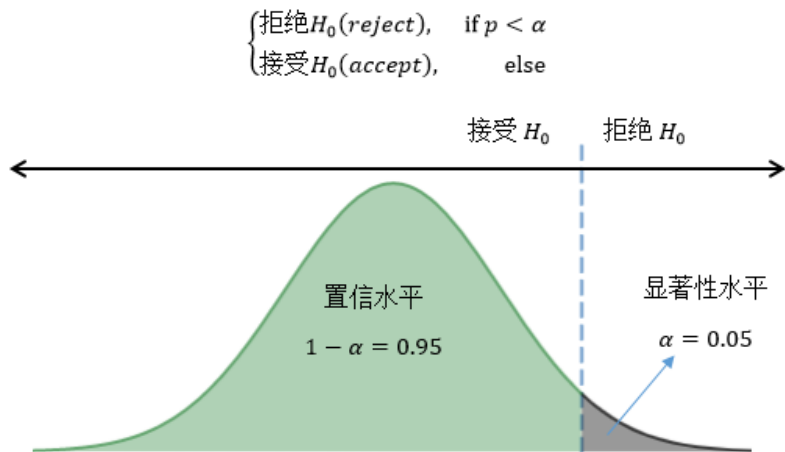
由于几乎每一局都是苦战，所以我并不服气，认为这纯属偶然，如果再打一次，结果完全不可预估。我的理由是：抛15次硬币，10次正面朝上完全有可能。对方则是持否定态度，认为这绝对是实力的差距。

我们二人的争论包含了两种截然相反的假设：

我的假设， H_0 ：双方势均力敌，获胜的概率都是0.5。

朋友的假设， H_1 ：这个结果表明实力的差距，他的获胜的概率大于0.5。

从朋友的立场来看，他想要证明自己是正确的，因此把我的 H_0 假设看成原假设（或者叫待验假设、虚无假设），也是他想要驳斥的假设。 H_1 是 H_0 的备择假设（或者叫对立假设、备选假设），是他试图肯定的假设。他的证明逻辑是：如果 H_0 成立，那么出现 H_1 这样的数据的概率将小于 α ，这个概率用 p 表示，称为 p 值（ p -value），作为阈值的 α 是上一节的显著性水平，取 $\alpha=0.05$ 。用 p 值是否大于 α 作为判断依据：



这有点类似于反证法，如果无法直接证明 H_0 是错的，就转换策略，暂且相信 H_0 是对的，这种情况下出现 H_1 的概率将小于0.05。这种小概率事件应该不容易在抽样中出现，但现在偏偏出现了，是个强有力的反驳证据，所有应当转而相信 H_1 。

现在来计算一下p值。按照原假设 H_0 ，我们二人实力相当，每一场比赛的胜率都相等。假设每场比赛都是独立的，用随机变量 X 表示朋友获胜的总数，那么 X 符合总数是15，胜率是0.5的二项分布 $X \sim B(15, 0.5)$ ，p值是：

$$p = P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{k=1}^{10-1} (C_{15}^k \times 0.5^k \times 0.5^{15-k})$$

可以用计算机直接计算：

```
from scipy import stats

c = stats.binom.cdf(10 - 1, n=15, p=0.5) # X~B(n, p)的累积分布
p = 1 - c
```

结果 $p \approx 0.151 > 0.05$ ，看起来是我赢了，应该相信 H_0 。

朋友当然不会满意这个结果，他反驳的理由是样本数量太小，并不能反应真正的胜败分布。于是他提出增加比赛次数，重新打100局。我欣然接受。

几天后终于决出了结果，我38胜62败。

还是沿用之前的原假设和备择假设，用随机变量 X 表示朋友获胜的总数，这次的p值是：

$$p = P(X \geq 62) = 1 - P(X < 62) \approx 0.0105 < 0.05$$

p值小于显著性水平，因此应当拒绝承认原假设 H_0 ，接受对立假设 H_1 ，从分布图上能够更清晰地看出这一结论。

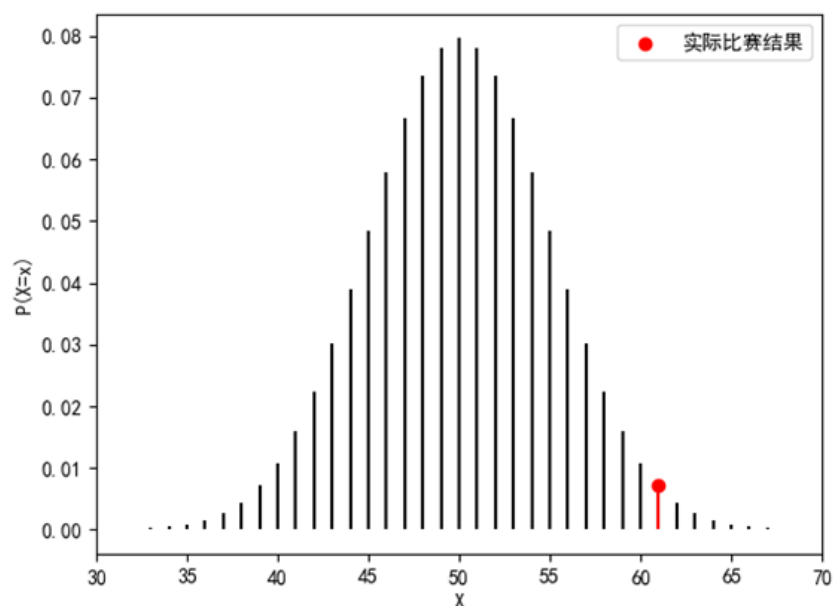
```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
```

```

n, p = 100, 0.5
x = 62 # 获胜次数
c = stats.binom.cdf(x - 1, n=n, p=p) #  $X \sim B(n, p)$  的累积分布,  $P(X \leq x-1)$ 
p_value = 1 - c
print(p_value)

xs = np.arange(0, n + 1)
ys = stats.binom.pmf(xs, n=n, p=p) #  $X \sim B(n, p)$ 
plt.vlines(xs, ymin=0, ymax=ys)
plt.vlines(xs[x - 1], ymin=0, ymax=ys[x - 1], colors='r')
plt.scatter(xs[x - 1], ys[x - 1], c='r', label='实际比赛结果')
plt.xlim((30, 70))
plt.xlabel('X')
plt.ylabel('P(X=x)')
plt.legend(loc='upper right')
plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签
plt.show()

```



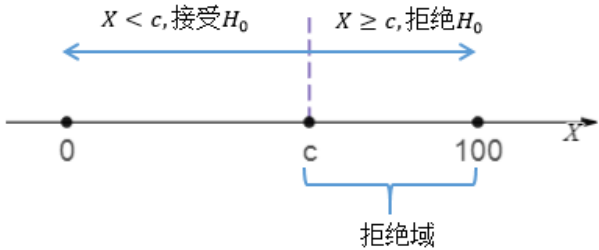
这次的样本较大，证据要比之前充分得多，所以我应当承认这次的失利绝非偶然。

值得注意的是，假设检验无法并不能给出绝对的证明，仅仅是用小概率事件作为证据，对原假设进行驳斥。基于这个原因，我依然不服气，提出再打100局。当然，朋友也坚决予以否定——1局也不行。

临界值

38胜62败的战绩为驳斥 H_0 提供了强有力的证据，我现在想要知道的是，到底胜多少局才能不被驳斥？也就是0.95置信水平和0.05显著性水平的临界点应该在什么位置？

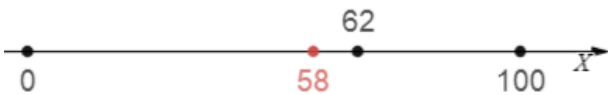
以 c 为临界点，将置信水平和显著性水平分隔开来， c 的值称为临界值。用 X 表示朋友获胜的数量，如果 $P(X \geq c) = 1 - P(X < c) \leq \alpha$ ，说明样本位于拒绝域中，应该转而相信备择假设 H_1 。



直接计算c值不太容易，需要使用分布的反函数，幸好可以通过计算机直接计算：

```
c = stats.binom.ppf(0.95, 100, 0.5) # 当P(X ≤ c) = 0.95时的c值
```

结果c=58，这意味着如果我输掉58局以上，就应该属于实力的差距。



假设检验与点估计

我们回到改善机器元件的故事，从抽样分布的角度去看假设检验。

已知元件改善前的功率服从均值为 μ_0 ，方差为 σ_0^2 的正态分布，以下是总经理和技术组的观点。

总经理：技术组纯粹是在说瞎话，质量毫无提升。

技术组：质量得到了极大的改善。

设改善后的总体均值是 μ_1 ，方差是 σ_1 ，现在用假设检验的逻辑翻译一下：

总经理，原假设 H_0 ：改善前与改善后是同一个正态分布， $\mu_0=\mu_1$ 、 $\sigma_0=\sigma_1$ 。

技术组，备择假设 H_1 ：改善前与改善后是不同的正态分布， $\mu_0 < \mu_1$ 。

公司用新技术制造了大量元件，从中多次抽取容量是m ($m \geq 30$) 的样本并计算出均值。根据中心极限定理，样本均值的抽样分布服从均值为总体均值，方差为总体方差1/m的正态分布：

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{m}\right)$$

对样本均值进行标准化处理：

$$Z = \frac{\bar{X} - \mu_1}{\frac{\sigma_1}{\sqrt{m}}}$$

按照原假设 H_0 ：

$$\mu_1 = \mu_0, \quad \sigma_1 = \sigma_0$$
$$\Rightarrow Z = \frac{\bar{X} - \mu_1}{\frac{\sigma_1}{\sqrt{m}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma_0}{\sqrt{m}}}$$

根据大数定律和点估计理论，样本标准差 σ_s 是总体标准差的合理估计量：

$$Z = \frac{\bar{X} - \mu_1}{\frac{\sigma_1}{\sqrt{m}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma_0}{\sqrt{m}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma_s}{\sqrt{m}}}$$

假设元件原来的功率均值是600mW，技术改进后抽取了30个样本进行对比，测得样本的功率均值是603mW，标准差是6。代入后可以求得具体的Z值：

$$z = \frac{\bar{X} - \mu_0}{\frac{\sigma_s}{\sqrt{m}}} = \frac{603 - 600}{\frac{6}{\sqrt{30}}} \approx 2.739$$

Z服从均值为0方差为1的标准正态分布，可以通过查表求得临界值：

$$P(Z < c) = 0.95 \xrightarrow{\text{查表求得}} c \approx 1.645$$

或者直接使用下面的代码计算：

```
stats.norm.ppf(0.95, 0, 1) # 1.6448536269514722
```

$z > c$ ，z落入了拒绝域，因此应当拒绝相信 H_0 ，原假设不成立，技术组的观点才是正确的。

接下来是非常绕的第一类错误和第二类错误（待续）。

出处：微信公众号“我是8位的”

本文以学习、研究和分享为主，如需转载，请联系本人，标明作者和出处，非商业用途！

扫描二维码关注作者公众号“我是8位的”



随笔

分类: 概率

标签: p值, 备则假设, 虚无假设, 假设检验

好文要顶

关注我

收藏该文





我是8位的

关注 - 5

粉丝 - 288

+加关注

1

推荐

0

反对

« 上一篇: 概率统计21——指数分布和无记忆性

» 下一篇: 概率统计23——假设检验理论（2）

posted on 2020-03-04 16:47

我是8位的

阅读(1155)

评论(0)

编辑

收藏

举报

[刷新评论](#) [刷新页面](#) [返回顶部](#)

 登录后才能查看或发表评论，立即 [登录](#) 或者 [逛逛](#) [博客园首页](#)

- 【推荐】华为 HWD 2022 故事征集，分享最打动你的科技女性故事
- 【推荐】华为开发者专区，与开发者一起构建万物互联的智能世界



编辑推荐:

- 革命性创新，动画杀手锏 @scroll-timeline
- 戏说领域驱动设计（十二）—— 服务
- ASP.NET Core 6框架揭秘实例演示[16]：内存缓存与分布式缓存的使用
- .Net Core 中无处不在的 Async/Await 是如何提升性能的？
- 分布式系统改造方案 —— 老旧系统改造篇



最新新闻:

- 乔布斯的创业搭档：他缺乏工程师才能，不得不锻炼营销能力来弥补
 - 美国大厂码农薪资曝光：年薪18万美元，够养家，不够买海景房
 - 两张照片就能转视频！Google提出FLIM帧插值模型
 - Android 再推“杀手级”功能，可回收 60% 存储空间
 - 溺在理财暴雷潮的投资人：本金63万，月兑25元不够卖菜
- » 更多新闻...

Powered by:
博客园
Copyright © 2022 我是8位的
Powered by .NET 6 on Kubernetes