t 分佈的歷史

• 在英語文學中，它需要

它的名字來自William Sealy Gosset的

1908年在Biometrika下的論文

筆名“學生”。

• 戈塞特 在吉尼斯世界紀錄

啤酒廠在都柏林，愛爾蘭，和

對小問題感興趣

樣品，例如化學品

大麥的性質，其中樣本量

可能低至 3。

續

• 一個版本的起源

化名是 戈塞特的 僱主

首選員工在以下情況下使用筆名

發表科學論文而不是

他們的真名 ，因此 他使用了

命名「學生」以隱藏他的身份。

• 另一個版本是吉尼斯沒有

希望他們的競爭對手知道他們

使用 t 檢驗來檢驗品質

原料。

續

• Gosset的 論文指的是分銷

作為「頻率分佈標準”

從

正常人群“。

• 通過工作而聞名

羅納德·A·費舍爾（Ronald A. Fisher），他稱

分佈 “學生的分佈”

並將該值稱為 t。

• 在上一次講座中，我們提到使用

以下公式可將

樣本均值（這是 一個隨機的

變數） 到 Z，並從 Z 分佈

（這是 標準正態分佈）

以定位 CI 的邊界。

• 要做到這一點，人們需要知道

標準偏差  的一般

人口。

n

x

跟

 /





x

• 在實踐中，我們可能不知道標準

一般人群的偏差。

• 相反，我們只有樣品標準

偏差 s。

• 因此，我們定義了一個新的隨機變數 t

以進行以下轉換。

• 與 Z 類似，t 有自己的概率

密度函數也是如此。與 Z 不同，t 不會有

正態分佈。

s n

x

t

/





學生的 t 分佈

• 在概率和統計中，學生的t-

分佈（或簡稱 t 分佈）是一個

連續概率分佈族

在估計

正態分佈人口

以下情況：

– 樣本量 n 很小

– 總體標準差  為

未知。

http：//en.wikipedia.org/wiki/Student's\_t-distribution

新的隨機變數 t

• 這被稱為學生的t

n1 度分佈

自由度（或 df = n1），並且經常是

用符號 tn1 表示

s n

x

t

/



 n取決於

，會有不同的t-

分佈。

自由度 （df）

• 自由度數為

最終計算中的值數

可以自由變化的統計資訊。

• 一般而言，自由度

參數的估計值等於

獨立分數的數量

到估計值減去

用作中間步驟的參數

在參數本身的估計中。

http://en.wikipedia.org/wiki/Degrees\_of\_freedom\_（統計）

DF – 示例

• 範例 1：估計平均值時

值  從 n 個數字開始，每個數位是

自由更改，無論

其他數位。因此 DF = n。

• 範例 2：估算標準時

偏離 n 個數字，我們需要

值 x̅ 這是平均值。一次 x̅

是固定的，只有 n-1 個可用數位

以進行更改。第 n 個數位是從屬的

在那些 n-1 數字和 x̅ 上。如

結果，DF = n-1。

續

• 在這裡，自由度測量

數據中可用的資訊量

可用於估計 2。

• 自由度為 n-1 而不是

n 因為我們失去了1個自由度

估計樣本均值 x̅。

• 回想一下，df = n1 是

我們把平方和除以

圍繞均值的偏差以獲得

方差。







 

n

我

十一 x

s n 1

2 （ ）2

（ 1）

1

• t 分佈看起來像標準正態分佈

曲線已被「踩踏」 - 它有點

更扁平，更胖

• t 分佈完全由其

自由度 （df） –

自由度，它越平坦。

0

正常 （t）

t15 （n=16）

t 分佈

學生的 t 分佈

n-1 自由度 （df），

用 tn-1 表示。

概率密度函數 （PDF）

各種自由度

 =df 以上。

• 回想一下，我們使用的公式

估計95%置信區間

• 這相當於說兩者

CI 的邊界是：

（ 1.96 ， 1.96 ）

n

x

n

x    

n

x 1.96 

將 1.96 替換為 2.58，表示 99% 置信區間。

標準誤差

均值（或 SEM）

• 當需要 t 校正時，

上一個方程變為

• 這相當於說兩者

CI 的邊界是：

( , )

n

s

x t

n

s

x t 

n

s

x t

t 的值將取決於 n 以及

您需要達到什麼水準的CI。

標準誤差

均值（或 SEM）

針對小樣本量進行調整

均值的 95% 置信區間的 T 值

df T df T

1 12.706 12 2.179

2 4.303 13 2.160

3 3.182 14 2.145

4 2.776 15 2.131

5 2.571 20 2.086

6 2.447 25 2.060

7 2.365 30 2.042

8 2.236 40 2.021

9 2.262 60 2.000

10 2.228 120 1.980

11 2.201 ~ 1.960

（n-1）

（此處 T 與 t 相同）

當不同的 CI 時，T 的值會有所不同

使用了間隔。例如，對於90%置信區間或99%置信區間。

2.306

df T df T

1 12.706 12 2.179

2 4.303 13 2.160

3 3.182 14 2.145

4 2.776 15 2.131

5 2.571 20 2.086

6 2.447 25 2.060

7 2.365 30 2.042

8 2.236 40 2.021

9 2.262 60 2.000

10 2.228 120 1.980

11 2.201 ~ 1.960

（n-1）

2.306

正如我們所記得的，可以得到1.960的值

由函數 norminv（0.975）同樣，一個

MATLAB 函數 tinv（0.975，df） 允許您

獲取這些提供 df 的 T 值。

df=1：100;

t=錫夫（0.975，df）;

情節（df，t）

當 DF=60 時，t 為

約2.0。

Q1：將是什麼

的價值

tinv（0.975， inf）？

Q2：將是什麼

的價值

tinv（0.5， 30）？

關於 t 修正的注意事項

• 您需要的 t 的值取決於

置信水準（例如，95% 或 90% 等）

您想要的樣本大小。

• 具有非常小的樣本量（n<15左右），

您還需要注意

數據在

樣本

– 需要“表現良好”（即，在某種程度上

正常）讓我們使用x̅± t\* SEM

創建置信區間

示例 1

• 隨機抽取10名兒童

來自接受以下疾病的嬰兒群體

抗酸劑（用於治療消化系統疾病）

含有鋁。

• 群體血漿的分佈

鋁大致正常

未知  和 。樣本均值和

標準差計算為 37.2

和 7.13.

• 我們想估計 95% 的置信區間

人口平均值 。

續

• 此處的自由度 DF = 10-1 = 9。

• 從下表中，我們知道t-

要使用的更正將是2.262，而不是

比通常使用的1.96。

n

s

x t

s n

x

t

 





/

df T df T 

1 12.706 12 2.179

2 4.303 13 2.160

3 3.182 14 2.145

4 2.776 15 2.131

5 2.571 20 2.086

6 2.447 25 2.060

7 2.365 30 2.042

8 2.236 40 2.021

9 2.262 60 2.000

10 2.228 120 1.980

11 2.201 ~ 1.960

2.306

續

（32.1， 42.3）

10

7.13

≤2.262~

 

x

n

s

x t 基於此 10 個子樣本

統計，我們有95%的信心

區間 （32.1， 42.3） 將

包含總體平均值。

例如，等離子鋁

42.0 將不予考慮

過量。

df T df T

1 12.706 12 2.179

2 4.303 13 2.160

3 3.182 14 2.145

4 2.776 15 2.131

5 2.571 20 2.086

6 2.447 25 2.060

7 2.365 30 2.042

8 2.236 40 2.021

9 2.262 60 2.000

10 2.228 120 1.980

11 2.201 ~ 1.960

2.306

評論

• 請注意，如果樣本數量增加到 n=31

（DF=30），t 值將變為 2.042。這

CI 將為 （32.6， 41.8）。

•如果人口已知為7.13，我們將

使用前面描述的公式來獲得CI，這

是 （32.8， 41.6）。[因此，42.0 將建議

過量。

• 在這兩種情況下，我們都看到更窄的CI（因為

更少或沒有更正）。

• 這裡我們使用的是雙面 CI。也就是說，要麼

多鋁或太少是壞的。

n

x 1.96 

在 z和 t 之間進行選擇？

在 z 和 t 之間進行選擇

• 試圖為總體均值構建 CI 

對於以下情況：

– n =150， x̅=100， s = 15， 並且總體有一個

偏分佈

– n =8， x̅=100， s = 15， 正態分佈

– n =8， x̅=100， s = 15， 一個非常偏斜的分佈

– n =150， x̅=100，  = 15， 偏分佈（幾乎

從不發生）

– n =8， x̅=100，  = 15， 一個非常偏斜的分佈

（幾乎 從不發生）

在 z 和 t 之間進行選擇

• 試圖為總體均值構建 CI 

對於以下情況：

– n =150， x̅=100， s = 15， 並且總體有一個

偏分佈 [t]

– n =8， x̅=100， s = 15， 正態分佈 [t]

– n =8， x̅=100， s = 15， 偏分佈 [？]

– n =150， x̅=100，  = 15， 偏分佈（幾乎

從不發生） [z]

– n =8， x̅=100，  = 15， 偏分佈（幾乎

從未發生過） [？]

9.4 置信度的應用

間隔

示例 2

• 一種叫做哌醋甲酯的藥物用於

治療20名兒童的注意力缺陷

紊亂（M組）。

• 另外20名患有該疾病的人被給予

安慰劑（P組）。

• 稱為父母評級量表（PRS）的測試是

用於評估他們的行為和

注意力狀態得到改善。一個更低的

分數意味著更好的改進。

• 這些PRS分數的分佈

兒童大致正常。這

總體均值和分數的變化

未知。

• 的均值和標準變化

M組分別為10.8和2.9。對於組 P

它們是14和4.8。

• 我們想要估計 95% 置信區間

服用藥物組的平均得分，如

以及服用安慰劑的小組。

溶液：

• N=20，因此 DF=19。要使用的 t 值為 2.093\*。

• 組的均值標準誤差 （SEM）

M 和 P 是：

或

最後，95% 的置信區間是

或

20

4.8

,

20

2.9 0.6485, 1.0733

10.82.0930.6485， 14.02.0931.0733

（9.44， 12.16）， （11.75， 16.25）

將向您展示如何

稍後再計算。

簡短摘要

• 很明顯，服用藥物的兒童是

顯然得分較低，暗示

提高注意力。

• 但是，兩者之間有一些重疊

兩個間隔。這稍微模糊了

這種藥物的有效性。

（9.44， 12.16）， （11.75， 16.25）

服用藥物 服用安慰劑

要麼有些人服用藥物可能沒有改善，要麼有些人沒有服用

藥物可以顯示改善。

示例 3

• 隨機抽取 81 名工人

公司表明他們工作

平均每月 100 小時，其中

標準偏差為27小時。

• 即 x̅=100，s=27 和 n=81。

• 計算 95% 置信區間

所有工人每月的平均工時

公司 工作。

答

• SEM = 27/sqrt（81） = 3。

• 所以 1.99SEM=5.97。（t 1.99

表，df= 80）

• 間隔將從1005.97=

94.03 到 100+5.97=105.97。

df T df T

1 12.706 12 2.179

2 4.303 13 2.160

3 3.182 14 2.145

4 2.776 15 2.131

5 2.571 20 2.086

6 2.447 25 2.060

7 2.365 30 2.042

8 2.236 40 2.021

9 2.262 60 2.000

10 2.228 120 1.980

11 2.201 ~ 1.960

示例 4

• 考慮以下隨機樣本

4 個觀測值 （n=4） 25， 47， 32， 56.

• 假設人口正常

分散式。請建譯 95%

均值的置信區間。

df T df T

1 12.706 12 2.179

2 4.303 13 2.160

3 3.182 14 2.145

4 2.776 15 2.131

5 2.571 20 2.086

6 2.447 25 2.060

7 2.365 30 2.042

8 2.236 40 2.021

9 2.262 60 2.000

10 2.228 120 1.980

11 2.201 ~ 1.960

答

• 首先計算此範例的平均值，

即 40，然後計算

標準偏差 s =14.07。

• 計算 s/sqrt（n=4）=14.07/2=7.035。

• 由於樣本量小，t-

校正適用，t=3.182

（df=3）。

• 因此間隔將來自

403.182\*7.035 = 17.615 至

40+3.182\*7.035 = 62.385。

MATLAB 函數 TPDF

TPDF 概率 密度函數 （pdf）

對於學生的 T 分佈

Y = TPDF（X，V） 傳回

學生的 T 分佈

V 自由度，在

X 中的值。

>> z=[-4：0.1：4];

>> X1=1/（sqrt（2\*pi））\*exp（-0.5\*z.^2）;

>>圖（z，X1）

>> A 標準正態分佈

或者您可以使用

X1 = normpdf（z）

>> t=[-4：0.1：4];

>> X2=tpdf（t，120）;

>>圖（t，X2）

>>

當 DF=120 時，這是

“幾乎相同”

標準正常

分配。

>> X2=tpdf（t，120）;

>> X3=tpdf（t，30）;

>> X4=tpdf（t，15）;

>> X5=tpdf（t，10）;

>> X6=tpdf（t，5）;

>>圖（t，X2，t，X3，t，X4，t，X5，t，X6）被“步進”

當 DF 下降時

MATLAB Function TCDF

TCDF學生的T累積分佈

函數 （cdf）。

P = TCDF（X，V） 計算

V 度為

自由，在X中的價值觀。P 的大小為

X 和 V 的常見大小。標量輸入

函數為相同的常量矩陣

大小作為其他輸入。

MATLAB 功能 TINV

TINV逆學生T累積

分佈函數 （cdf）。

X=TINV（P，V） 傳回

具有V學位的學生T cdf

的自由，在P中的價值觀。

>> 錫（0.95，inf）

年 =

1.6449

>> 錫（0.975，inf）

年 =

1.9600

>> 錫（0.995，inf）

年 =

2.5758

90% 間隔

（每邊5%）

95% 間隔

（每邊2.5%）

99% 間隔

（每邊0.5%）

各種 t 值（不同

DFs） 表示 95% 置信區間

>> tinv（0.975，inf） ans = 1.9600

>> tinv（0.975，120） 年 = 1.9799

>> tinv（0.975.60） 年 = 2.0003

>> tinv（0.975.30） 年 = 2.0423

>> tinv（0.975，15） 年 = 2.1314

>> tinv（0.975，10） 年 = 2.2281

>> tinv（0.975.5） 年 = 2.5706

df T df T

1 12.706 12 2.179

2 4.303 13 2.160

3 3.182 14 2.145

4 2.776 15 2.131

5 2.571 20 2.086

6 2.447 25 2.060

7 2.365 30 2.042

8 2.236 40 2.021

9 2.262 60 2.000

10 2.228 120 1.980

11 2.201 ~ 1.960

2.306

>> 錫（0.975，8）

年 =

2.3060

各種 t 值（不同

DFs） 表示 90% 置信區間

>> 錫（0.95，inf）

>> 錫（0.95，120）

>> 錫（0.95，60）

>> 錫（0.95，30）

>> 錫（0.95，15）

>> 錫（0.95，10）

>> 錫（0.95，5）

以下 MATLAB 命令

生成各種 t 值

自由度。

>>數據集=[25 47 32 56];

>> datamean=mean（dataset）

數據含義 = 40

>> L=長度（資料集）

L = 4

>> q = tinv（0.975，L-1）\*std（dataset）/sqrt（L）

q = 22.3905

>> 數據意義-q

年 = 17.6095

>> 數據含義+q

年 = 62.3905

>>

示例 4 - 重新存取

n

s

x t

標準誤差

手段

MATLAB 函數 參數

參數置信區間。

CI = PARAMCI（PD） 傳回一個 2 x N 陣列

CI 包含 95% 置信區間

概率分佈 PD 的參數。

CI = PARAMCI（PD， ALPHA） 傳回 100\*（1-

ALPHA）% 置信區間。預設值為

0.05 表示 95% 置信區間。

MATLAB 函數 擬合器

FITDIST 擬合到數據的概率分佈。

PD = FITDIST（X，DISTNAME） 適合

概率分佈 DISTNAME 到 中的數據

列向量 X，並返回一個物件

表示擬合分佈的 PD。 PD是一個

物件從 ProbDist 類派生的類。

DISTNAME 是此發行版的名稱，

可以是“二項式”，“正常”，“泊松”等。

>>數據集=[25 47 32 56];

>> PD=fitdist（dataset， 'normal'）;

???在 102 處使用 ==>擬合度時出錯

X 必須是數位列向量。

>> X=資料集'

>> PD=fitdist（X， 'normal'）;

>> CI=paramci（PD， 0.05）

CI =

17.6095 7.9712

62.3905 52.4653

>>

示例 5

• 構建 95% 置信區間來估計平均值

以下數字的值：

1903 1935 1910 2496 2108 1961 2060

1444 1612 1316 1511

• 我們需要首先評估分佈

以決定使用 t、z 或無。

>> X=[1903 1935 1910 2496 2108 1961 2060 1444 1612

1316 1511];

>>歷史（X）

>>平均值（X）

年 =

1.8415e+003

>>標準（X）

年 =

342.7373

>>

x̅=1841.5

s = 342.7

n = 11

近似正態分佈。

>> M=均值（X）;

>> S=std（X）;

>> N=11;

>> SEM=S/sqrt（N）

掃描電鏡 = 103.3392

>> T=錫（0.975， N-1）

T = 2.2281

>> 下限=M-T\*SEM

下限 = 1.6112e+003

>> 上部=M+T\*SEM

上限 = 2.0717e+003

>>

n

s

x t

答：

CI = [1611.2 2071.7]

使用 fitdist 和 paramci

>> X=[1903 1935 1910 2496 2108 1961 2060 1444

1612 1316 1511]';

>> PD=fitdist（X， 'normal'）;

>> CI=paramci（PD， 0.05）

CI =

1.0e+003 \*

1.6112 0.2395

2.0717 0.6015

>>

同樣的答案：

CI = [1611.2 2071.7]

章節回顧 - 1

• 解釋點和點之間的區別

總體均值  的區間估計。

• 什麼是均值的標準誤差 （SEM）？

• 描述

人口平均值 。間隔如何

解釋？

• 影響長度的因素有哪些

樣本平均值的置信區間？

• 單側置信區間與雙面置信區間？

章節回顧 - 2

• 描述相似之處和不同之處

在 t 分佈和標準之間

正態分佈 （z）。

• 如果您嘗試建構 CI，則當

你會使用一個而不是另一個（z vs

t)?