

芯片制程統計基礎篇

统计之六爻(PPSSVD)

SRS, Population, Statistics and Bias

王不老說半导

芯片制程統計基礎篇

统计之六爻(PPSSVD)

SRS, Population, Statistics and Bias

王不老說半导

世界其實很小：六次握手規則

- The six-handshake rule: 地球上的任何人，都或多或少彼此有些關聯，例如
- 王不老隨機找個路人A，竟然就可以經過A的關係網(B們)，在經過他的關係網的關係網(C們)...如此最多六次(理論上是~6.6)左右的握手介紹(A→Bs → Cs → Ds → Es → Fs)，最後可以連絡上原來毫無一毛錢關係的馬雲



其實中國古人早就知道了：八卦六爻

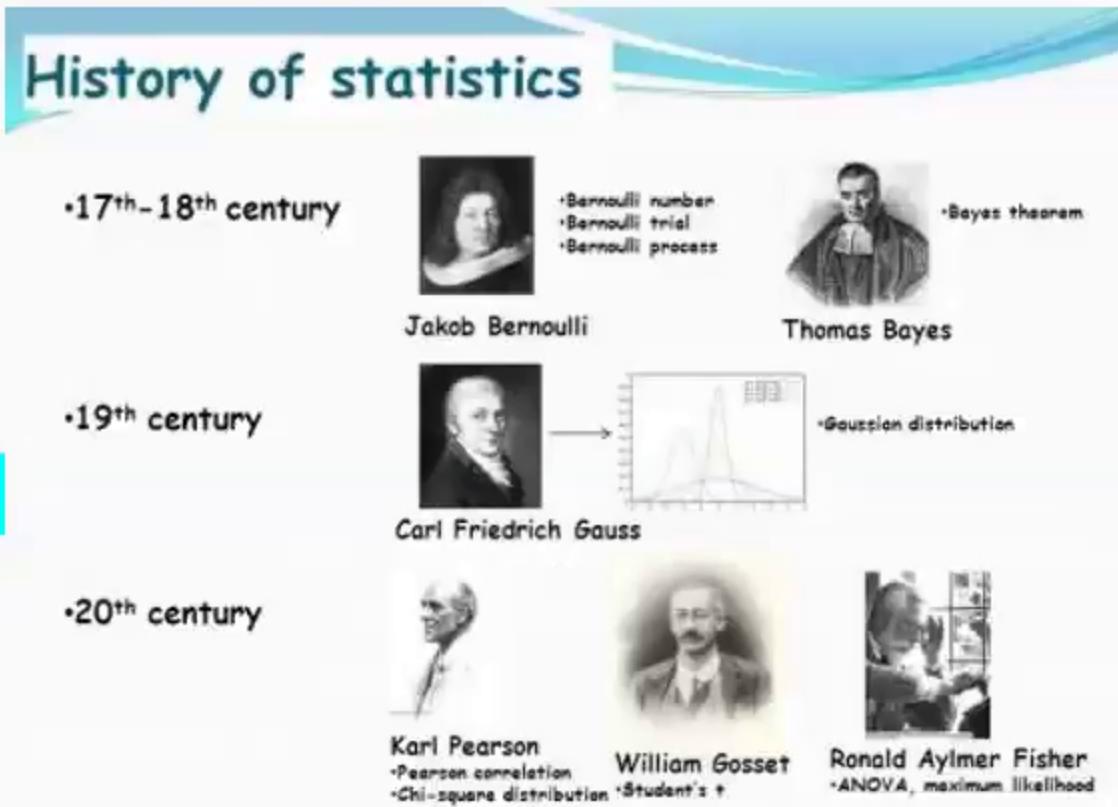
六爻歌(爻音搖：乾三連、坤六斷、兌上缺、巽下斷、離中虛、坎中滿、震仰盂、艮覆碗)

- 八卦其實乃我們中國古人藉觀察大自然，學習其中的法則，進而藉此演譯人間世(解說與預測驅吉避兇)的一種獨特的統計分析手段
- 現代統計學其實也有六爻 (= PPSSVD)，我們用此架構來分析並優化少量所謂代表性數據，並以之得出整體面貌



Statistics 的起源

- Statistics 字源 = Stat(e) + istics
 - Stat(e) = systematic collection of demographic/economic data by *states*
 - istics = 高大上的字根(代表特性, e.g., character → characteristics)
- 与数学不同, 统计学起源于17-18世纪, 方便执政者→统治麾下百姓
 - 应用在人口 (population) 和经济数据 (economic data) 的收集与分析
 - 抽税, 打仗, 赌博
 - 插播: 明朝灭亡原因 = 税收不公, 银子太多而不均(皇帝不懂统计被忽悠)



統計學六爻 (PPSSVD)

- PPSSVD:

Sample (样本, 抽樣, 子體?)	Population (总体, 母體)
Statistics (统计量)	Parameters (参量, 参数),
Data (數據)	data
Variables (变量, numerical or categorical.)	Variables (变量, numerical or categorical)

- 有猫腻的V:变量

- Numerical variable:**数值**变量，大家熟知的數的變化，“可量得的”量”，所以有數量單位
- Categorical variable:**分类**变量，“分门别类”竟然也算一種变量，例如性别，不同的颜色等等，他不好被量化，所以沒有數量單位

統計學六爻 (PPSSVD)

- PPSSVD:

Sample (样本, 抽樣, 子體?)	Population (总体, 母體)
Statistics (统计量)	Parameters (参量, 参数),
Data (數據)	data
Variables (变量, numerical or categorical.)	Variables (变量, numerical or categorical)

6

- 有猫腻的V:变量

- Numerical variable: **数值**变量，大家熟知的數的變化，“可量得的”量”，所以有數量單位
- Categorical variable: **分类**变量，“分门别类”竟然也算一種变量，例如性别，不同的颜色等等，他不好被量化，所以沒有數量單位

PPSSVD: Parameters are for Population

Population 有两大类:

- (I)量雖大，但大約是清楚的，稍微穩定，與時間的變化不是太大
- (II)量也大(包含过去现在及未来的量)，與時間的變化相當大些

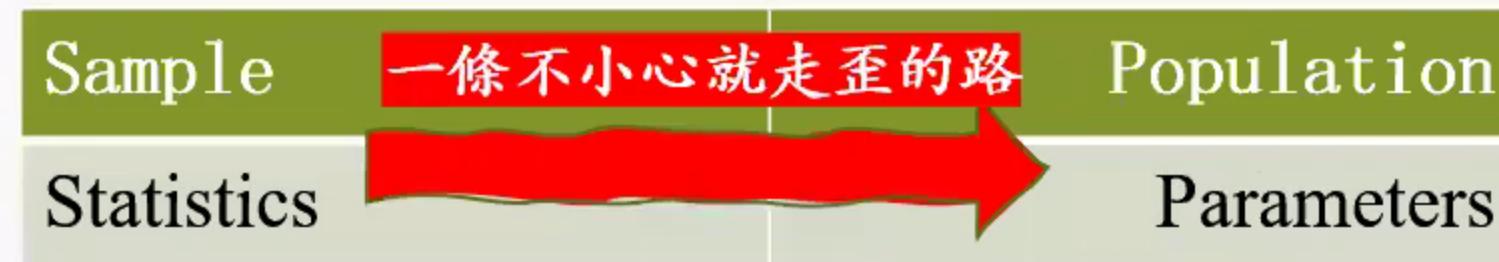
Parameters: 參數

- Population 内比较有代表性特质的某種量化(numerical characteristics of the population)

Population I	Parameters	Population II	Parameters
北京人口	平均身高 平均收入 平均结婚年龄	京东面板 中芯晶片 病毒疫苗	平均良率 平均销售额 平均有效率

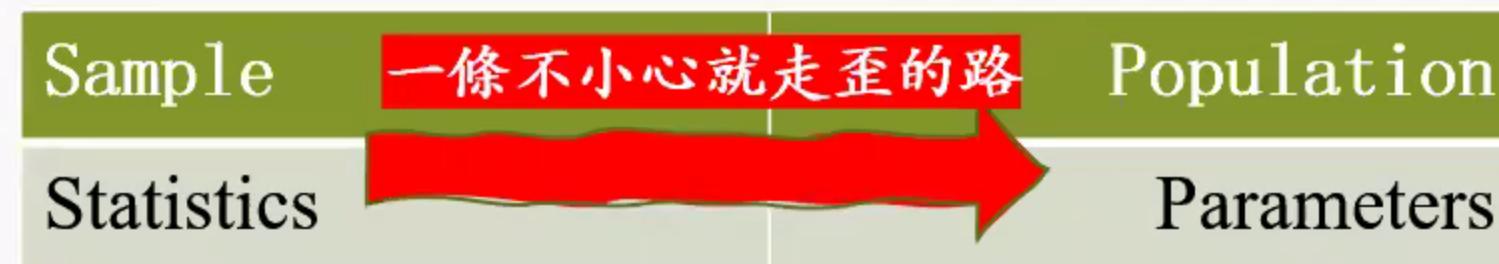
PPSSVD: Sample and Statistics

- Sample: 就是偏要見微知著，偏要以管窺天(因為時不我予而且鈔票也不夠)，但是盜亦有道，這"微"，這"管"卻有諸般講究，
 - 可以偷懶(樣本數可以商量)，但不能有私心(必須隨機抽樣)，所以是給你方便，但不是隨便的SRS (簡單隨機抽樣)，然而看似簡單，但知易行難
 - *A simple random sample (SRS) of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.*
- Statistics: 以特異招數分析收集到高質量的"微"，最後找出那見微知著的"著"，此過程，吾人曰之 "statistical inference" "(統計推斷)"
 - 若收集到的是偏心的"微"，那預測的結果就是牛頭不對馬嘴



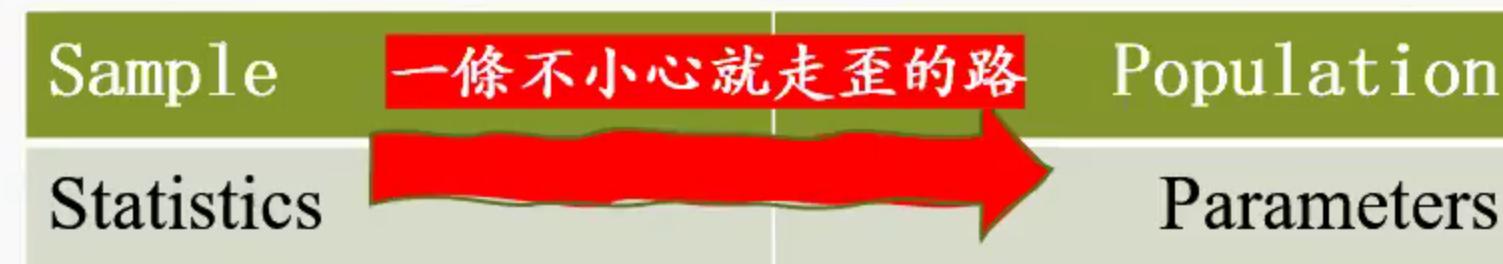
PPSSVD: Sample and Statistics

- Sample: 就是偏要見微知著，偏要以管窺天(因為時不我予而且鈔票也不夠)，但是盜亦有道，這"微"，這"管"卻有諸般講究，
 - 可以偷懶(樣本數可以商量)，但不能有私心(必須隨機抽樣)，所以是給你方便，但不是隨便的SRS (簡單隨機抽樣)，然而看似簡單，但知易行難
 - *A simple random sample (SRS) of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.*
- Statistics: 以特異招數分析收集到高質量的"微"，最後找出那見微知著的"著"，此過程，吾人曰之 "statistical inference" "(統計推斷)"
 - 若收集到的是偏心的"微"，那預測的結果就是牛頭不對馬嘴



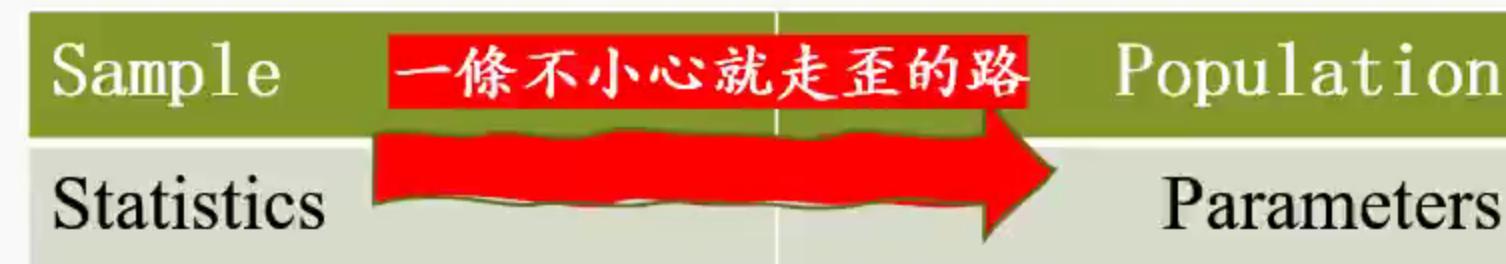
PPSSVD: Sample and Statistics

- Sample: 就是偏要見微知著，偏要以管窺天(因為時不我予而且鈔票也不夠)，但是盜亦有道，這"微"，這"管"卻有諸般講究，
 - 可以偷懶(樣本數可以商量)，但不能有私心(必須隨機抽樣)，所以是給你方便，但不是隨便的SRS (簡單隨機抽樣)，然而看似簡單，但知易行難
 - *A simple random sample (SRS) of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.*
- Statistics: 以特異招數分析收集到高質量的"微"，最後找出那見微知著的"著"，此過程，吾人曰之 "statistical inference" "(統計推斷)"
 - 若收集到的是偏心的"微"，那預測的結果就是牛頭不對馬嘴



PPSSVD: Sample and Statistics

- Sample: 就是偏要見微知著，偏要以管窺天(因為時不我予而且鈔票也不夠)，但是盜亦有道，這"微"，這"管"卻有諸般講究，
 - 可以偷懶(樣本數可以商量)，但不能有私心(必須隨機抽樣)，所以是給你方便，但不是隨便的SRS (簡單隨機抽樣)，然而看似簡單，但知易行難
 - *A simple random sample (SRS) of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.*
- Statistics: 以特異招數分析收集到高質量的"微"，最後找出那見微知著的"著"，此過程，吾人曰之 "statistical inference" "(統計推斷)"
 - 若收集到的是偏心的"微"，那預測的結果就是牛頭不對馬嘴



王不老開餐廳

試問:王不老愛燒菜，每次燒菜都拿去請同事鄰居吃，人人稱讚，退休無所事事，竟然突發奇想開餐廳，結果只紅了一陣子就收了攤，**它百思不得其解，明明人人稱讚，為何後來都沒了下文？**

解答:公正抽样真的很難，因為偏心很簡單

Bias (社會學的，我們的常識) is defined as “prejudice in favor of or against one thing, person, or group compared with another, usually in a way considered to be unfair.”

統計偏差(**Statistical Bias**)包括：

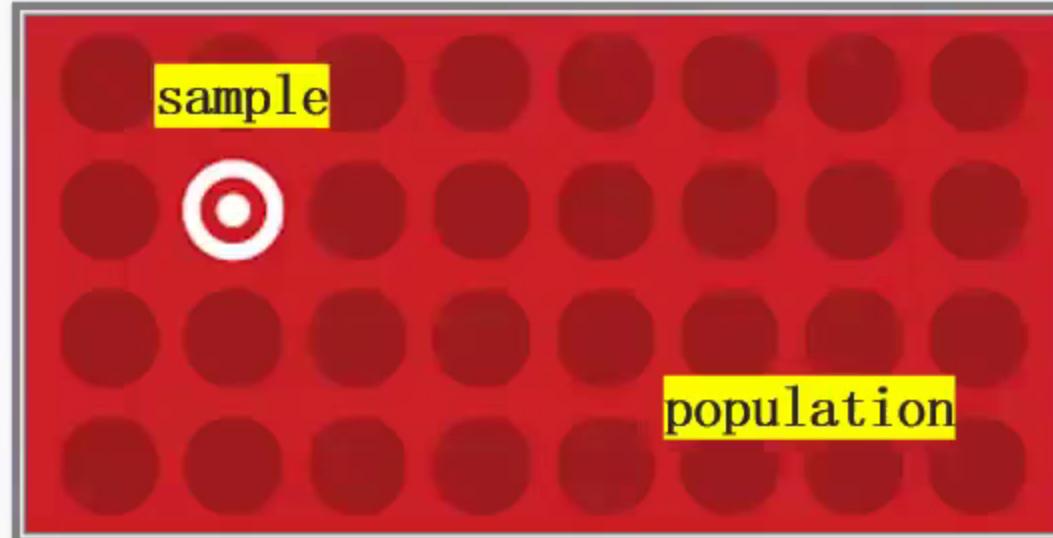
1. Selection bias
2. Survivorship bias
3. Omitted variable bias
4. Recall bias
5. Observer bias
6. Funding bias (給錢就是大爺)
7. Hindsight Bias (馬後炮)

這見微知著的"微"之諸般講究

以知: Population (母體) = 中国所有賣川菜的餐廳

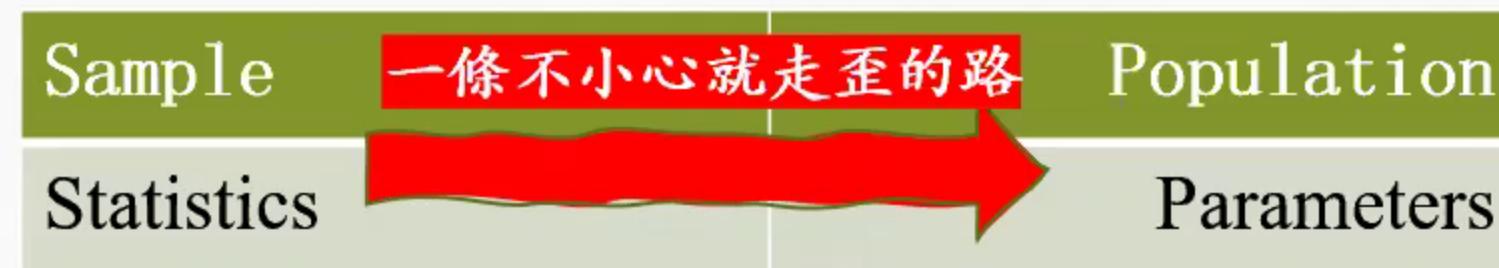
試問: 以SRS的精神, 以搖獎方式(表示公平)抽取某樣本(目標是要了解以上母體), 以下何者比較正確?

1. 搖獎取樣本1, 將中獎者放回, 再搖獎取樣本2
2. 搶獎取樣本1, 不將中獎者放回, 再搖獎取樣本2



PPSSVD: Sample and Statistics

- Sample: 就是偏要見微知著，偏要以管窺天(因為時不我予而且鈔票也不夠)，但是盜亦有道，這"微"，這"管"卻有諸般講究，
 - 可以偷懶(樣本數可以商量)，但不能有私心(必須隨機抽樣)，所以是給你方便，但不是隨便的SRS (簡單隨機抽樣)，然而看似簡單，但知易行難
 - *A simple random sample (SRS) of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.*
- Statistics: 以特異招數分析收集到高質量的"微"，最後找出那見微知著的"著"，此過程，吾人曰之 "statistical inference" "(統計推斷)"
 - 若收集到的是偏心的"微"，那預測的結果就是牛頭不對馬嘴

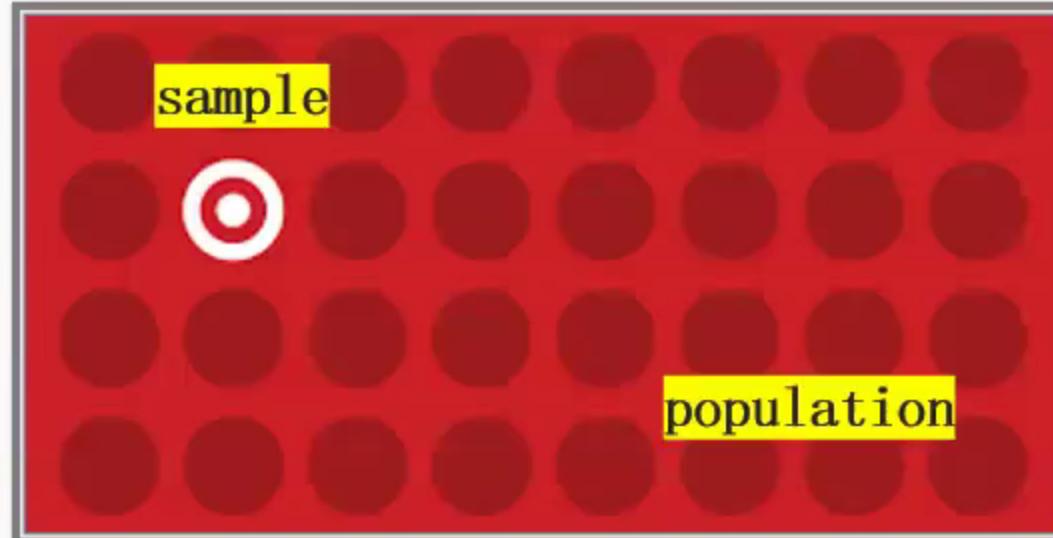


這見微知著的"微"之諸般講究

以知: Population (母體) = 中国所有賣川菜的餐廳

試問: 以SRS的精神, 以搖獎方式(表示公平)抽取某樣本(目標是要了解以上母體), 以下何者比較正確?

1. 搖獎取樣本1, 將中獎者放回,
再搖獎取樣本2
2. 搶獎取樣本1, 不將中獎者放回,
再搖獎取樣本2

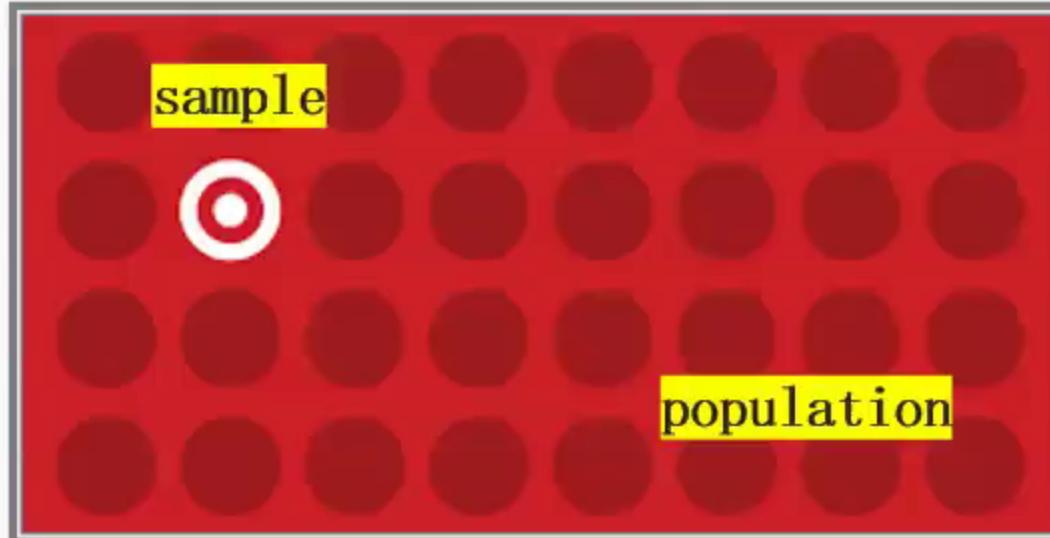


這見微知著的"微"之諸般講究

以知: Population (母體) = 中国所有賣川菜的餐廳

試問: 以SRS的精神, 以搖獎方式(表示公平)抽取某樣本(目標是要了解以上母體), 以下何者比較正確?

1. 搖獎取樣本1, 將中獎者放回,
再搖獎取樣本2
2. 搶獎取樣本1, 不將中獎者放回,
再搖獎取樣本2



简单随机抽样SRS 準則

SRS: 每个样本单位被抽中的概率相等，样本的每个单位完全独立，彼此间无一定的关联性和排斥性

- **重复抽样**: 每次抽中的单位仍放回总体(**with replacement**)，样本中的单位可能不止一次被抽中
- **不重复抽样**: 抽中的单位不再放回总体(**without replacement**)，样本中的单位只能抽中一次。社会调查采用不重复抽样。

但简单随机抽样的必须有一个完整的**抽样框**，即总体各单位的清单。总体太大时，制作这样的抽样框工作量巨大，加之有许多情况，使总体名单根本无法得到(若題目改為**世界所有賣川菜的餐廳**)。故在大规模社会调查中很少采用纯随机抽样

- 所幸在硅芯片制程世界中，這個問題不大，因為機台會不停地收集各種制程訊息，而且已經擁有許多所謂baseline資料，我們關切的其實是背離此baseline的諸般變化

閒聊：包公判案有無偏心(Bias)?

- 无罪开释 真-正阴 $(1-\alpha)$ 包公无罪释秦香莲
- 碰瓷陷罪 假-假阳 (α)
- 助人脱罪 假-假阴 (β)
- 罪有应得 真-正阳 $(1-\beta)$ 包公怒铡陈世美

凡人断案时不准

		无罪?	有罪?
神仙断案金标准	无罪	无罪开释 真-正阴 $(1-\alpha)$ True Negative	碰瓷陷罪 假-假阳 (I,α) False Positive
	有罪	助人脱罪 假-假阴 (II,β) False Negative	罪有应得 真-正阳 $(1-\beta)$ True Positive

课程抽考

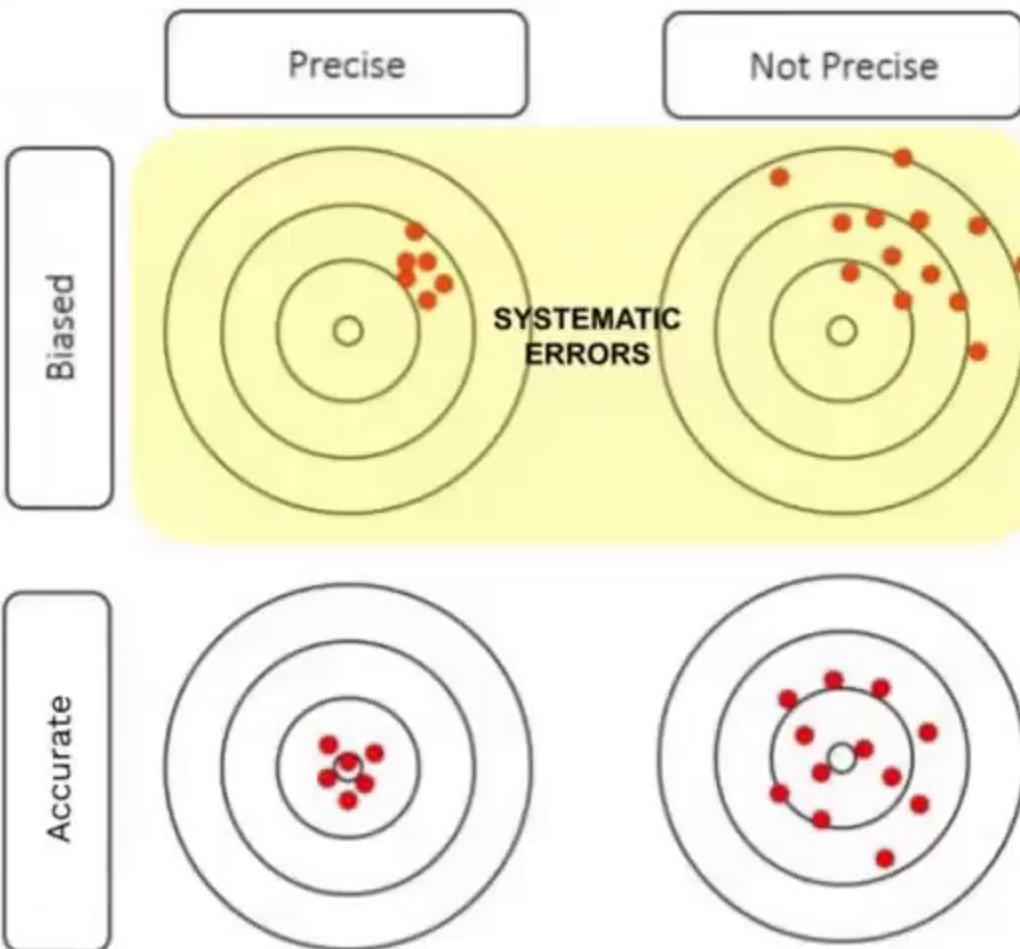
- 秦香莲是____，陈世美是____
- 辛普森谋杀案是____ $(\alpha?\beta?)$

話說：芯片SPC有關的Bias

Bias = systematic error (不是前面所說的常識偏差，而是與制程有關的偏差)

Statistical Process Control (SPC) is to enable HMV (high volume manufacturing) by

- Identifying and **removing** bias (要徹底幹掉它)
- Identifying and **reducing** (cannot eliminate) random error (有些無奈，但要想法子把它縮小)

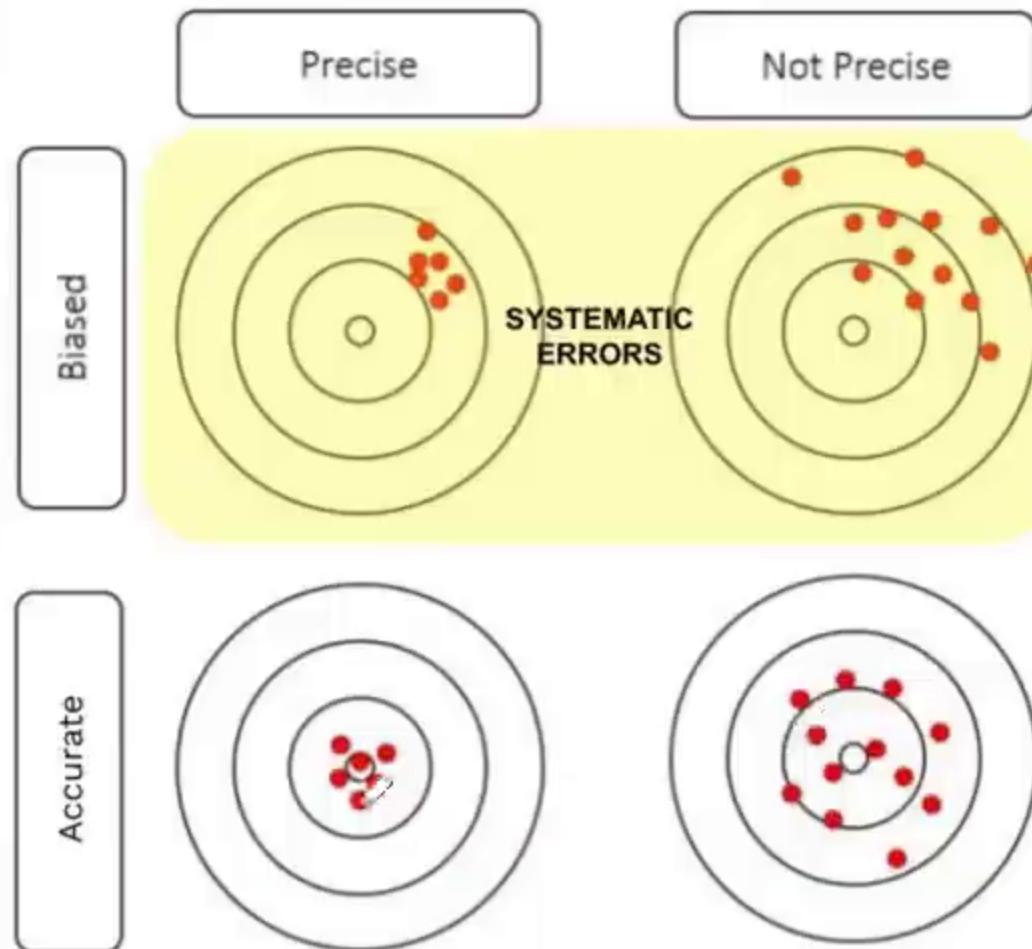


話說：芯片SPC有關的Bias

Bias = systematic error (不是前面所說的常識偏差，而是與制程有關的偏差)

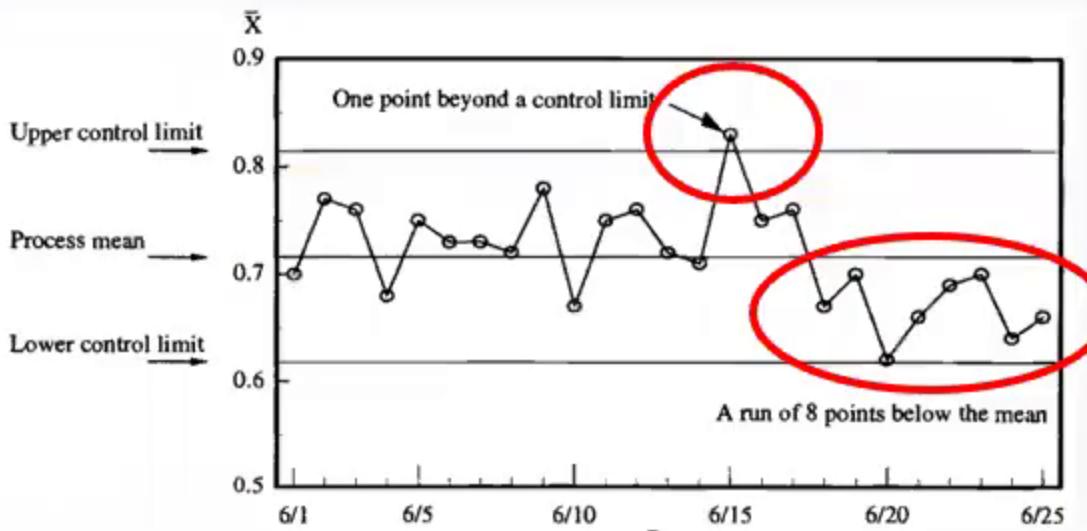
Statistical Process Control (SPC) is to enable HMV (high volume manufacturing) by

- Identifying and **removing** bias (要徹底幹掉它)
- Identifying and **reducing** (cannot eliminate) random error (有些無奈，但要想法子把它縮小)



晶片良率兩大殺手: WIW and WTW

- WTW (wafer to wafer uniformity, 晶圆间均匀度) 的要求，是以SPC控制图(如右)監控每一個重要製程，能夠保持品質在穩定範圍(UCL/LCL)之內
- 若WTW不稳定，例如出現許如右圖所示紅圈(bias)，工程机器必须暂时停止运转，修正毛病以后，才能够继续开工，影响出货的速度甚至品質
- 製程不健康，則根本無法量產



附錄：晶片良率兩大殺手：WIW and WTW

已知：右表為某半導製程收集的統計數據

試問：選擇其中各標準差 σ ，計算WTW最適當
UCL/LCL參數值($\mu \pm 3\sigma$)， $\mu=8000\text{Å}$

解答： σ_3

- 我們知道標準差 $\sigma_{\bar{x}} = \frac{3\sigma}{\sqrt{n}}$
- 但是在右表有四個標準差，哪一個才是對的？
- 首先必須排除的是WIW (σ_1)，因WTW與WIW的 σ ，有根本物理本源及統計觀念上的不同
 - WIW是在反應室裡面，各式物化反應造成晶圓內的不均勻(across the wafer variation)
 - 但是WTW卻是有關長期晶圓間均勻度的漂移(a long-term drift of the wafers)
 - 雖每一數據是取此晶圓內的平均值

WIW σ_1	9Å
WTW σ_2 (10 consecutive runs)	6Å
WTW σ_3 (3 months average)	15Å
σ_4 from all data including WIW	17Å

解答：

- 其次必須排除的是WTW (σ_2)，10個晶圓間基本上太短(not representative)，這短短的時間取得的數據，不能代表整體(統計約之為 cannot infer)
- σ_4 包含WIW，所以必須排除
- σ_3 為適當(because they are from a long-term representative drift of the designated wafers)，therefore UCL/LCL = $8000 \pm 3 \times 15\text{Å}$

附錄：晶片良率兩大殺手：WIW and WTW

已知：右表為某半導製程收集的統計數據

試問：選擇其中各標準差 σ ，計算WTW最適當
UCL/LCL參數值($\mu \pm 3\sigma$)， $\mu=8000\text{Å}$

解答： σ_3

- 我們知道標準差 $\sigma_{\bar{x}} = \frac{3\sigma}{\sqrt{n}}$
- 但是在右表有四個標準差，哪一個才是對的？
- 首先必須排除的是WIW (σ_1)，因WTW與WIW的 σ ，有根本物理本源及統計觀念上的不同
 - WIW是在反應室裡面，各式物化反應造成晶圓內的不均勻(across the wafer variation)
 - 但是WTW卻是有關長期晶圓間均勻度的漂移(a long-term drift of the wafers)
 - 雖每一數據是取此晶圓內的平均值

WIW σ_1	9Å
WTW σ_2 (10 consecutive runs)	6Å
WTW σ_3 (3 months average)	15Å
σ_4 from all data including WIW	17Å

解答：

- 其次必須排除的是WTW (σ_2)，10個晶圓間基本上太短(not representative)，這短短的時間取得的數據，不能代表整體(統計約之為 cannot infer)
- σ_4 包含WIW，所以必須排除
- σ_3 為適當(because they are from a long-term representative drift of the designated wafers)，therefore UCL/LCL = $8000 \pm 3 \times 15\text{Å}$

芯片制程統計基礎篇

统计之六爻(PPSSVD)

SRS, Population, Statistics and Bias

王不老說半导