

# 統計分析

## 第一講

王寧寧, Ph.D

[oliningning@qq.com](mailto:oliningning@qq.com)

2022/10/24



# 課程安排

- 第一講：基本原理；參數估計；單變量描述性分析
- 第二講：假設檢驗；t檢驗；方差分析
- 第三講：卡方檢驗和非參數檢驗
- 第四講：相關與回歸，多變量分析

## 參考教材（可選）

### 《醫學統計學》

人民衛生出版社（第7版）主編：李康、賀佳



# 今天主要内容

- 醫學統計學的基本定義
- 單變量描述行分析
- SPSS的基本操作
- 中期教學檢驗

# 爲什麼要學習統計學

什麼是科學的問題？


証偽性


假設檢驗的證偽系統

# 醫學論文中關於統計學P值使用的爭議

**nature**

[View all Nature Research journals](#)

[Search](#) 

[Explore our content](#) 

[Journal information](#) 

[Subscribe](#)

[nature](#) > [comment](#) > [article](#)

**COMMENT** • 20 MARCH 2019

## Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

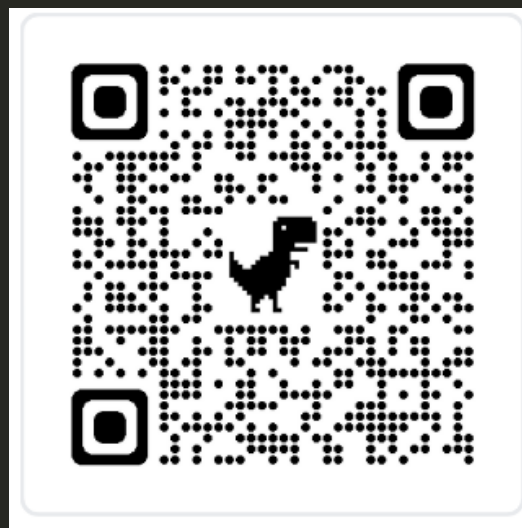


# 爭論的結果

- 反對者：“P值就是是個屁！”
- 支持者：“論文中，沒有P值，連屁都不是！”
- 假設檢驗是至今為止最有效的證偽系統
- 參考：<https://statmodeling.stat.columbia.edu/2019/03/20/retire-statistical-significance-the-discussion/>

# 爲什麼要學習統計

<http://opinion.people.com.cn/n1/2017/0609/c1003-29327938.html>



近年来，中药注射液一直饱受争议，甚至屡屡被“黑”，其使用范围也受到限制。根据2017版国家基本医保药品目录，限制基层使用的中药注射液品种从2009年的6个增加到26个，且只能在二甲以上医院使用。事实上，能否进入医保目录，决定着一个药品在市场上的生死。如果不进入医保目录，必然遭遇毁灭性打击。

中药注射液受限，是因为其安全性受到质疑。那么，中药注射液果真不安全吗？2015年全国药品不良反应监测网络显示，在发生不良反应的药品中，化学药占81.2%、中药占17.3%。与西药相比，中药注射液的不良反应率并不高。从用药原则来说，能口服不肌注、能肌注不输液，因为静脉注射风险最高，中药注射液也不例外。但是，中药注射液在临床上起效迅速，如果医生合理使用，病人获益大于风险。国家药品不良反应监测数据分析显示，基层成为中药注射液安全风险高发区，主要原因在于不合理使用。在中药注射液发生的不良反应中，合并用药超四成。这意味着，发生不良反应，并非单纯是中药注射液的问题。再加上基层医疗机构的救治设备、设施较有限，医护人员救治经验和能力相对缺乏，影响抢救效果及预后。因此，中药注射液的安全性问题主要是使用不当，而不完全是药物本身的毛病。

# 醫學統計學的定義

醫學統計學(medical statistics):臨床醫學、基礎醫學、公共衛生學和醫療衛生服務服務研究中的一門**基礎學科**，是關於搜集數據、分析數據和由數據得出結論的一組概念、原則和方法。

# 學習醫學統計學的目的

- 培養正確的統計思維
- 掌握基本的統計設計方法和如何收集準確可靠的數據
- 運用基本統計分析方法正確分析和處理數據
- 掌握操作統計軟件的基本技能
- 正確解釋和表達分析結果

# 變量和數據類型

## 定量數據：

也稱計量資料，變量的觀測值是定量的，其特點是能用數值的大小衡量其水平的高低，一般有計量單位。

## 定性數據：

也稱計數資料，變量的觀測值是定性的，表現為互不相容的類別或屬性。血型分為A、B、O、AB

## 有序數據:

也稱半定量數據或等級資料。變量的觀測值是定性的，但各類別（屬性）之間有程度或順序上的差別，如尿糖的化驗結果分爲-、+、++、+++。

統計方法的選用與數據類型有密切關係。根據分析的需要，不同類型的變量或數據之家可以進行轉換。



# 數據的描述性分析

## 例：140名成年男子紅細胞數

4.76	5.26	5.61	5.95	4.46	4.57	4.31	5.18
4.92	4.27	4.77	4.88	5.00	4.73	4.47	5.34
4.70	4.81	4.93	5.04	4.40	5.27	4.63	5.50
5.24	4.97	4.71	4.44	4.94	5.05	4.78	4.52
4.63	.....	5.02	4.76				

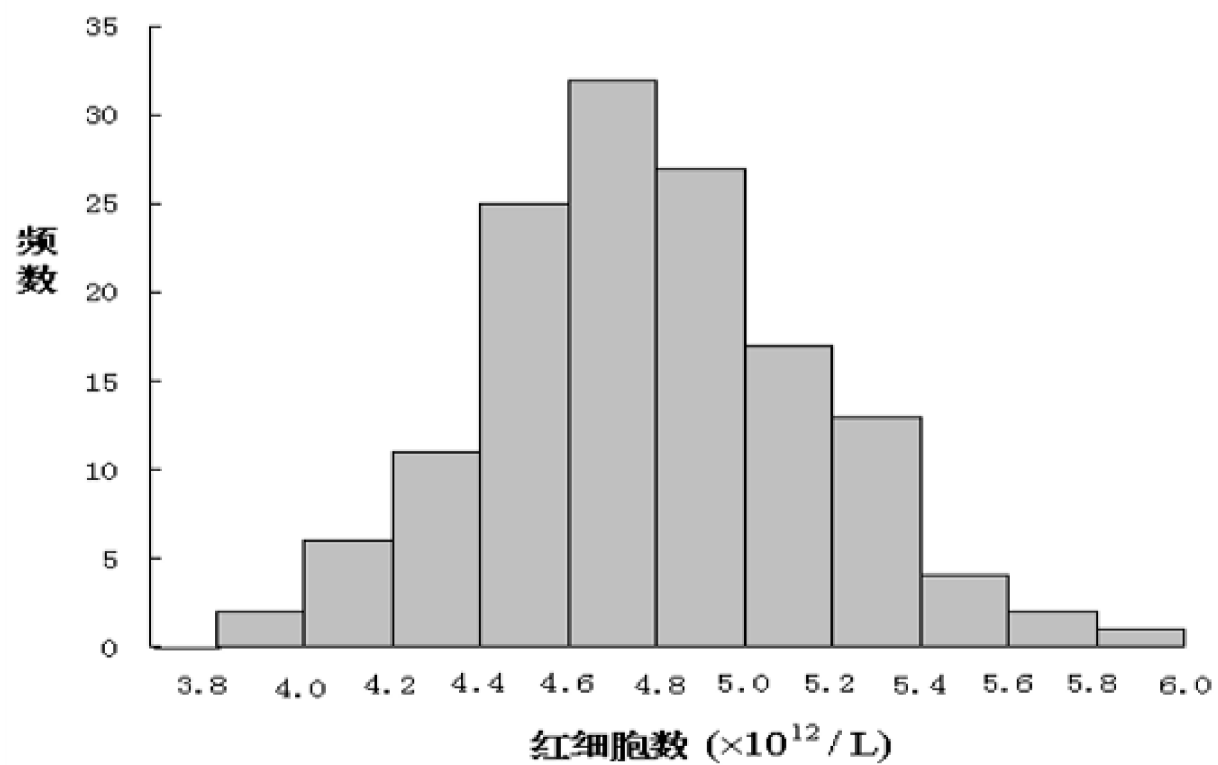
- 目的：了解數據分佈的情況
- 假定：來自正態分佈  $N(\mu, \sigma^2)$

# 頻數表

確定組數->確定組距-> 確定組限 ->確定頻數

紅細胞數	頻 數	頻 率 (%)	累積頻率 (%)
3.80~	2	1.4	1.4
4.00~	6	4.3	5.7
4.20~	11	7.9	13.6
4.40~	25	17.9	31.5
4.60~	32	22.9	54.4
4.80~	27	19.3	73.7
5.00~	17	12.1	85.8
5.20~	13	9.3	95.1
5.40~	4	2.9	98.0
5.60~	2	1.4	99.4
5.80~ 6.00	1	0.7	100.0
合計	140	—	100.0

# 直方圖



## 集中趨勢的指標

- 算術均值（mean）： $\bar{X}(\mu)$
- 中位數（median）： $M$
- 百分位數（percentile）
- 眾數（mode）

## 離散趨勢的指標

- 極差 (range) : R
- 四分位間距 (interquantile range, IQR)
- 方差 (variance) :  $S^2(\sigma^2)$
- 標準差 (standard deviation S.D.) :  $S(\sigma)$
- 變異係數(coefficient of variation, C.V.): 標準差/算術均值

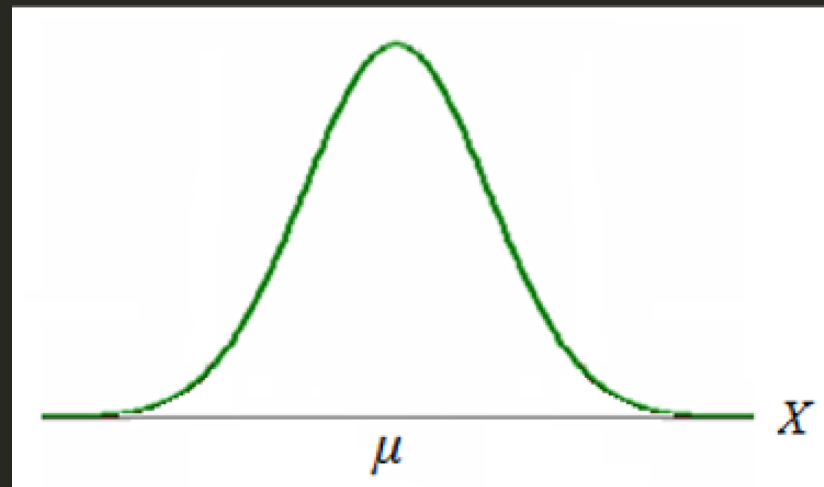
使用SPSS繪製140名成年男子紅細胞  
數的頻數表和直方圖並計算上述各個  
指標

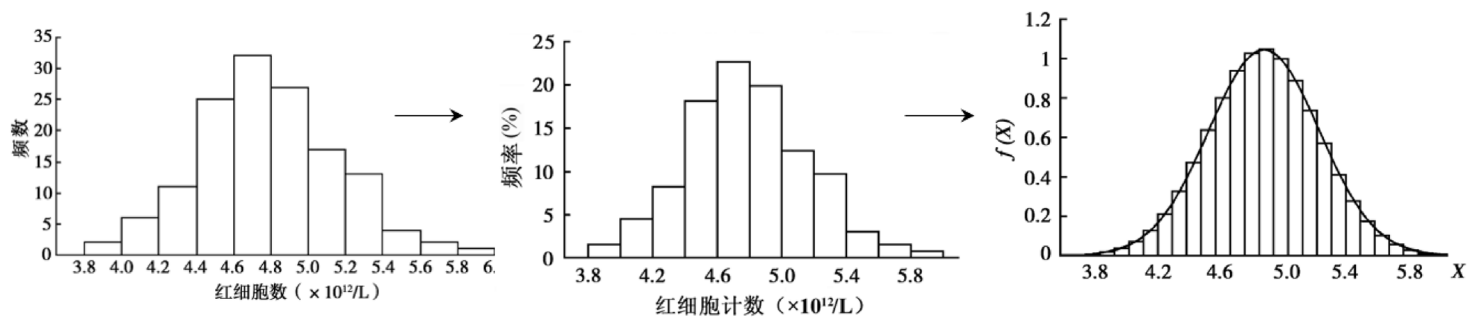
# 如何描述定量變量

- 判斷是否正態分佈
- 具體方法：頻數表、直方圖、QQ圖、正態檢驗
- 準則：
  - 正態分佈：均值（標準差）
  - 非正態分佈：中位數（四分位間距）



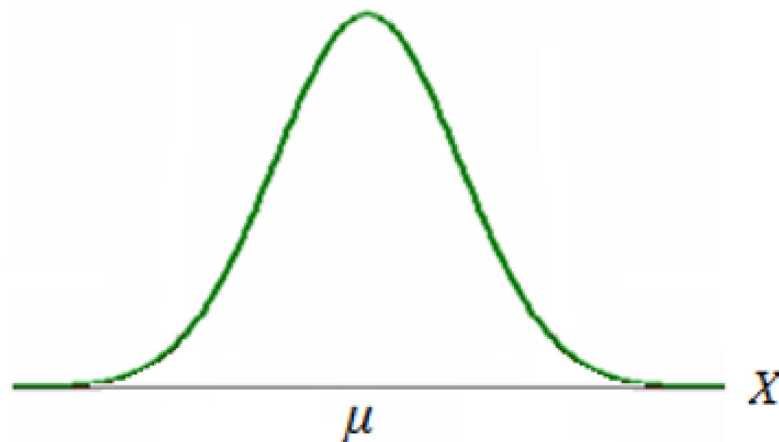
# 正態分佈





某地正常成年男子紅細胞數的分佈情況

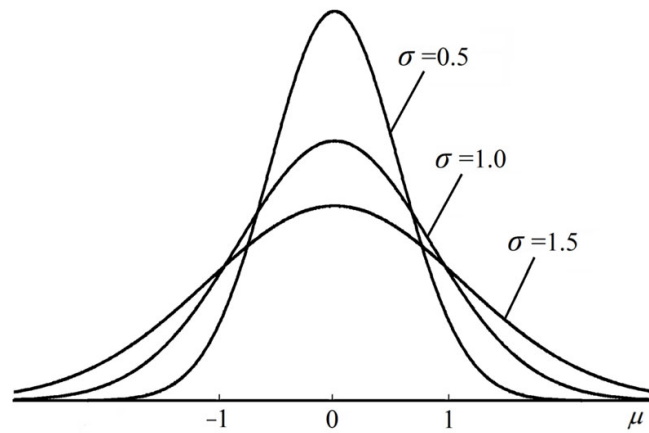
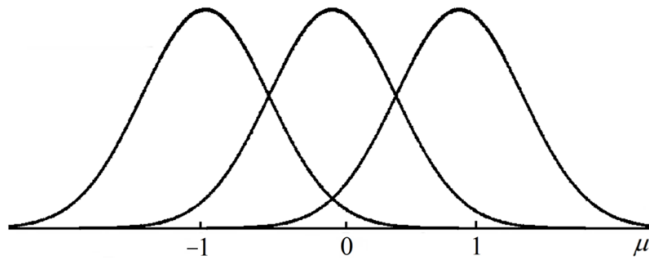
連續型隨機變量 $X$ 服從正態分佈:  $X \sim N(\mu, \sigma^2)$



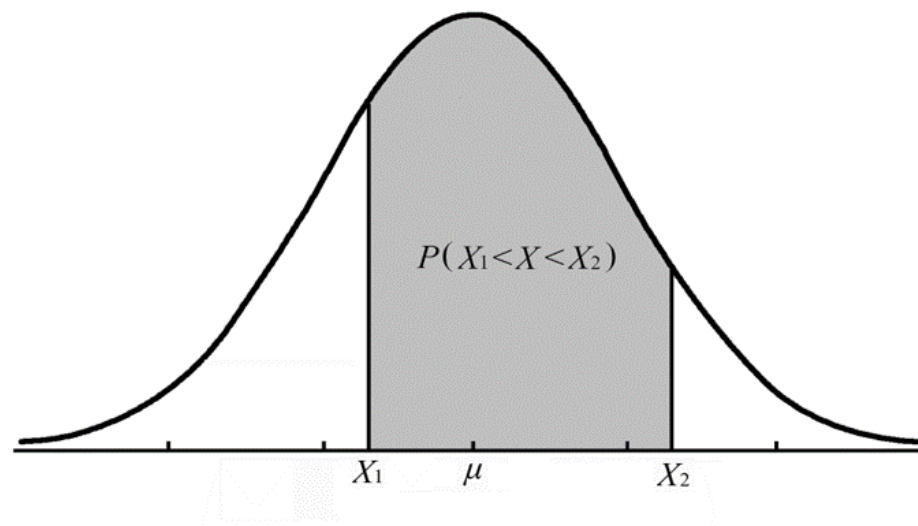
單峰鐘型分佈，以  $X = \mu$  為對稱中心

在  $X = \mu \pm \sigma$  有拐點

$\mu$  是位置參數，  $\sigma$  是形狀參數

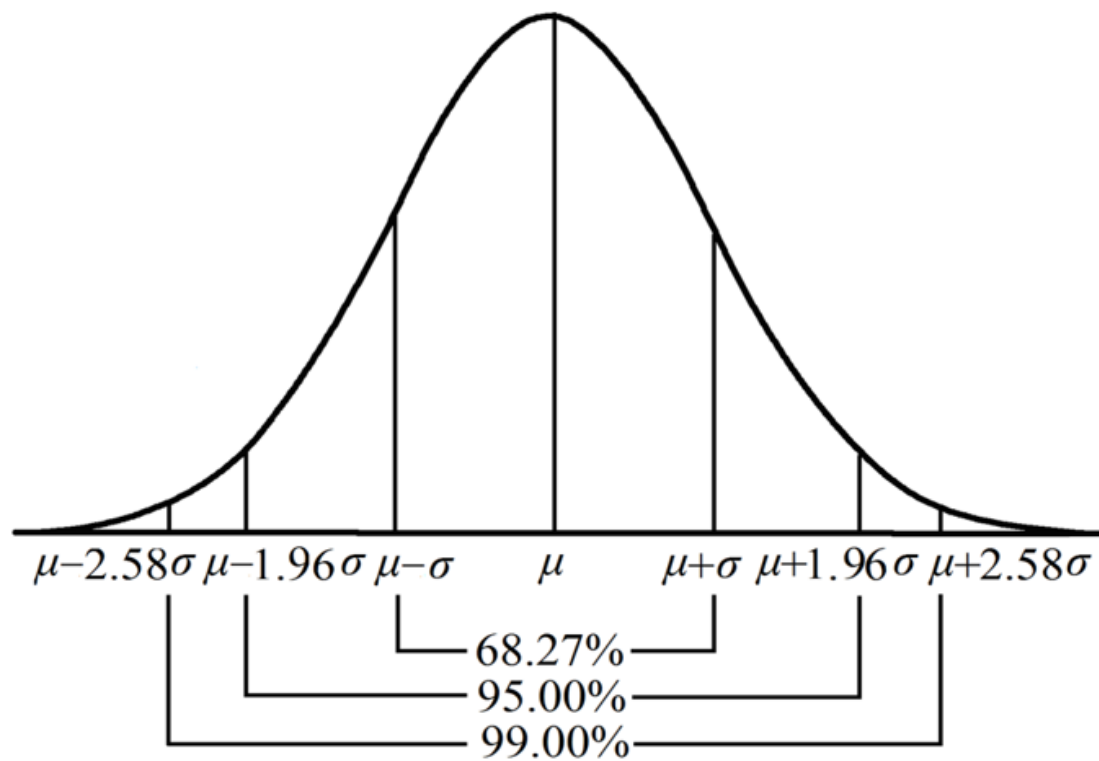


## 密度曲綫下方的面積含義



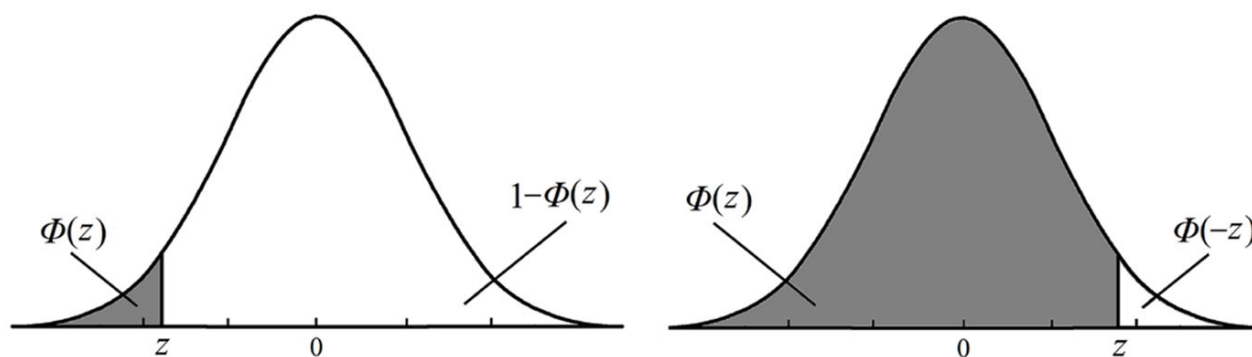
$$P(X_1 < X < X_2)$$

# 正態分佈密度曲線的分佈律



# 標準正態分佈

$\mu = 0, \sigma = 1$ 時:  $Z \sim N(0, 1)$



標準化:  $z = \frac{X - \mu}{\sigma}$

已知某地140名成年男子紅細胞計數近似服從正態分佈:

$$\bar{X} = 4.78 \times 10^{12}/L, S = 0.38 \times 10^{12}/L$$

求:

- 計算該地正常成年男子紅細胞計數在  $4.0 \times 10^{12}/L$  以下者佔該地正常成年男子總數的百分比。
- 計算紅細胞計數在  $4.0 \times 10^{12}/L \sim 5.5 \times 10^{12}/L$  佔該地正常成年男子總數的百分比。



# 醫學論文中常用的分析方法

# 單因素分析

**TABLE 1** Basal characteristics of patients with diminished ovarian reserve and comparison between poor ovarian responder (POR) and non-POR groups.

Characteristics	All (n=931)	POR (n=561)	Non-POR (n=370)	P value
Age, y	36.65 ± 4.96	38.05 ± 5.25	31.2 ± 4.24	<0.001
Duration of infertility, y	4.98 ± 3.86	5.11 ± 4.13	4.78 ± 3.42	0.183
BMI	22.53 ± 3.10	22.66 ± 3.10	22.33 ± 3.11	0.107
Basal FSH level, IU/L	10.61 ± 5.73	11.02 ± 5.72	9.98 ± 5.70	0.007
Basal LH level, IU/L	4.92 ± 5.84	5.29 ± 7.23	4.35 ± 2.42	0.005
Basal estradiol level, pg/mL	53.00 ± 55.12	51.53 ± 54.47	55.23 ± 56.10	0.319
AFC	4.14 ± 1.52	4.00 ± 1.48	4.36 ± 1.54	<0.001
AMH, ng/mL	0.54 ± 0.29	0.52 ± 0.29	0.57 ± 0.29	0.030
Type of infertility				0.001
Primary infertility	357 (38.3)	253 (45.1)	104 (28.1)	
Secondary infertility	574 (61.7)	308 (54.9)	266 (71.9)	
Cause of infertility				<0.001
Tubal factor	371 (39.9)	235 (41.9)	136 (36.8)	
Endometriosis	174 (18.7)	143 (25.5)	31 (8.4)	
Male factor	82 (8.8)	36 (6.4)	46 (12.4)	
Multiple causes	304 (32.7)	147 (26.2)	157 (42.4)	

<https://obgyn.onlinelibrary.wiley.com/doi/10.1002/ijgo.13174>

# 單因素分析

自变量	因变量	方法
定量	定量	简单相关分析/秩相关
定量	定性	T检验/方差分析/非参数
有序	有序	秩相关 /非参数
定性	定性	卡方检验
定性	定量	T检验/方差分析/非参数
定性	时间长度	Log-rank检验

# 多因素分析

- 多重綫性回歸模型
- logistics 回歸模型
- 泊松回歸
- COX回歸

# 多因素分析

**TABLE 3** Logistic regression model of the predictors of cumulative live birth rate (CLBR).

Variables	Regression coefficient	Standard error	P value	OR	95% CI
No. of good-quality embryos	0.275	0.108	0.011 <sup>a</sup>	1.316	1.068–1.631
Age	−0.142	0.018	<0.001 <sup>a</sup>	0.868	0.837–0.899
No. of oocyte retrieval cycles	−0.195	0.060	0.001 <sup>a</sup>	0.823	0.730–0.922
No. of transferrable embryos	0.015	0.095	0.873	1.015	0.841–1.222
AMH	0.310	0.315	0.325	1.363	0.735–2.526
AFC	0.069	0.059	0.249	1.071	0.953–1.204
Constant	3.324	0.718	<0.001 <sup>a</sup>	27.770	6.878–115.106

Abbreviations: CI, confidence interval; OR, odds ratio; CLBR, cumulative live birth rate; AMH, anti-Müllerian hormone; AFC, antral follicle count.

<sup>a</sup>P<0.05.

<https://obgyn.onlinelibrary.wiley.com/doi/10.1002/ijgo.13174>

謝謝大家