考試只考 Binomial 與 Poisson 分佈

Binomial distribution (二項分佈)

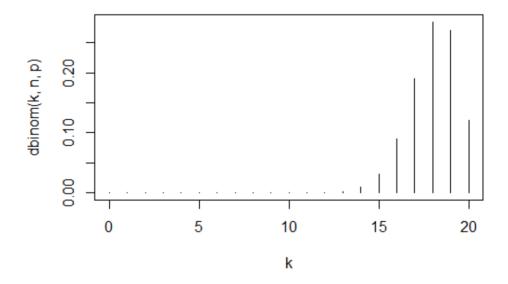
 $X \sim Binomial(n, p)$

$$Pr(X = x) = {n \choose x} p^{x} (1-p)^{n-x} \text{ for } x = 0,1,...,n$$

Example: Germination of Seeds

- Germination rate = 90% \leftarrow claimed by the store manager
- If you bought n = 20 seeds, what is the expected number of seeds to germinate?
- If you planted the 20 seeds and there were 15 seeds which eventually germinate, can you say that the store manager lied to you?
 - A naïve method:

dbinom(0:20, n=20, p=0.9)



> n=20; p=0.9; k=seq(0,n)
> plot(k, dbinom(k,n,p), type='h', main='dbinom(0:20, n=20, p=0.9)', x
lab='k')

Let X = # of seeds that germinate ~ Binomial(n=20,p)

Based on the manager's argument: p = 0.9

$$E(X) = 20*0.9 = 18$$

$$Pr(X \le 15) = \sum_{x=0}^{x=15} \frac{20!}{x!(20-x)!} 0.9^{x} (0.1)^{20-x} \implies \text{sum of 16 terms}$$

$$= 1 - \sum_{x=16}^{x=20} \frac{20!}{x!(20-x)!} 0.9^{x} (0.1)^{20-x} \implies \text{sum of 5 terms}$$

$$= 0.0432$$

統計意義 (之後會學)

- $Pr(X \le 15) = 0.0432$ is related to "p-value" which provides evidence to check whether p = 0.9 is correct.
- Since 0.0432 is very low (compared with 0.05 which is the level of significance), p = 0.9 is not reasonable.

Remarks

- 1. You can use the attached table to find the probability for selected values of (n, p).
- 2. For any (n, p), use the online calculator

https://stattrek.com/online-calculator/binomial.aspx

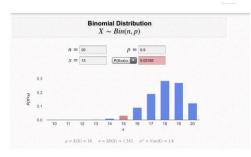
Example: When circuit boards used in the manufacture of compact disc players are tested, the long-run percentage of defectives is p = 5%.

Let X = the number of defective boards in a random sample of size n = 25.

1. Determine $Pr(X \le 2)$, $Pr(X \ge 4)$ and $Pr(1 \le X \le 4)$

$$\Pr(X \le 2) = \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) = \sum_{x=0}^{2} {25 \choose x} \cdot (0.05)^{x} \cdot (0.95)^{20-x}$$





Find the values of E(X), Var(X) and $\sqrt{Var(X)}$

$$E(X) = np = 25 \times 0.05 = 1.25$$

$$Var(X) = npq = 25 \times 0.05 \times 0.95 = 1.1875$$

$$\sqrt{Var(X)} = \sqrt{1.1875} = 1.09$$

Example:

Suppose that only 20% of all drivers come to a complete stop at an intersection having flashing red lights in all directions when no other cars are visible.

What is the probability that, of 20 randomly chosen drivers coming to an intersection under these conditions:

- 1. At most five will come to a complete stop?
- 2. Exactly five will come to a complete stop?
- 3. At least five will come to a complete stop?
- 4. How many of the next 20 drivers do you expect to come to a complete stop?

Sol:
$$n = 20, p = 0.2$$

X = number of cars come to a complete stop among 20 chosen cars

$$X \sim Binomial(n = 20, p = 0.2)$$

- 1. $Pr(at most 5) = Pr(X \le 5)$
- 2. Pr(exactly 5) = Pr(X = 5)
- 3. $Pr(\text{at least 5}) = Pr(X \ge 5)$
- 4. $E(X) = np = 20 \times 0.2 = 4$
- 5. $Var(X) = np(1-p) = 20 \times 0.2 \times 0.8 = 3.2$

II. Poisson random variable (波耳松分配)

Consider a stochastic process which records the occurrence of random events

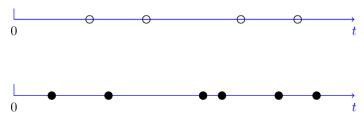
Examples of random events:

car accidents, earthquakes, receiving a phone call, disease occurrence

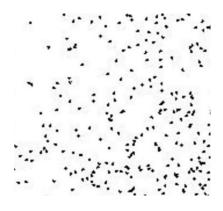
Poisson process: satisfies the following conditions

- The probability that *exactly one event* occurs in a given interval of length h is proportional to the length, say λh , where λ = the average number of events in a unit of time (space).
- When h is small, the probability that two or more events occur = 0.
- The numbers of events in two *disjoint* intervals are independent.

Plots of Poisson process (time)



Plots of Poisson process (space)



在零膨脹資料結構下進行卜瓦松迴歸模型與遞延分配模型之模… ndltd.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi/login?e=dnclcdr&s=id...

... 藉由1998-2008年高雄地區登革熱病例案件數及每日溫度資料進行模擬,在對立假設 下將登革熱病例資料與溫度的關係由零膨脹卜瓦松分布(Zero-inflated Poisson ...

南台灣登革熱族群傳輸動態及感染風險評估 - 臺灣博碩士論文... ndltd.ncl.edu.tw/handle/44847339654105523541

2012年6月21日 - 本研究亦發展以統計指標為基礎之卜瓦松迴歸模式(Poisson regression model)評估影響<mark>登革熟</mark>每月發生率之潛在因子,最後並以曲棍球桿..

國立臺灣大學生工系生物模擬與控制研究室: 研究主題-環境毒... homepage.ntu.edu.tw/~cmllao/research_7.htm * 本研究以台灣2001 - 2008年登革熱盛行地區之流行病學調查資料,利用波以松迴歸分析(Poisson regression analysi)探討登革熟發生率與氣候因子溫度、相對濕度...

[PDF] 2051KB - 行政院環境保護署

20JTND - 1 D以内が成場に設合 www.epa.gov.tw/cpDownloadCtt.asp?id=17104 ▼ 1994年12月31日 - 的一致性則沒有無象因子對於登革熱流行來的明顯與明確,後續進一步大量...... 進一步利用Poisson regression with Generalized Additive Model ...



Q: How do we judge whether the Poisson process is appropriate?

- Example 1: the number of car accidents in Kuan-Fu Road within a month
 - Yes \rightarrow if probabilities in two disjoint time intervals are independent
 - No → if the situation changes (more policemen, change of regulations)

 after a big accident within the same time interval
- Example 2: the number of phone calls received by an operator within 4 working hours
 - Yes \rightarrow if the patterns of phone calls are roughly the same

And also the chance of getting two calls at the same time is rare

Example3: the number of flaws in a book of 30 pages

Usually yes, unless some pages require special printing techniques.

Example 4: the number of customers coming to a store within 4 working hours

Usually no: since more than one person may come to the shop together.

The patterns may not be the same within the 4-hour interval.

Remarks:

- We may restrict the time or space which provides a *homogeneous* condition.
- More examples:
 - the location of users in a wireless network;
 - the requests for individual documents on a web server;
 - photons landing on a photodiode;
 - The number of deaths attributed to typhoid fever (傷寒).
- The Poisson process is connected in various interesting ways to a number of special distributions, including the *Poisson, exponential, Gamma, Beta, uniform, binomial, and the multinomial.* These embracing connections and wide applications make the Poisson process a very special topic in probability.

Poisson Random Variables

From a Poisson process, pick an interval or space.

Define the number of events occurring in the interval or space.

Definition 1: Poisson random variable

Let X be the number of events in the time interval [0,t] with length t.

Let λ = the average number of events in a unit of time.

Then $E(X) = \lambda t = \mu$.

Definition 1: Poisson random variable

 $X \sim \text{Poisson}(\mu = \lambda t)$ with the probability function:

$$\Pr(X = x) = \frac{\exp(-\lambda t)(\lambda t)^x}{x!} = \frac{\exp(-\mu)(\mu)^x}{x!} \quad (x = 0, 1, 2, \dots) \Rightarrow \text{ no upper bound}$$

 $E(X) = \lambda t$ & $Var(X) = \lambda t$ (Poisson: the mean and variance are the same)

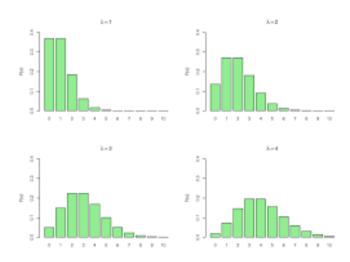
Definition 2: Poisson random variable

$$\Pr(X = x) = \frac{\exp(-\mu)\mu^{x}}{x!}, \quad x = 0,1,2,...$$

$$E(X) = \sum_{x=0}^{\infty} x \frac{\exp(-\mu)\mu^{x}}{x!} = \mu. \text{ (the definition of } \mu \text{ is the mean)}$$

$$Var(X) = \sum_{x=0}^{\infty} (x - \mu)^{2} \frac{\exp(-\mu)\mu^{x}}{x!} = \mu.$$

Poisson: Probability functions: x axis: x, y axis: Pr(X = x)

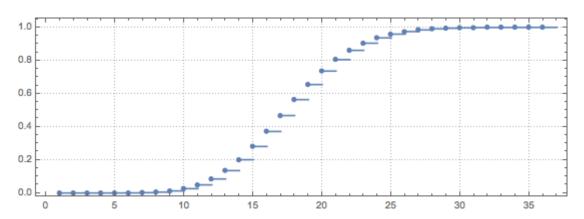


Cumulative distribution function:

$$\Pr(X \le a) = \sum_{x=0}^{x=a} \frac{\exp(-\lambda t)(\lambda t)^x}{x!} = \sum_{x=0}^{x=a} \frac{\exp(-\mu)(\mu)^x}{x!} \rightarrow \text{sum of finite series}$$

→ no explicit formula

Poisson: cumulative distribution function: $Pr(X \le x) \leftarrow right$ continuous



Example: Births in a hospital occur randomly at an average rate of 1.8 births per hour.

Let $X = \text{No. of births in a given hour} \sim \text{Poisson}(1.8)$

1. What is the probability of observing 4 births in a given hour at the hospital?

$$\Pr(X=4) = \frac{e^{-1.8}(1.8)^4}{4!} = \frac{0.1653 \times 10.5}{24} = 0.072$$

2. What about the probability of observing more than or equal to 2 births in a given

hour at the hospital?

$$Pr(X \ge 2) = 1 - Pr(X = 0) - Pr(X = 1) \approx 0.5347$$

$$\Pr(X=0) = \frac{e^{-1.8}(1.8)^0}{0!} = 0.1653$$

$$\Pr(X = 1) = \frac{e^{-1.8}(1.8)^1}{1!} \approx 0.30$$

3. What about the probability of **observing no birth** within two hours at the hospital?

Hint: Let Y = No. of births in a two-hour period

$$E(Y) = 1.8 \times 2 = 3.6$$

$$\Pr(Y=0) = \frac{e^{-3.6}(3.6)^0}{0!} = 0.027$$

Comparison between Binomial and Poisson distributions

	Binomial	Poisson
fixed	total # of experiment (n)	length or space
definition	count the number of	count the number of events
	successes	
starting value	0	0
upper bound	n	no upper bound

Derivations of Poisson probability: from Binomial (補充, 不考)

Cut [0,t] into n small non-overlapping intervals, each with length $\Delta t = \frac{t}{n}$



Imagine that in each interval Δt , perform a Bernoulli random experiment, each with success probability $p = \frac{E(X)}{n} = \frac{\mu}{n}$.

$$X = \text{ # of successes in the } n \text{ trials } \Pr(X = x) = \binom{n}{x} (\frac{\mu}{n})^x (1 - \frac{\mu}{n})^{n-x}$$

$$= \frac{n(n-1)...(n-x+1)(n-x)...(1)}{[(x)(x-1)...1][(n-x)(n-x-1)...1]} \cdot (\frac{\mu}{n})^x \cdot (1 - \frac{\mu}{n})^{n-x}$$

$$= \frac{1}{x!} \cdot \left(\frac{n}{n} \frac{n-1}{n}...\frac{n-x+1}{n}\right) \cdot \mu^x \cdot (1 - \frac{\mu}{n})^{n-x} = (1) \cdot (2) \cdot (3) \cdot (4)$$

Note: when $n \rightarrow \infty$

(1):
$$\frac{1}{x!}$$

(2):
$$\frac{n}{n} \frac{n-1}{n} \dots \frac{n-x+1}{n} \xrightarrow[n \to \infty]{} 1$$

(3):
$$\mu^{x}$$

(4):
$$(1-\frac{\mu}{n})^{n-x} = (1-\frac{\mu}{n})^n (1-\frac{\mu}{n})^{-x} \xrightarrow[n \to \infty]{} e^{-\mu}$$

(4.1):
$$(1 - \frac{\mu}{n})^n \xrightarrow[n \to \infty]{} e^{-\mu}$$
 (4.2): $(1 - \frac{\mu}{n})^{-x} \xrightarrow[n \to \infty]{} 1$

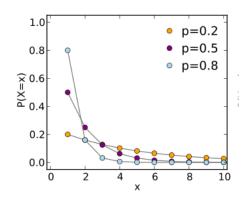
Probability of Poisson random variable (taking $n \to \infty$) $Pr(X = x) = \frac{1}{x!}e^{-\mu}\mu^x$

Other Discrete Random Variables (will not be on the exam)

* Geometric distribution (幾何分配)

X = # of Bernoulli trials in order to obtain the first success

$$Pr(X = x) = (1 - p)^{x-1} p \quad (x = 1, 2, 3, ...)$$



Example:

- A couple really wants to have a girl. They will keep having babies until they get a girl (and then stop). X = # of children they have

* Multinomial distribution (多項式分配)

In each trial, there are K possible outcomes.

(When
$$K = 2 \rightarrow Binomial$$
)

 X_k = number of occurrences for outcome k (k = 1,...,K);

 p_k = the probability of occurrence for outcome k;

Note:
$$\sum_{k=1}^{K} X_k = n$$
, $\sum_{k=1}^{K} p_k = 1$

$$Pr(X_1 = x_1, ..., X_K = x_K) = \frac{n!}{x_1! ... x_K!} (p_1)^{x_1} ... (p_K)^{x_K}$$

Example: 100 NCTU students are selected to answer the question:

"Do you agree that NCTU and Yang Ming University should merge?"

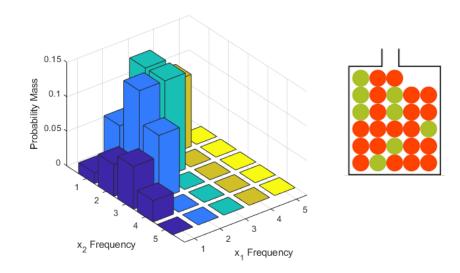
There are 3 choices: "agree", "neutral" & "disagree"

$$n = 100$$
, $K = 3$, $p_1 = Pr(agree)$; $p_2 = Pr(neutral)$; $p_3 = Pr(disagree)$;

For
$$K = 3 \Rightarrow \text{plot } (x_1, x_2, \Pr(X_1 = x_1, X_2 = x_2))$$

since $x_3 = n - x_1 - x_2$, $\Pr(X_1 = x_1, X_2 = x_2, X_3 = n - x_1 - x_2)$

補充三:*



*Hypergeometric distribution (超幾何分配)

Total: A + B balls (Red: A balls; Green: B balls)

draw n balls from the pool which contains A + B balls

Let X =the number of red balls (suppose n < A and n < B)

$$\Pr(X = x) = \frac{\binom{A}{x} \binom{B}{n-x}}{\binom{A+B}{n}} \quad x = 0, ..., n$$

Remarks:

- This is an example of sampling from a <u>finite population</u>.
- The probability of obtaining a red ball <u>keeps changing</u> and <u>depends on</u> $\frac{\text{the previous outcome}}{\text{the previous outcome}} \rightarrow \text{violate the assumptions of constant } p \text{ and}$ independence for Binomial distributions.
- When A+B>>> n, Hypergeometric distribution is close to Binomial distribution. \rightarrow since different values of p are still very close and previous results have very little effect in which independence is still plausible.

離散隨機變數概念小測驗



假設清華大學同學喜歡 Alisasa 的機率 = p

- 1. 隨機抽 n 個人, 喜歡 Alisasa 的人數 (只有<u>喜歡與不喜歡</u>的選項): 二項式分配
- 2. 隨機找路人來問,直到問到一位喜歡 Alisasa,總共會問多少人: 幾何分配
- 3. 桶子裡裝著許多同學投入(兩個顏色)的紙片, 紅色A張代表喜歡 Alisasa,黑色B張代表不喜歡,均勻混合。 從中抽出n 張紙片,裡面紅色的張數: 超幾何分配
- 4. 隨機抽 n 個人, 問對Alisasa的看法 (喜歡、不喜歡、沒意見三種選項, 機率 p1,p2,p3) 三種意見的個數: 多項式分配

Poisson 分佈

5. 如果「喜歡」Alisasa是「罕見事件」,可以在某天上學日的早上九點到十一點, 問進入校園的同學並計算喜歡的人數

補充不考

幾何分配

A representative from the National Football League's Marketing Division randomly selects people on a random street in Kansas City, Missouri until he finds a person who attended the last home football game. Let p, the probability that he succeeds in finding such a



person, equal 0.20. And, let X denote the number of people he selects until he finds his first success. What is the probability mass function of X?

超幾何分配

In a small pond there are 50 fish, 10 of which have been tagged. A fisherman's catch consists of 7 fish (assume his catch is a random selection done without replacement). What is the probability that exactly 2 tagged fish are caught?

多項式分佈

Suppose that in a three-way election for a large country, candidate A received 20% of the votes, candidate B received 30% of the votes, and candidate C received 50% of the votes. If six voters are selected randomly, what is the probability that there will be exactly one supporter for candidate A, two supporters for candidate B and three supporters for candidate C in the sample?

Approximation: Use Poisson to approximate Binomial (very important)

Given $X \sim Binomial(n, p)$,

$$Y \sim Poisson(\mu = \lambda t)$$

There exists a situation that

when
$$E(X) = np = E(Y) = \mu = \lambda t$$
,

$$Pr(X = t) \approx Pr(Y = t)$$
 for most $t = 0, 1, 2, ...$

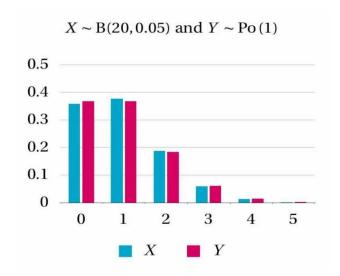
Question: When does the above situation happen?

Answer: for large n and small p (for rare events)

$$\Pr(X = t) = \binom{n}{t} p^t (1 - p)^{n - t} \approx \Pr(Y = t) = \frac{e^{-np} (np)^t}{t!}$$

Specific criteria for the approximation:

- $n \ge 20$, $p \le 0.05$ \rightarrow good approximation
- $n \ge 100, np \le 10$ \rightarrow very good approximation



Remarks:

- Computing a Binomial probability is more complicated since it involves

$$\frac{n!}{x!(n-x)!}$$
 and p^x (or $(1-p)^{n-x}$) if p is very small or large

- We usually use the Poisson distribution to approximate the corresponding Binomial distribution.

Example: bacteria mutate to become resistant to antibiotic.

$$n = 2 \times 10^9$$
 cells

 $p = 10^{-9}$ (the probability of mutation of each cell)

X = # of mutation ~ Binomial($n = 2 \times 10^9$, $p = 10^{-9}$)

$$Pr(X = x) = {2 \times 10^9 \choose x} (10^{-9})^x (1 - 10^{-9})^{2 \times 10^{-9} - x} \implies \text{based on Binomial}$$

→ computationally impossible

Use Poisson for approximation

$$E(X) = np = 2 \rightarrow \text{approximated by } X \sim \text{Poisson}(2)$$

$$\Pr(X=x) \approx \frac{e^{-2}(2)^x}{x!}.$$

$$Pr(X = 0) \approx e^{-2} = 0.1353 = Pr(\text{no mutation})$$

$$Pr(X \ge 1) = 1 - Pr(X = 0) = 0.7247 = Pr(at least one mutation)$$

Example: Airline Overbook

- An airline knows that overall 3% of passengers do not turn up for flights .
- The airline decides to adopt a policy of selling more tickets than there are seats on a flight. → overbooked (超賣)
- For an aircraft with 196 seats, the airline sold 200 tickets for a particular flight. By using a suitable approximation, find
 - the probability that there is at least one empty seat on this flight
 - the probability that at least one passenger will have to be bumped from the flight?"

Let X = number of passengers who buy the tickets but will not show up

= 有票但不會搭機的旅客人數

 $X \sim Binomial(200, 0.03) \rightarrow large \ n \ and small \ p$

 $X \sim Binomial(200, 0.03) \approx Y \sim Poisson(6)$

Pr(there is at least one empty seat on this flight)

- = Pr(at most 195 passengers show up)
- = Pr(at least 5 persons who buy the tickets but will not show up)

$$= \Pr(X \ge 5) = 1 - \Pr(X \le 4)$$

$$=1-\sum_{x=0}^{x=4} {200 \choose x} 0.03^x 0.97^{200-x}$$

= $0.7190 \rightarrow$ sometimes very difficult to compute

$$\approx \Pr(Y \ge 5)$$

$$=1-\sum_{x=0}^{x=4}\frac{e^{-6}6^x}{x!}=1-(0.002479)\sum_{x=0}^{x=4}\frac{6^x}{x!}$$

 $= 1 - 0.2851 = 0.71494 \rightarrow$ easier to obtain

Pr(at least one passenger will have to be bumped)

- = Pr(at least 197 passengers show up)
- = Pr(at most 3 persons who buy the tickets but will not show up)

$$= \Pr(X \leq 3)$$

$$\Pr(X \le 3) = \sum_{x=0}^{x=3} {200 \choose x} (0.03)^x (0.97)^{200-x} = .14715 \leftarrow \text{ more difficult}$$

$$\approx \sum_{x=0}^{x=3} \frac{e^{-6} 6^x}{x!} = 0.1512 \leftarrow \text{ easier}$$

www.thenewslens.com > 國際 ▼ Translate this page

美國聯合航空超賣機位,亞裔男子竟遭航警強拖下機還被撞得...

Apr 11, 2017 — 因機位起賣、無法起飛·於是航空公司以800美元(約24000元台幣)希望徵求自願下機乘客·仍無人願意·只好用電腦隨機抽出四人。 標籤: 聯合 ...

www.cw.com.tw › ... › 經濟學人 ▼ Translate this page

聯航讓亞裔醫師被爆打一年後美國的航空變了 | 天下雜誌

Apr 14, 2018 — 如果某架班機已<mark>超賣</mark>或可能<mark>超賣</mark>。有加人的人就會在飛機起飛5天前就收到通知。並選擇自己要... 延伸閱讀:消費者是健忘的<mark>聯合航空好輝</mark>煌).

www.storm.mg > article 🔻 Translate this page

聯航為何低頭道歉?業界律師:這不是超賣機位趕人下機要負民...

Apr 12, 2017 — 美國聯合航空公司日前一架準備從芝加哥飛往路易斯維爾的航班上,一名乘客因為拒絕改養被安全人員強行拖下飛機,這個事件被同機乘客拍攝...

補充: 有關臨床試驗

為何國產疫苗只需要二期試驗過關,就可緊急授權?【陳建仁專欄】

疫苗是現代解決傳染病的最好辦法,醫學上,如何透過三期試驗,證明安全有效?進度落後的國產疫苗,有機會緊急授權?

疫苗在 20 世紀成為撲滅傳染病的最佳策略,疫苗的研發要先經過動物實驗證實安全性和有效性,再進行第一、二、三期人體臨床試驗,甚至進一步進行上市後的第四期試驗。第一期臨床試驗通常會選擇健康且免疫功能健全的成人當受試者,主要是評估人體施打疫苗的安全性及耐受性。

第二期是探討疫苗的免疫激發力與安全性,找出最適合施打的劑量、劑型與時程。

第三期要確認疫苗的有效性與安全性,一組打疫苗、一組打安慰劑,隨機分配,醫生與 病人都不知道誰打什麼(雙盲),最後解盲才能正確得知保護力。

疫苗的保護力是指,接種疫苗的人群與沒有接種的人群相較,可減少多少罹病、重症或死亡風險,可經由第三期試驗,或大規模接種後的第四期試驗得知。

默沙東COVID-19口服藥是疫情扭轉者?效果、 副作用一次看懂

2021/10/7 14:49 (10/7 15:10 更新)

● Molnupiravir是何方神聖

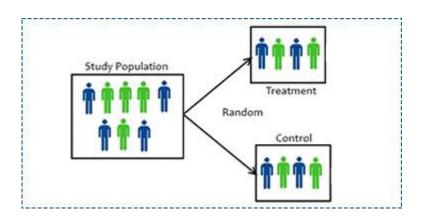
Molnupiravir是一款原本研發用於治療流感口服膠囊的化學藥名。它能藉由誘發「致命突變」抑制新型冠狀病毒複製。簡單來說,藥理機制會讓病毒複製遺傳物質時出錯,使病毒複製出現缺陷。

■ Molnupiravir效果如何

默沙東10月1日的聲明指出,隨機試驗數據的期中分析顯示,Molnupiravir可將確診住院風險降低約50%。試驗期間,385名服藥的受試患者僅28人住院,比例7.3%;377名服安慰劑的受試患者卻有53人住院,比例達14.1%。

Development of Clinical Trials (臨床試驗)

- ▶ RA Fisher: Experimental design (1922-1926)
 - Randomization
 - Replication
 - o Blocking
- First randomized curative trial: 1946–1947
 - test the efficacy of the <u>chemical streptomycin</u> for curing pulmonary tuberculosis (肺結核)
 - Double-blind & placebo controlled
 - Compute Relative risk $\frac{\Pr(D \mid E)}{\Pr(D \mid E^c)}$



Published: 06 February 2007

R. A. Fisher and his advocacy of randomization

Nancy S. Hall

✓

Journal of the History of Biology 40, 295–325 (2007) | Cite this article 529 Accesses | 31 Citations | 12 Altmetric | Metrics

Abstract

The requirement of randomization in experimental design was first stated by R. A. Fisher, statistician and geneticist, in 1925 in his book *Statistical Methods for Research Workers*. Earlier designs were systematic and involved the judgment of the experimenter; this led to possible bias and inaccurate interpretation of the data. Fisher's dictum was that randomization eliminates bias and permits a valid test of significance. Randomization in

Ronald Fisher

British mathematical statistician





His contributions to statistics include promoting **the method of maximum likelihood** and deriving the properties of maximum likelihood estimators, fiducial inference, the derivation of various sampling distributions, founding principles of the design of experiments, and much more.

Mathematics Genealogy Project 數學族譜

https://www.genealogy.math.ndsu.nodak.edu/id.php?id=73510

