Lecture 6 From Discrete to Continuous Random Variables

10/21/2020

- **Exercise (Poisson)**

The mean number of bacteria per milliliter of a liquid is known to be 4.
Assuming that the number of bacteria follows a Poisson distribution,
Find the probability that, in 1ml of liquid, there will be
(a) no bacteria
(b) 4 bacteria
(c) fewer than 3 bacteria
Sol:
X ~ Poisson(4)

(a) $\Pr(X=0)=\dfrac{e^{-4}(4)^0}{0!}=e^{-4}=0.0183$

(b) $\Pr(X=4)=\dfrac{e^{-4}(4)^4}{4!}=\dfrac{0.0183\times 256}{24}=0.1954$

(c)

$$\Pr(X<3)=\Pr(X=0)+\Pr(X=1)+\Pr(X=2)$$
$$=\frac{e^{-4}(4)^0}{0!}+\frac{e^{-4}(4)^1}{1!}+\frac{e^{-4}(4)^2}{2!}=0.2381$$

Poisson calculator:
https://stattrek.com/online-calculator/poisson.aspx

Find the probability that
(d) in 3ml of liquid there will be less than 2 bacteria
X ~ Poisson(4*3=12)

$$\Pr(X<2)=\Pr(X=0)+\Pr(X=1)=\frac{e^{-12}(12)^0}{0!}+\frac{e^{-12}(12)^1}{1!}=0.00008$$

(e) in 0.5ml of liquid there will be more than 2 bacteria
X ~ Poisson(4*0.5=2)

$$\Pr(X>2)=1-\{\Pr(X=0)+\Pr(X=1)+\Pr(X=1)\}$$
$$=1-\{\frac{e^{-2}(2)^0}{0!}+\frac{e^{-2}(2)^1}{1!}+\frac{e^{-2}(2)^2}{2!}\}=0.3233$$

- **Approximation: Binomial & Poisson**

Given $X \sim Binomial(n, p)$, $Y \sim Poisson(\mu = np)$

For large n and small p:

$$\Pr(X = t) = \binom{n}{t} p^t (1-p)^{n-t} \approx \Pr(Y = t) = \frac{e^{-np}(np)^t}{t!} \quad \text{for } t = 0,1,\ldots,n$$

Example

In a large town, one person in 80, on average, has blood type X.

(a). If 200 blood donors are taken at random, what is the probability that they include at least five persons having blood type X. Just write down the formula

$$X \sim Binomial(200, \frac{1}{80})$$

$$\Pr(X \ge 5) = \sum_{x=5}^{x=200} \binom{200}{x} (\frac{1}{80})^x (\frac{79}{80})^{200-x} = 0.1076 \text{ (from 網路計算機)}$$

(b) Find an approximation to the probability that they include at least five persons having blood type X.

$$Y \sim Poisson(\frac{200}{80} = 2.5)$$

$$\Pr(Y \ge 5) = 1 - \sum_{x=0}^{x=4} \frac{e^{-2.5}(2.5)^x}{x!} = 0.1088$$

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

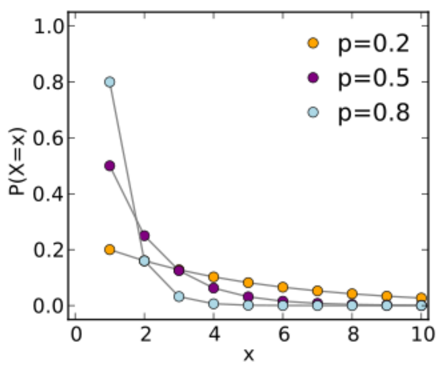
*** Geometric distribution (幾何分配)**

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

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

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



2. Multinomial distribution (多項式分配)

In each trial, there are K possible outcomes.

(When K = 2 → Binomial)

X_k = number of occurrences for outcome k $(k = 1, \ldots, K)$;

p_k = the probability of occurrence for outcome k;

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

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

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

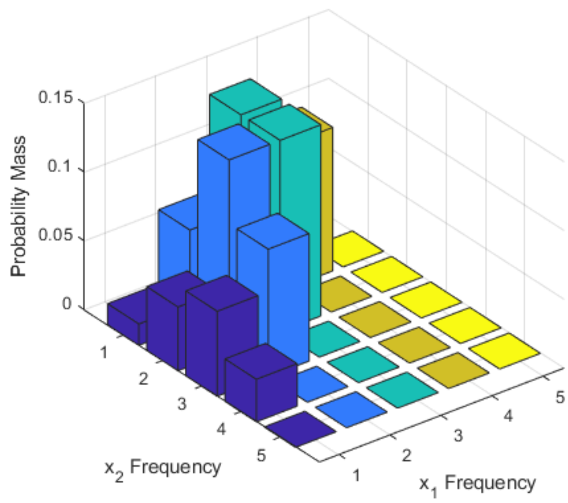
"Do you agree that NCTU and Yang Ming University should merge?"

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

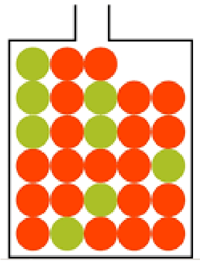
$n = 100$, $K = 3$, p_1 = Pr(agree); p_2 = Pr(neutral); p_3 = Pr(disagree);

For $K = 3$ → 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)

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

Remarks:

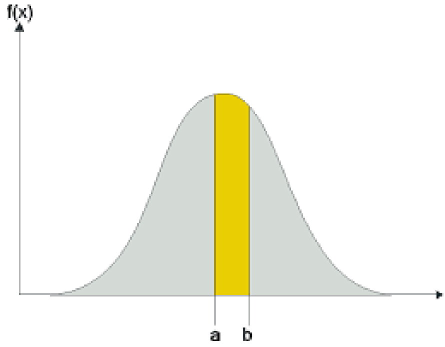
- This is an example of sampling from a <u>finite population</u>.
- The probability of obtaining a red ball <u>keeps changing</u> and <u>depends on the previous outcome</u> → violate the assumptions of constant p and independence for Binomial distributions.
- When $A+B >>> n$, Hypergeometric distribution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

Topic: Continuous Random Variables

Density function: $f(x) = \lim_{\Delta \to 0} \dfrac{\Pr(X \in [x, x+\Delta))}{\Delta}$

Probability for a continuous random variable:

$$\Pr(a \le X \le b) = \int_a^b f(x)dx$$ → area under the density function



Note: $\int_{-\infty}^{\infty} f(x)dx = 1$

Cumulative density function (CDF)

$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(u)du$$

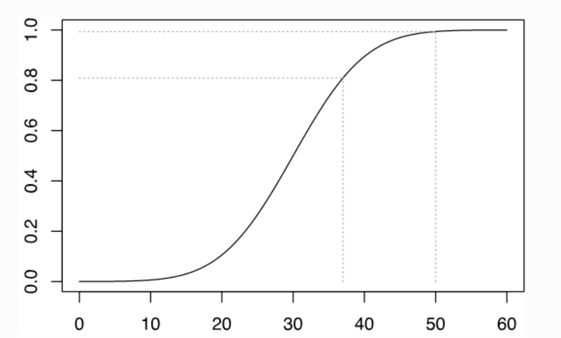
Remarks:

i. The relationship between CDF 和 the density function

$$\frac{\partial F(x)}{\partial x} = f(x)$$

ii. $F(x)$ is an increasing function satisfying $0 \le F(x) = \Pr(X \le x) \le 1$

$$Pt(a \le X \le b) = F(b) - F(a)$$



iii. $f(x) > 0$ but it is possible that $f(x) > 1$

iv. If X is continuous, then $\Pr(X = a) = 0$. (important)

$$\Pr(X = a) = \Pr(X \le a) - \Pr(X < a)$$

$$= \int_{-\infty}^{a} f(x)dx - \int_{-\infty}^{a-} f(x)dx = \int_{a-}^{a} f(x)dx = 0.$$

Definitions of Parameters

- Mean: $E(x)=\int_{-\infty}^{\infty} xf(x)dx=\mu$

- Variance

$$Var(x)=\sigma^2$$
$$=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx=\int_{-\infty}^{\infty}x^2 f(x)dx-\mu^2=E(X^2)-[E(X)]^2$$

- Quartiles

a. Q_1: the first quartile satisfying $\int_{-\infty}^{Q_1} f(x)dx=\frac{1}{4}$

b. M: the median satisfying $\int_{-\infty}^{M} f(x)dx=\frac{1}{2}$

c. Q_3: the third quartile satisfying $\int_{-\infty}^{Q_3} f(x)dx=\frac{3}{4}$

Remarks:

1. Discrete version: $\sum$ (summation)

   Continuous version: $\int$ (integration)

2. $F(x)$, $\mu=E(X)$ and $\sigma^2=Var(X)$ may not have closed forms so that numerical integrations are usually required.

3. For the Normal random variable with $\mu=0$ and $\sigma=1$, we write $F(z)=\Phi(z)$ and provide the table which contains the values of $\Phi(z)$ for many selected values of z.

Exercise: Use the formula of area to find the probability

$f(x)=\frac{200x}{9}$ $\quad 0\le x\le k$ (Note: $f(x)$ is a straight line passing through $(0,0)$)

a. prove $k=0.3$

$$\int_0^k f(x)dx=\int_0^k \frac{200}{9}xdx=1$$

Use area of a triangle: $\frac{1}{2}k\cdot\frac{200}{9}k=1\Rightarrow k^2=\frac{9}{100}\Rightarrow k=\frac{3}{10}$

Integration formula: $\int xdx=\frac{1}{2}x^2$, $\int x^k dx=\frac{1}{k+1}x^{k+1}$ $\quad (k=1,2,...)$

b. Find $\Pr(X \le 0.1)$

Use area: $\Pr(X \le 0.1) = \frac{1}{2} 0.1 \times (\frac{200}{9} \cdot \frac{1}{10}) = \frac{1}{9}$

Use integration: $\int_0^{0.1} \frac{200}{9} x dx = \frac{200}{9} \frac{1}{2} (\frac{1}{10})^2$

c. Find $\Pr(0.1 \le X \le 0.2)$ $= \Pr(X \le 0.2) - \Pr(X \le 0.1)$ → Trapezoid

when $x = 0.1$, $f(0.1) = \frac{200}{9} \cdot \frac{1}{10} = \frac{20}{9}$

when $x = 0.2$, $f(0.2) = \frac{200}{9} \cdot \frac{2}{10} = \frac{40}{9}$

Area = $(\frac{20}{9} + \frac{40}{9}) \cdot \frac{1}{10} \cdot \frac{1}{2} = \frac{3}{9} = \frac{1}{3} = \Pr(0.1 \le X \le 0.2)$

d. $E(X) = \int_0^{0.3} x f(x) dx = \int_0^{0.3} \frac{200}{9} x^2 dx = \frac{200}{9} \cdot \frac{1}{3} x^3 |_0^{0.3}$

$= \frac{200}{9} \cdot \frac{1}{3} \frac{27}{1000} = \frac{1}{5}$

e. $Var(X) = E(X^2) - \mu^2$

$E(X^2) = \int_0^{0.3} x^2 f(x) dx = \int_0^{0.3} \frac{200}{9} x^3 dx = \frac{200}{9} \cdot \frac{1}{4} x^4 |_0^{0.3}$

$= \frac{200}{9} \cdot \frac{1}{4} \frac{81}{10000} = \frac{9}{200}$

$Var(X) = E(X^2) - \mu^2 = \frac{9}{200} - \frac{1}{25} = \frac{1}{200}$

Example: X = the time (in minutes) that a pheromones trail persists after the last secretion of hormone. (“pheromones”: 一種螞蟻用來標示路徑的荷爾蒙)

Density function:

$$f(x) = \frac{1}{4} x \quad \text{for} \quad 0 \le x \le 2.$$

$$f(x) = -\frac{1}{4} x + 1 \quad \text{for} \quad 2 \le x \le 4.$$

求 $\Pr(1 < X < 3)$, $E(X)$, $Var(X)$

Solution:

$\Pr(1 < X < 3) = \int_1^2 \frac{1}{4} x dx + \int_2^3 (-\frac{1}{4} x + 1) dx$ (use integration or the area formula)

$$=\frac{1}{4}\frac{1}{2}x^2|_1^2+(-\frac{1}{4})\frac{1}{2}x^2|_2^3+x|_2^3 \quad =\frac{1}{8}(4-1)+(-\frac{1}{8})(9-4)+1=\frac{6}{8}$$

$$E(X)=\int_0^2 x\left(\frac{1}{4}x\right)dx+\int_2^4 x(-\frac{1}{4}x+1)dx=\mu$$

$$Var(X)=\int_0^2 x^2\left(\frac{1}{4}x\right)dx+\int_2^4 x^2(-\frac{1}{4}x+1)dx-\mu^2$$

Special Continuous Random Variables

A. Uniform Random Variable: $X \sim U(a,b)$

符號意義: U – uniform 分配, a: 下界; b: 上界

density 表示法: $f(x)=\frac{1}{b-a}$ if $a \le x \le b$

$f(x)=0$ otherwise.

Proof:

$$\int_{-\infty}^{\infty} f(x)=\int_{-\infty}^{a}0dx+\int_a^b\frac{1}{b-a}dx+\int_b^{\infty}0dx=\frac{1}{b-a}\int_a^b 1dx=\frac{1}{b-a}(b-a)=1.$$

Area of a rectangle: $(b-a)\cdot\frac{1}{b-1}=1$

Example: $X \sim U(5,10)$ → plot

a. 求 $\Pr(X \le 5)$ b. $\Pr(X \le 6)$ c. $\Pr(X > 8)$

補充: 交大校務資料分析 (化鏡為窗)

Normal distribution (常態分佈)

$X \sim N(\mu, \sigma^2)$, where $E(X) = \mu$, $Var(X) = \sigma^2$

density: $f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp(-\dfrac{(x-\mu)^2}{2\sigma^2}) \quad -\infty < x < \infty$

Remarks:

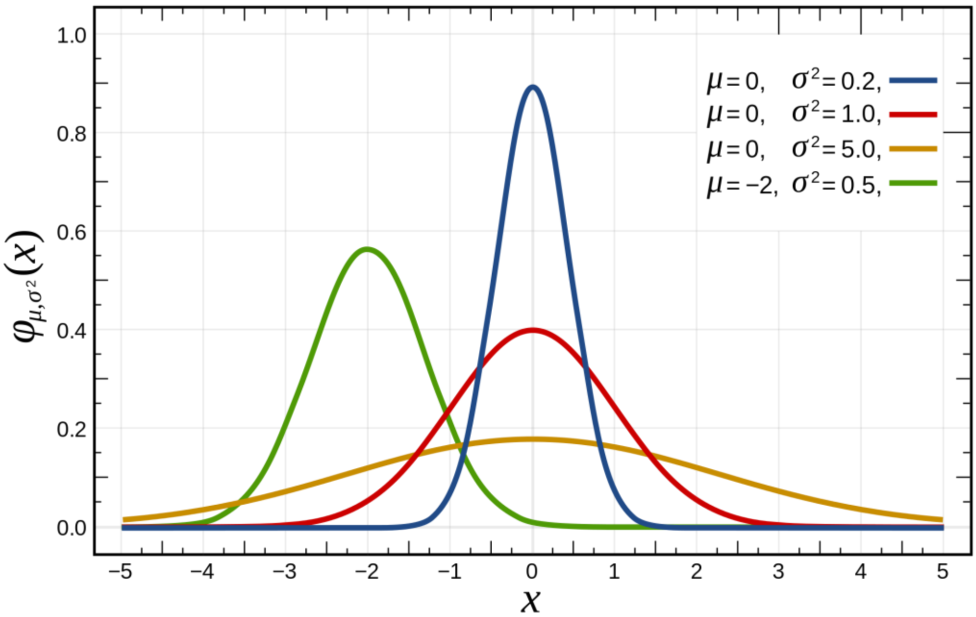
- The normal distribution is the most important distribution is statistics
- $f(x)$ is symmetric at $x = \mu$

Important parameters

- $\mu = E(X)$ satisfying

$$\mu = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$$

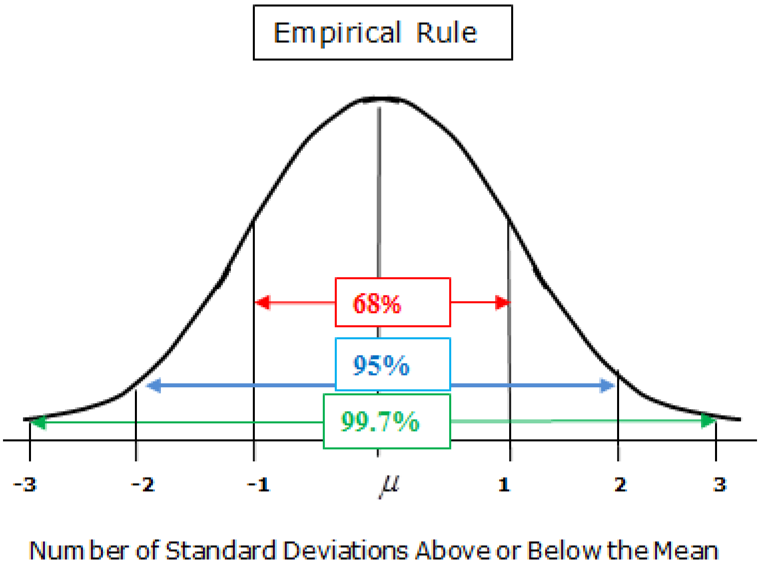
- $\sigma^2 = Var(X)$ satisfying
- $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} \dfrac{x^2}{\sqrt{2\pi\sigma^2}} \exp(-\dfrac{(x-\mu)^2}{2\sigma^2})dx - \mu^2$



Important Special case (標準常態) :

$Z \sim Normal(0,1)$, where $E(Z) = 0$, $Var(Z) = 1$

density: $f(z) = \dfrac{1}{\sqrt{2\pi}} \exp(-\dfrac{z^2}{2}) \quad -\infty < x < \infty$



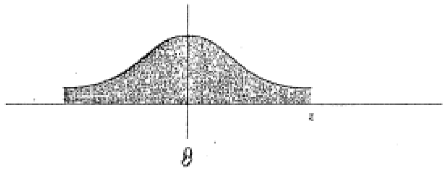
Use the Z table for N(0,1) to compute the probability

$$\Pr(-1 \le Z \le 1) = \Pr(Z \le 1) - \Pr(Z < -1) = 0.8413 - (1 - 0.8413)$$
$$= 0.6826$$

$$\Pr(-2 \le Z \le 2) = \Pr(Z \le 2) - \Pr(Z < -2) = 0.9772 - (1 - 0.9772)$$
$$= 0.9544$$

$$\Pr(-3 \le Z \le 3) = \Pr(Z \le 3) - \Pr(Z < -3) = 0.9987 - (1 - 0.9987)$$
$$= 0.9974$$

$F_Z(z) = P[Z \le z]$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1921	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148

Question: How do we find the probability for $X \sim N(\mu, \sigma^2)$ **using the Z table**

- **Standardization (標準化)**

Fact: $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

$$E(Z) = E(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma}\{E(X)-\mu\} = 0$$

$$Var(Z) = Var(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma^2}Var(X-\mu) = \frac{\sigma^2}{\sigma^2} = 1$$

To compute $\Pr(a \le X \le b)$, it follows that

$$\Pr(a \le X \le b) = \Pr(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}) = \Pr(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}) \rightarrow \text{Z table}$$

Remarks on notations: Pay attention to $X \sim N(\mu, \sigma^2)$ or $X \sim N(\mu, \sigma)$

Exercise: $X \sim N(\mu = 37, \sigma = 2)$ (sometimes $X \sim N(\mu = 37, \sigma^2 = 4)$)

$$\Pr(32 < X < 39) = \Pr(\frac{32-37}{2} < \frac{X-37}{2} < \frac{39-37}{2})$$

$$= \Pr(-2.5 < Z < 1) \quad = \Pr(Z < 1) - \Pr(Z \le -2.5) \quad = 0.8413 - 0.0062$$

Exercise: $X \sim N(\mu = 37, \sigma = 2)$ Find x such that $\Pr(36 < X < x) = 0.4$

$$\Pr(\frac{36-37}{2} < \frac{X-37}{2} < \frac{x-37}{2}) = 0.4$$

$$\Rightarrow \Pr(-0.5 < Z < \frac{x-37}{2}) = 0.4$$

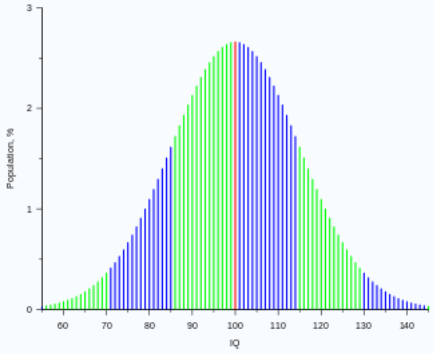
$$\Rightarrow \Pr(Z < \frac{x-37}{2}) - \Pr(Z \le -0.5) = 0.4$$

$$\Rightarrow \Pr(Z < \frac{x-37}{2}) - 0.3085 = 0.4$$

$$\Rightarrow \Pr(Z < \frac{x-37}{2}) = 0.7085$$

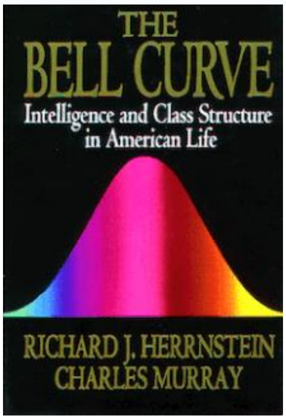
$$\Rightarrow \frac{x-37}{2} \approx 0.55 \rightarrow x \approx 38.1$$

Application of normal distribution – IQ test



- An intelligence quotient (IQ) is a total score derived from several standardized tests designed to assess human intelligence.
- When current IQ tests were developed, the median raw score of the norming sample is defined as IQ 100 and scores each standard deviation (SD) up or down are defined as 15 IQ points greater or less, although this was not always so historically.
- By this definition, approximately two-thirds of the population scores are between IQ 85 and IQ 115. About 2.5 percent of the population scores above 130, and 2.5 percent below 70.

Story about James Watson and the Bell Curve – See the last 2 pages



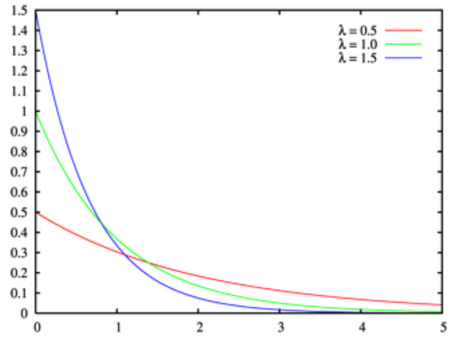
The Bell Curve: Intelligence and Class Structure in American Life is a 1994 book by psychologist Richard J. Herrnstein and political scientist Charles Murray, in which the authors argue that human intelligence is substantially influenced by both inherited and environmental factors and that it is a better predictor of many personal dynamics, including financial income, job performance, birth out of wedlock, and involvement in crime than are an individual's parental socioeconomic status. They also argue that those with high intelligence, the "cognitive elite", are becoming separated from those of average and below-average intelligence. The book was controversial, especially where the authors wrote about racial differences in intelligence and discussed the implications of those differences.

Exponential Distribution (will not be on the exam)

Example: Consider a Poisson process with λ being the average rate of occurrence within a unit time interval

Poisson distribution: $X \sim$ the number of events within $[0,t]$ interval

$$\Pr(X=x)=\frac{e^{-\lambda t}(\lambda t)^x}{x!} \quad (x=0,1,2,\ldots)$$



Exponential distribution:

$T \sim$ the waiting time to the first event

$$\Pr(T>t)=\Pr(X=0)=\frac{e^{-\lambda t}(\lambda t)^0}{0!}=e^{-\lambda t}$$

$$F(t)=\Pr(T\le t)=1-\Pr(T>t)=1-e^{-\lambda t}$$

The density of T :

$$f(t)=\frac{\partial F(t)}{\partial t}=\frac{\partial(1-e^{-\lambda t})}{\partial t}=(-1)(-\lambda)e^{-\lambda t}=\lambda e^{-\lambda t} \quad \text{for} \quad t>0$$

a. show $f(x)$ is a density staying $\int_0^\infty \lambda e^{-\lambda t}dt=1$

Formula: $\frac{\partial e^x}{\partial x}=e^x$

$$\int_0^\infty \lambda e^{-\lambda t}dt=\int_0^\infty e^{-\lambda t}d\lambda t=\int_0^\infty e^{-y}dy=-e^{-y}\mid_0^\infty=(-1)(e^{-\infty}-1)=0+1=1$$

b. Find $\Pr(T\le 2)$

Change variable: $y=\lambda t$

$$\Pr(T\le 2) = \int_0^2 \lambda e^{-\lambda t}dt=\int_0^{2\lambda} e^{-y}dy=-e^{-y}\mid_0^{2\lambda}=1-e^{-2\lambda}$$

c. Find $E(T)=\int_0^\infty t(\lambda e^{-\lambda t})dt=\frac{1}{\lambda}$

$$E(T)=\frac{1}{\lambda}\int_0^\infty \lambda t(e^{-\lambda t})d\lambda t=\frac{1}{\lambda}\int_0^\infty ye^{-y}dy$$

Use integration by parts $\int udv=uv-\int vdu$

$$\int_0^\infty ye^{-y}dy=-\int_0^\infty yde^{-y}=(-1)\cdot\{(ye^{-y})\mid_0^\infty-\int_0^\infty e^{-y}dy\}=1$$

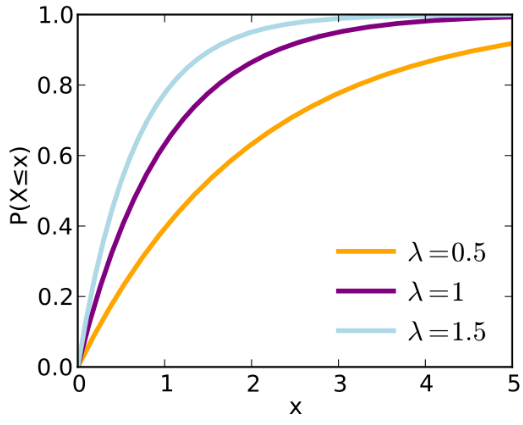
Random Number Generation (will not be on the exam)

Important fact:

$X \sim$ any continuous random variable with $F(x) = \Pr(X \le x)$

Then

$$F(X) \sim Uniform(0,1)$$



Example: $X \sim \text{Exponential}(\lambda)$

$F(x) = 1 - \exp(-\lambda x)$ →

Step 1: First generate U from $U(0,1)$

Step 2: Then solve

$U = F(X) = 1 - \exp(-\lambda X)$

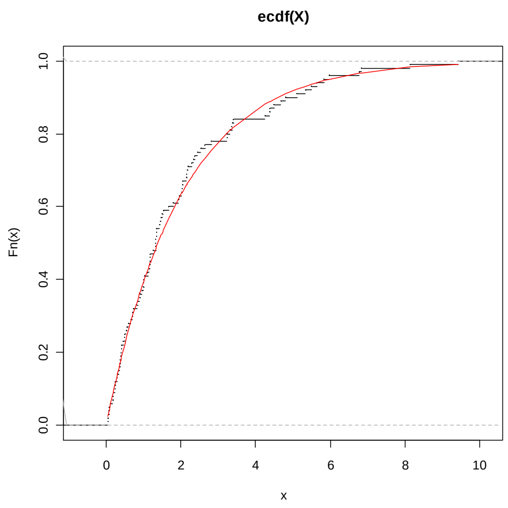
→ $\log(1-U) = -\lambda X$

→ $X = -\dfrac{\log(1-U)}{\lambda}$

Algorithm:

1. generate $U_i \sim \text{iid } U(0,1)$ for $i = 1,\ldots,100$
2. Set the value of λ
3. Let $X_i = -\dfrac{\log(1-U_i)}{\lambda}$
4. Based on $(X_1,\ldots,X_{100})$, plot $\bar{F}(x) = \sum_{i=1}^{100} I(X_i \le x)/100$
5. Plot $F(x) = 1 - \exp(-\lambda x)$.

Plot two curves $\bar{F}(x), F(X)$ $\lambda = 0.5, n = 100$ (red: $F(x) = 1 - \exp(-\lambda x)$)



NCTU 成績比較的分布圖

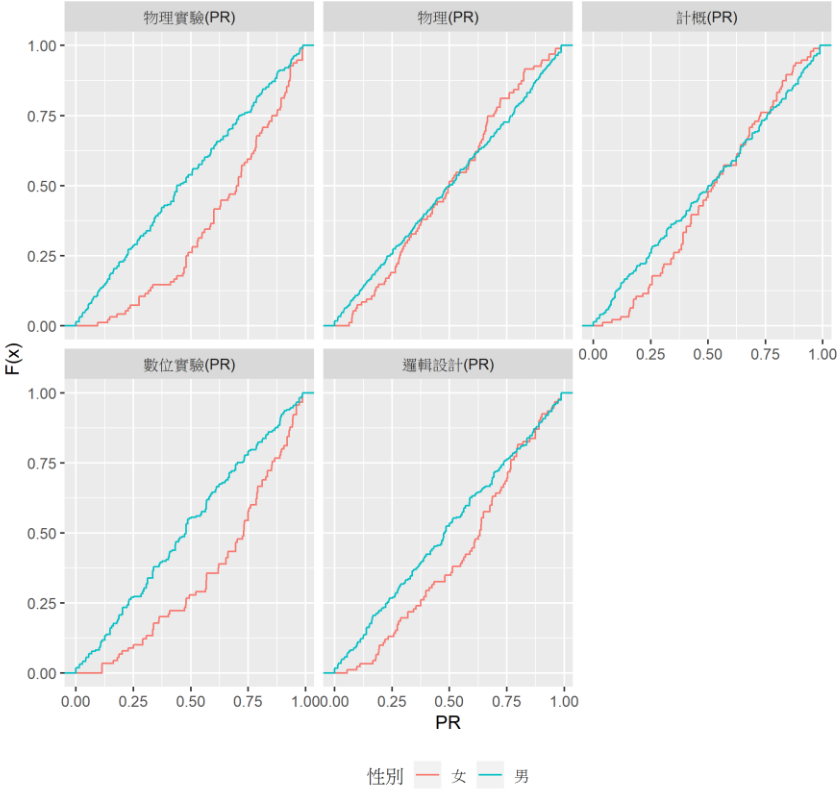
【決策支援】

第五章 大數據精準招生開展共贏局面　　鄭朝陽、林珊如
第六章 多元學習歷程與招生就學表現初探　陳思光、常善媚、林珊如
第七章 多元入學管道交大生學習成效盤點　　王維菁、謝維軒
第八章 由交大再探大學均質化及不同生源學業表現　毛靖嵐、洪慧念
第九章 正確解讀世界大學排名　　陳瑋真

如何控制 confounding variables:

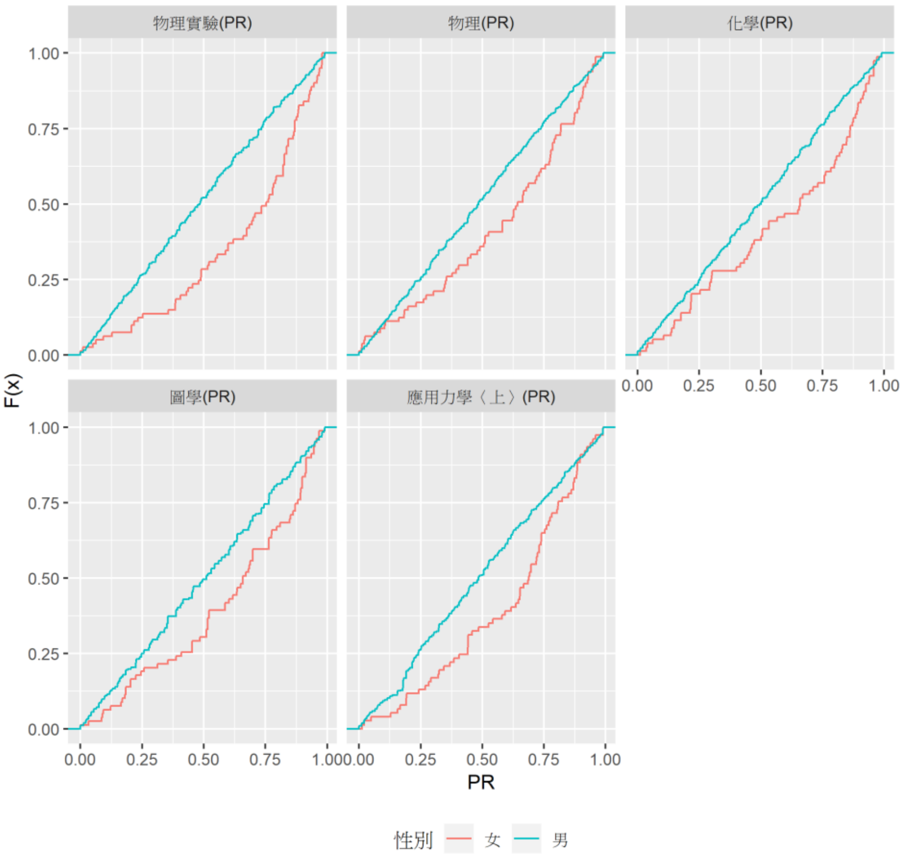
- 修課科目不盡相同 → 用系訂必修
- 不同科目老師給分的 "甜度" 不同 → 用 Percentile score
- 令 X 為百分位成績，其分佈為 $F(x)=\Pr(X\le x)=x$，$(x,F(x))$ 的圖形為 45 度的直線。可將百分位成績表示為 PRq，滿足 $F(\mathrm{PR}q)=q/100$ $(0\le q\le 100)$，q 值越高代表相對成績越好，例如 PR90 代表贏過百分之 90 的人。
- 亦可定義組內的分佈函數：$F_i(x)=\Pr(X\le x\,|\,\text{group i})$，圖形上的點若落在 45 度的右邊代表表現佳。例如 $F_i(\mathrm{PR}75)=0.5$，會對應圖形上的 $(0.75,0.5)$，代表在全班 PR75 的百分位成績在第 i 組只贏 50%的人，顯示這組競爭比較激烈，所以較優。
- 以不同顏色標示組別畫出 $(x,F_i(x))$ $(i=1,\ldots,I)$，可清楚的比較各組細部的差異。當縱軸的 q 值越靠近 100，比較的是各組最前段的學生；q 值越靠近 50，比較的是各組的中段生；q 值越靠近 0，對象是各組的末段生。若 $F_i(x)$ 與 $F_j(x)$ 的圖形有交錯，代表第 i 組與第 j 組在不同區段的 q 值互有勝負；若某一組和另一組完全無交錯，代表 45 度線右邊的組完勝(左邊的完敗)。

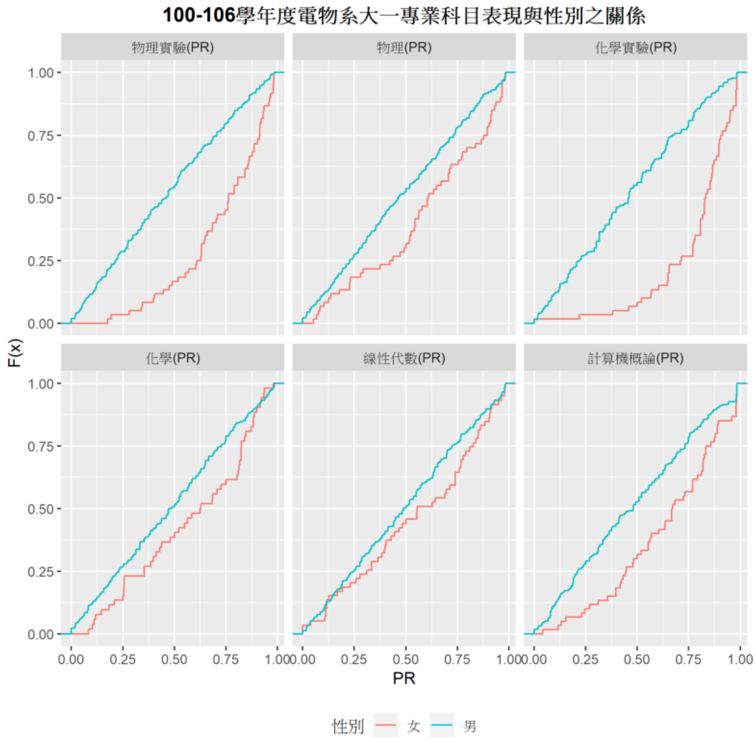
100-106學年度電工系大一專業科目表現與性別之關係

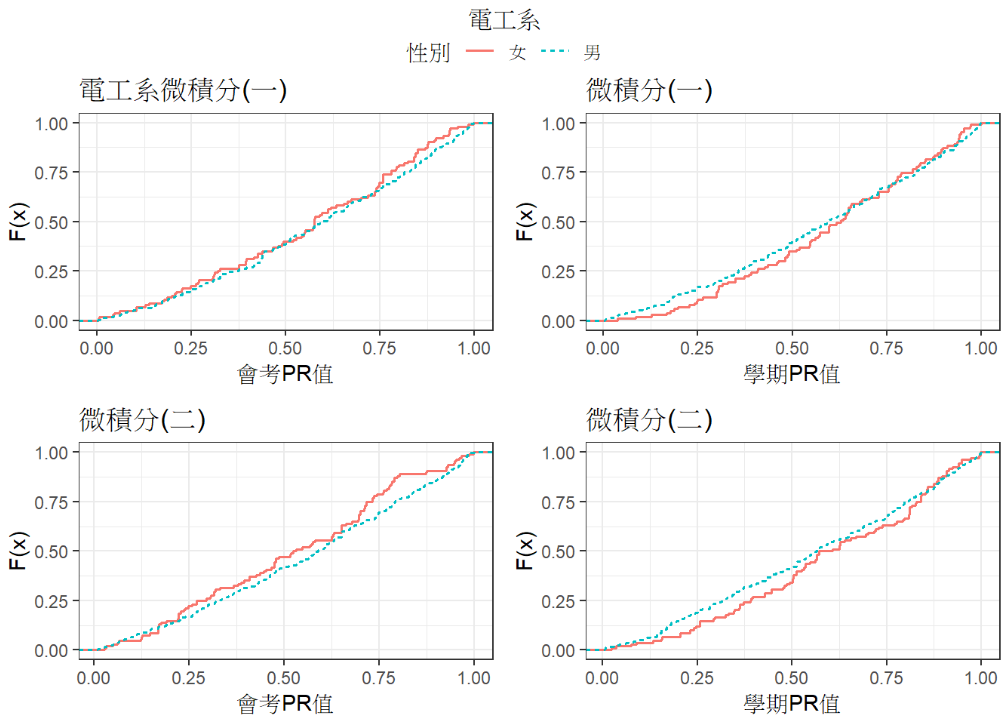


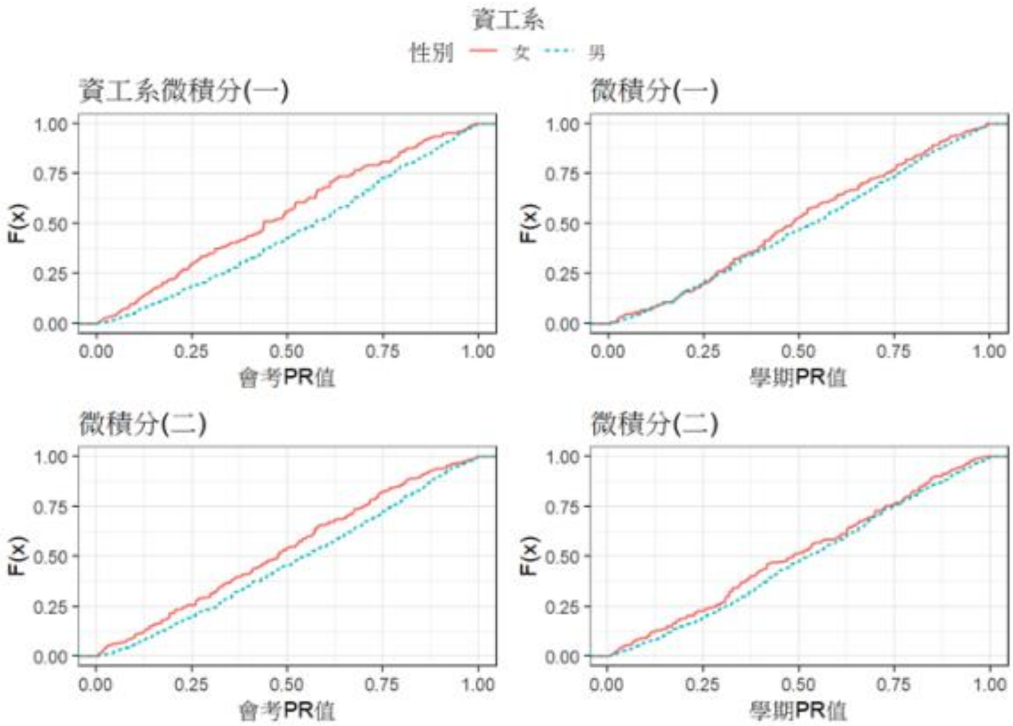
100-106學年度資工系大一專業科目表現與性別之關係

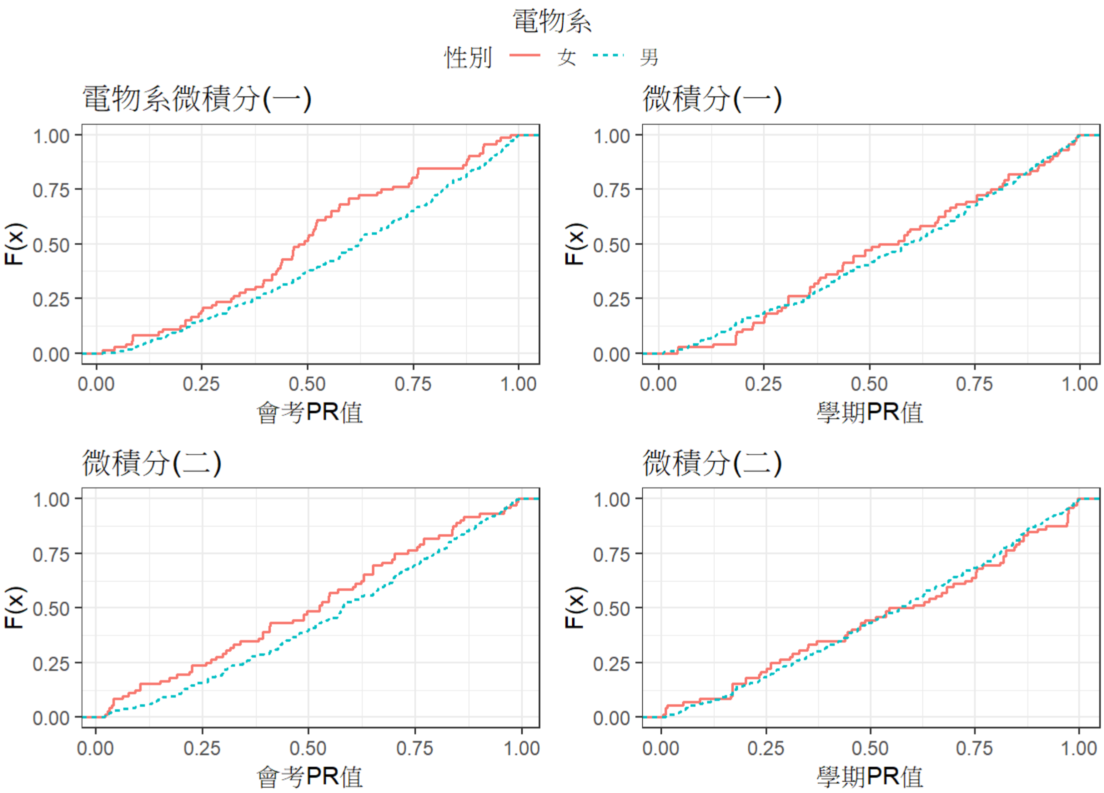
100-106學年度機械系大一專業科目表現與性別之關係

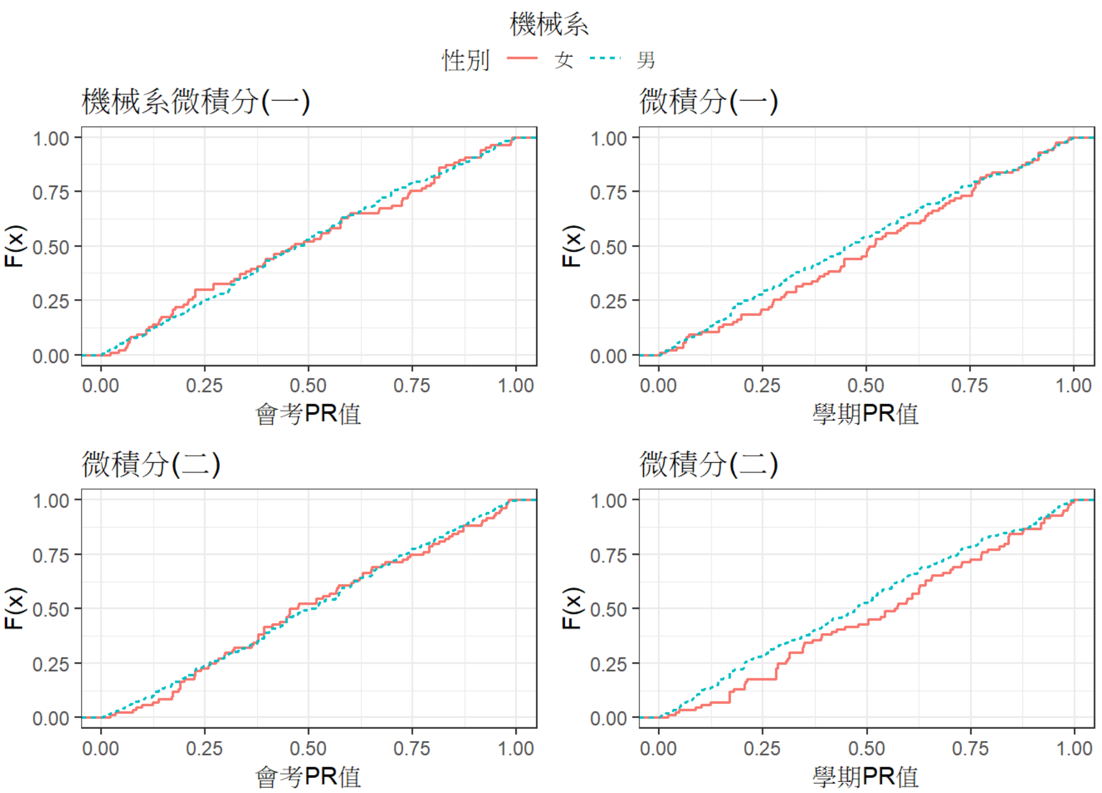



100-106學年度電物系大一專業科目表現與性別之關係
物理實驗(PR)
物理(PR)
化學實驗(PR)
化學(PR)
線性代數(PR)
計算機概論(PR)
F(x)
PR
性別
女
男


電工系
性別 女 男
電工系微積分(一)
微積分(一)
微積分(二)
微積分(二)
F(x)
會考PR值
學期PR值
0.00 0.25 0.50 0.75 1.00

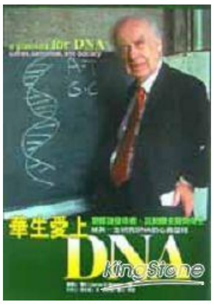

資工系
性別 女 男
資工系微積分(一)
微積分(一)
微積分(二)
微積分(二)
F(x)
會考PR值
學期PR值
0.00 0.25 0.50 0.75 1.00

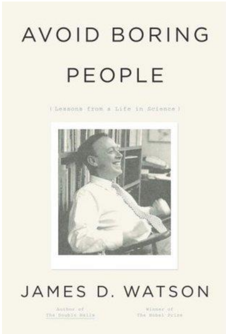

電物系
性別 女 男
電物系微積分(一)
微積分(一)
微積分(二)
微積分(二)
F(x)
1.00
0.75
0.50
0.25
0.00
會考PR值
學期PR值


機械系
性別 女 男
機械系微積分(一)
微積分(一)
微積分(二)
微積分(二)
F(x)
1.00
0.75
0.50
0.25
0.00
會考PR值
學期PR值

Nobel scientist snubbed after racism claims

By Stephen Adams 2:21PM BST 17 Oct 2007





Dr Watson is no stranger to controversy

The Science Museum last night cancelled a talk by Nobel Prize winning scientist Dr James Watson after he **was accused of making "racist" comments implying Africans were not as intelligent as whites.**

DNA pioneer Dr Watson, who discovered the double helix with Briton Francis Crick, has been roundly condemned for saying he was "inherently gloomy about the prospect of Africa" because "all our social policies are based on the fact that their intelligence is the same as ours – whereas all the testing says not really".

The 79-year-old American was due to talk at the Science Museum's Dana Centre on Friday but last night a spokesman said Dr Watson's comments had gone "beyond the point of acceptable debate". He announced the Musuem was cancelling the sold-out talk as a result.

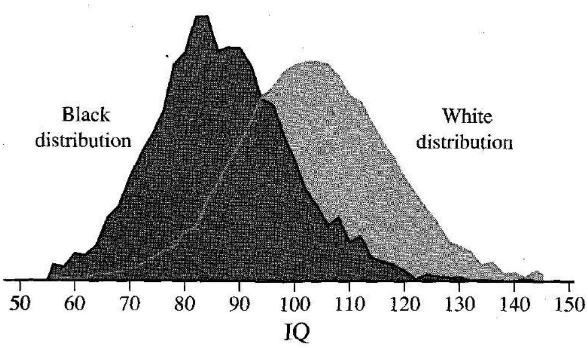
…

Despite the fierce barracking received by those who have put forward the theory of a racial basis for intellectual difference, the idea has refused to die.

IQ testing has consistently shown that racial groups perform differently, say advocates. In 1994 the publication of Richard Hernnstein and Charles Murray's book The Bell Curve

Both sides agree there is far more variation in intelligence and genetics within racial groups than between them.

However, opponents argue **IQ tests are culturally biased and say lower average scores among blacks can also be explained by social rather than genetic factors.**



James Watson Quits Post After Remarks on Races

By CORNELIA DEAN 26 Oct 2007

James D. Watson, the eminent biologist who ignited an uproar last week with remarks about the intelligence of people of African descent, **retired yesterday as chancellor of the Cold Spring Harbor Laboratory** on Long Island, and from its board.

In a statement, Dr. Watson noted that, at 79, he was "overdue" to surrender leadership positions at the laboratory, which he joined as director in 1968. He was president from 1994 until 2003. But he said the circumstances of his resignation "are not those which I could ever have anticipated or desired."

Dr. Watson, who shared a 1962 Nobel Prize for describing the double-helix structure of DNA, and later headed the American government's part in the international Human Genome Project, is leaving in a swirl of denunciations and canceled speaking invitations, including one from Rockefeller University, which yesterday canceled a lecture he was to have given next week.

Dr. Watson, who has long had a reputation for challenging scientific orthodoxy and for sometimes incendiary off-the-cuff remarks, **became a center of controversy last week after he was quoted in The Times of London as suggesting that, over all, people of African descent are not as intelligent as people of European descent.**

In the ensuing uproar, he issued a statement apologizing "unreservedly" for the comments, adding, "There is no scientific basis for such a belief."…

But a spokesman for the laboratory said yesterday that **while Dr. Watson would keep an office at the laboratory and would live there "in a house he built on land the laboratory owns," he would no longer have a job there.**

The furor erupted in England, where Dr. Watson was touring to promote his new book, "Avoid Boring People: Lessons from a Life in Science" (Knopf). **After his remarks were reported, the Science Museum in London canceled a sold-out speech he was scheduled to deliver and his comments were widely denounced as racist.**

The lecture Rockefeller University canceled yesterday was in connection with Dr. Watson's receipt of this year's Lewis Thomas Prize, which the university awards annually to scientists whose books help bridge the gap between the laboratory and the wider world. Dr. Watson is being honored for "The Double Helix," the book he wrote about the elucidation of DNA. Though he will still receive the prize, and the $5,000 it carries, "there were some members of **the university community who had expressed reservations about Dr. Watson coming here to speak after the controversy over his remarks in the U.K.**," Joseph Bonner, Rockefeller's director of communications, said yesterday. Mr. Bonner said that just as Rockefeller's president, Sir Paul M. Nurse, had decided to cancel the event, Dr. Watson called to suggest the same thing. Mr. Bonner said Sir Paul made the decision after consulting "various members of the university community and university leadership."