抽樣分配

前言

全球歷經2008金融海嘯後經濟普遍不景氣，報章雜誌報導台灣剛大學畢業的社會新鮮人月收入平均25,000元。若想探討這個主題，最好的方式就是針對台灣所有剛大學畢業的社會新鮮人進行月收入調查。根據教育部統計資料顯示，99年6月的大學畢業生人數是227,174人，要逐一的調查22萬人的月收入除了有人力物力的限制外更有時效性的考量。因此利用抽樣方法由22萬人抽出有效樣本成為探討社會新鮮人平均月收入的有效工具。

1

由22萬多名大學畢業生中隨機抽出1000人的好幾十萬種可能組合中當然就包含好幾十萬的樣本平均數、樣本中位數、樣本眾數、樣本全距與樣本標準差等統計量。因此好幾十萬的樣本平均數可以形成樣本平均數的機率分配；好幾十萬的樣本中位數可以形成樣本中位數的機率分配；好幾十萬的樣本標準差可以形成樣本標準差的機率分配。由樣本統計量所形成的機率分配統稱為抽樣分配。

2

抽樣方法

- 必須使用抽樣的方式從母體中抽出有限且具代表性的樣本，蒐集資料取得有效的數據對母體加以評估與預測。
- 抽樣的方式大致有四種：簡單隨機抽樣、系統抽樣分層抽樣、與集群抽樣。(參閱第二章抽樣)
- 抽樣的最終目的：

  以樣本資料推測母體趨勢，因此樣本平均數或樣本變異數的機率分配成為在建立統計推論前所應探討的主題。

3

樣本平均數的機率分配

假若 $X_1, X_2, \cdots, X_n$ 為取自於母體具有常態分配 $N(\mu, \sigma^2)$之一組大小為 n 的隨機樣本，則樣本平均數 $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ 具有常態分配 $N(\mu, \dfrac{\sigma^2}{n})$。

$$E(\overline{X}) = \mu \text{、} Var(\overline{X}) = \sigma^2 / n \text{；}$$

並以 $\overline{X} \sim N(\mu ; \sigma^2/n)$表示

4

例 題

一母體 1,3,5,7,9 含五數以抽出放回取 2 數，令隨機變數 $\overline{X}$ = 樣本空間內各樣本點的算術平均數。

(1)計算母體 μ、σ^2、σ

5

一母體 1,3,5,7,9 含五數以抽出放回取 2 數，令隨機變數 $\overline{X}$ = 樣本空間內各樣本點的算術平均數。

(2)列出樣本空間並計算樣本空間各樣本點 $\overline{X}$

樣本空間各樣本點 $\overline{X}$ 值與機率值

Ω	$\overline{X}$	f(Ω)	Ω	$\overline{X}$	f(Ω)
(1,1)	1	1/25	(5,7)	6	1/25
(1,3)	2	1/25	(5,9)	7	1/25
(1,5)	3	1/25	(7,1)	4	1/25
(1,7)	4	1/25	(7,3)	5	1/25
(1,9)	5	1/25	(7,5)	6	1/25
(3,1)	2	1/25	(7,7)	7	1/25
(3,3)	3	1/25	(7,9)	8	1/25
(3,5)	4	1/25	(9,1)	5	1/25
(3,7)	5	1/25	(9,3)	6	1/25
(3,9)	6	1/25	(9,5)	7	1/25
(5,1)	3	1/25	(9,7)	8	1/25
(5,3)	4	1/25	(9,9)	9	1/25
(5,5)	5	1/25			

例 題

一母體 1,3,5,7,9 含五數以抽出放回取 2 數，令隨機變數 $\overline{X}$ = 樣本空間內各樣本點的算術平均數。

(3)寫出 $\overline{X}$ 的機率分配

一母體 1,3,5,7,9 含五數以抽出放回取 2 數，令隨機變數 $\overline{X}$ = 樣本空間內各樣本點的算術平均數。

(4)求 $E(\overline{X})$、$Var(\overline{X})$、$\sigma_{\overline{X}}$

例 題

已知一常態母體$\mu = 50$、$\sigma^2 = 16$，由其中以抽出放回方式分別取16、25、64個樣本試分別

(1)求 $E(\overline{X})$、$Var(\overline{X})$、$\sigma_{\overline{x}}$　(2)寫出$\overline{X}$的機率分配

例 題

某藥商研發一種新藥物為了治療心臟疾病，這種新藥一顆重量具有平均數為 5.8mg，標準差 1.1 mg 的常態分配。今從製程中隨機抽取一樣本數為 16 的樣本，試求其平均重量介於 5.1mg 與 6.2mg 之間的機率為何？

由題意知，此樣本平均數的分配為一平均數 5.8 mg，標準差 (1.1/4)mg 的常態分配。令 $\overline{X}$ 為該地區中隨機抽取 16 顆新藥物平均重量，因此，

中央極限定理

(中央極限定理(Central Limit Theorem , CLT))

設 $X_1, X_2, \cdots$ 爲彼此互相獨立且具有相同分配的隨機變數序列。令 $E\,[X_i] = \mu$, $Var[X_i] = \sigma^2 < \infty$, $\forall i$，則

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{n \to \infty} N(\mu, \frac{\sigma^2}{n})$$

當母體夠大或所抽出的樣本數大於等於30時，不論母體的實際分配及抽樣的方式，樣本平均數機率分配的分佈趨勢均會近似於常態分配。

例 題

根據研究一般電子公司員工不支薪假期為常態分配，假設母體平均就是 8.5 星期，而標準差為 2 星期。

⑴隨機抽出 16 名員工，試寫出平均不支薪假期的機率分配。

⑵計算 16 名員工平均不支薪假期大於 9.5 星期的機率是多少？

⑶計算 16 名員工平均不支薪假期介於 8 到 9 星期的機率是多少？

例 題

1992 年美國報導鮪魚是美國人最喜愛的食用魚，每人平均每年吃 3.6 磅，且標準差為 1.5 磅。問：

⑴隨機抽出 100 人，顯示其平均消費鮪魚 $\overline{X}$ 的抽樣分配。

13

例 題

1992 年美國報導鮪魚是美國人最喜愛的食用魚，每人平均每年吃 3.6 磅，且標準差為 1.5 磅。問：

⑵隨機抽出 100 人，顯示其平均消費鮪魚 $\overline{X}$ 會大於或等於 4 磅的機率是多少？

⑶隨機抽出 100 人，顯示其平均消費鮪魚 $\overline{X}$ 會介於 3.2 到 4 磅的機率是多少？

14

例 題

過去吳郭魚一直被視爲廉價漁產品，由於近年來養殖技術的提升，吳郭魚在肉質的方面，受到消費者的青睞，因而大受歡迎，且在外銷市場也佔有一席之地，現今有一批養殖吳郭魚出口，已知吳郭魚魚身長爲常態分配具有平均數 19cm 及標準差 2cm，請問從這批吳郭魚中抽取 11 隻，問平均身長小於 20cm 的機率爲何？

15

t 分配

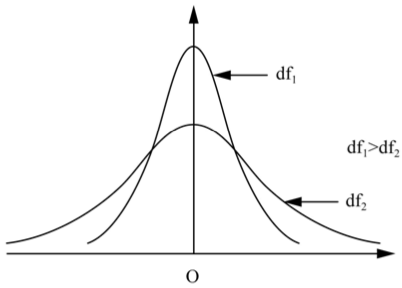
*t*分配(Student's t-distribution)為英國學者*Gosset*(1876-1937)於1908年以筆名*Student*所發表，故取名為*t*分配，這是一種專門討論母體變異數未知與小樣本(n<30)的分配。

與常態分配比較起來，它也是一種對稱於平均數的機率分配，且其平均數和標準常態分配一樣是0，但與常態分配最大的不同點是*t*分配的形狀隨著自由度(Degree of freedom，簡稱*d.f.*)的不同而改變。(*t*分配中的自由度為樣本數減1，即*d.f.*=*n*-1)

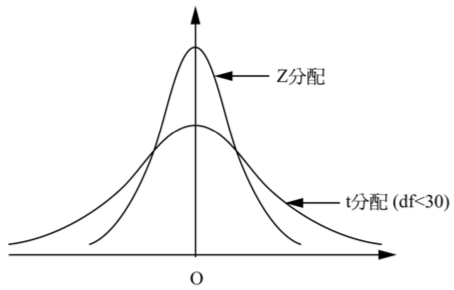
16

t 分配

1. t分配為一連續的機率分配，其曲線近似鐘型曲線
2. 曲線下面積為1，且以0為中心，二側互相對稱
3. $\infty < t < \infty$



不同自由度下之 t 分配圖形



t 分配與標準常態分配之圖形比較

t 分配

z 分配　$z = \dfrac{\overline{X} - \mu}{\delta / \sqrt{n}}$

t 分配：以樣本標準差取代母體標準差

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}}$$

令隨機變數：$t = \dfrac{\overline{X} - \mu}{S / \sqrt{n}}$

其中$\overline{X}$ 為樣本平均數，μ 為母體平均數，
S為樣本標準差，n 為樣本數，

則t所形成的分配為自由度d.f.=n-1的t分配；

$$E(t) = 0; \quad Var(t) = \frac{d.f.}{d.f. - 2} \quad d.f. > 2$$

18

t分配中的自由度

- t分配中的自由度意指在樣本標準差已知的情況下，則 $\overline{X}$ 必已知，因此研究者只需隨機抽取n-1個樣本資料後，即可知全部n個資料值。

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

t−分配表（右尾機率值）

df	α 0.1000	0.0500	0.0250	0.0100	0.0050	0.0010	0.0005
1	3.0777	6.3137	12.7062	31.8210	63.6559	318.2888	636.5776
2	1.8856	2.9200	4.3027	6.9645	9.9250	22.3285	31.5998
3	1.6377	2.3534	3.1824	4.5407	5.8408	10.2143	12.9244
4	1.5332	2.1318	2.7765	3.7469	4.6041	7.1729	8.6101
5	1.4759	2.0150	2.5706	3.3649	4.0321	5.8935	6.8685
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.2075	5.9587
7	1.4149	1.8946	2.3646	2.9979	3.4995	4.7853	5.4081
8	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0414
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.2969	4.7809
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5868
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.0248	4.4369
12	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2209

t－分配表（右尾機率值）

df	α = 0.1000	0.0500	0.0250	0.0100	0.0050	0.0010	0.0005
14	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1403
15	1.3406	1.7531	2.1315	2.6025	2.9467	3.7329	4.0728
16	1.3368	1.7459	2.1199	2.5835	2.9208	3.6861	4.0149
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9217
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.5793	3.8833
20	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8496
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
40	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510
50	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.4960
100	1.2901	1.6602	1.9840	2.3642	2.6259	3.1738	3.3905
∞	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905

21

例 題

求出n=10的情況下，

(1)P(t>4.2969)=

(2)P(t<2.2622)=

(3)P(2.2622<t<4.2969)=

22

例 題

求出n=10的情況下，

(4)$P(t>3)=$

(5)$P(t<1)=$

(6)$P(-1<t<2)=$

(7)$P(|t|<1)=$

利用Excel求 t 分配機率

TDIST 函數

傳回百分比點數的 (機率) 的 Student t-分配，該數值 (x) 是 t-分配的計算值，此計算值是以百分比來計算。

t-分配是用於小樣本資料組的假設檢定。使用此函數就不用建立 t-分配的臨界值表格。

語法

TDIST(x,degrees_freedom,tails)

X　是要用來評估分配的數值。 **Degrees_freedom**　是用來指出自由度的整數值。

Tails　指定要傳回的分配尾數的個數。如果 tails = 1，則 TDIST 傳回單尾分配。如果 tails = 2，則 TDIST 傳回雙尾分配。

備註

- 如果 tails = 1，則 TDIST 以 TDIST = P(X>x) 來計算，其中 X 為 t 分配後面的隨機變數)。

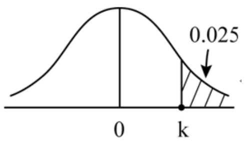
  如果 tails = 2，則 TDIST 以 TDIST = P(|X| > x) = P(X > x or X < -x) 來計算。
- 因為 x < 0 是不允許的，要在 x < 0 時使用 TDIST，請注意 TDIST(-x,df,1) = 1 - TDIST(x,df,1) = P(X > -x)

  且 TDIST(-x,df,2) = TDIST(x df,2) = P(|X| > x)。

例 題

求出在樣本數為26的情況下，

(1) $P(t > k)$=0.025的臨界值k=？ (2) $P(|t| < k)$=0.9的臨界值k=？



卡方分配

- 重複抽樣，並計算各別變異數。這些變異數的分配並不是常態，而是一種不對稱的右偏分配。
- 進而分別對各個變異數值，乘以n-1（n為樣本數），再除以母體變異數（常數），便形成一個新的統計量，稱為卡方變數，即

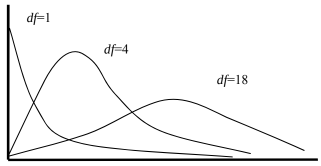
$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum^{n}(X_i - \bar{X})^2}{\sigma^2}$$

- 卡方統計量的分配型態，係由其自由度所決定，記作

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

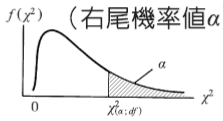
卡方分配的性質

- 卡方分配的機率密度函數十分複雜。在此，僅介紹卡方分配的一些重要特性：
  - 卡方統計量之值必為正數（因為卡方公式的分子、分母均為平方之故）。卡方分配圖的橫軸，定義於0~∞的範圍。
  - 卡方分配之平均數和變異數為 $E(\chi^2) = df$
    $Var(\chi^2) = 2df$
  - 卡方分配為一右偏分配，其形狀因自由度而異，當自由度愈大，向右偏斜程度愈小。



不同自由度的卡方分配圖

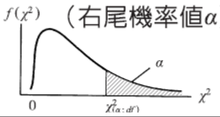
卡方分配表



（右尾機率值α）

df	α 0.9950	0.9900	0.9750	0.9500	0.9000	0.1000	0.0500	0.0250	0.0100	0.0050	0.0010	0.0005
1	0.0000	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415	5.0239	6.6349	7.8794	10.8274	12.1153
2	0.0100	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915	7.3778	9.2104	10.5965	13.8150	15.2014
3	0.0717	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147	9.3484	11.3449	12.8381	16.2660	17.7311
4	0.2070	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1433	13.2767	14.8602	18.4662	19.9977
5	0.4118	0.5543	0.8312	1.1455	1.6103	9.2363	11.0705	12.8325	15.0863	16.7496	20.5147	22.1057
6	0.6757	0.8721	1.2373	1.6354	2.2041	10.6446	12.5916	14.4494	16.8119	18.5475	22.4575	24.1016
7	0.9893	1.2390	1.6899	2.1673	2.8331	12.0170	14.0671	16.0128	18.4753	20.2777	24.3213	26.0179
8	1.3444	1.6465	2.1797	2.7326	3.4895	13.3616	15.5073	17.5345	20.0902	21.9549	26.1239	27.8674
9	1.7349	2.0879	2.7004	3.3251	4.1682	14.6837	16.9190	19.0228	21.6660	23.5893	27.8767	29.6669
10	2.1558	2.5582	3.2470	3.9403	4.8652	15.9872	18.3070	20.4832	23.2093	25.1881	29.5879	31.4195
11	2.6032	3.0535	3.8157	4.5748	5.5778	17.2750	19.6752	21.9200	24.7250	26.7569	31.2635	33.1382
12	3.0738	3.5706	4.4038	5.2260	6.3038	18.5493	21.0261	23.3367	26.2170	28.2997	32.9092	34.8211
13	3.5650	4.1069	5.0087	5.8919	7.0415	19.8119	22.3620	24.7356	27.6882	29.8193	34.5274	36.4768
14	4.0747	4.6604	5.6287	6.5706	7.7895	21.0641	23.6848	26.1189	29.1412	31.3194	36.1239	38.1085
15	4.6009	5.2294	6.2621	7.2609	8.5468	22.3071	24.9958	27.4884	30.5780	32.8015	37.6978	39.7173
16	5.1422	5.8122	6.9077	7.9616	9.3122	23.5418	26.2962	28.8453	31.9999	34.2671	39.2518	41.3077
17	5.6973	6.4077	7.5642	8.6718	10.0852	24.7690	27.5871	30.1910	33.4087	35.7184	40.7911	42.8808
18	6.2648	7.0149	8.2307	9.3904	10.8649	25.9894	28.8693	31.5264	34.8052	37.1564	42.3119	44.4337

卡方分配表



df	α											
	0.9950	0.9900	0.9750	0.9500	0.9000	0.1000	0.0500	0.0250	0.0100	0.0050	0.0010	0.0005
19	6.8439	7.6327	8.9065	10.1170	11.6509	27.2036	30.1435	32.8523	36.1908	38.5821	43.8194	45.9738
20	7.4338	8.2604	9.5908	10.8508	12.4426	28.4120	31.4104	34.1696	37.5663	39.9969	45.3142	47.4977
21	8.0336	8.8972	10.2829	11.5913	13.2396	29.6151	32.6706	35.4789	38.9322	41.4009	46.7963	49.0096
22	8.6427	9.5425	10.9823	12.3380	14.0415	30.8133	33.9245	36.7807	40.2894	42.7957	48.2676	50.5105
23	9.2604	10.1957	11.6885	13.0905	14.8480	32.0069	35.1725	38.0756	41.6383	44.1814	49.7276	51.9995
24	9.8862	10.8563	12.4011	13.8484	15.6587	33.1962	36.4150	39.3641	42.9798	45.5584	51.1790	53.4776
25	10.5196	11.5240	13.1197	14.6114	16.4734	34.3816	37.6525	40.6465	44.3140	46.9280	52.6187	54.9475
26	11.1602	12.1982	13.8439	15.3792	17.2919	35.5632	38.8851	41.9231	45.6416	48.2898	54.0511	56.4068
27	11.8077	12.8785	14.5734	16.1514	18.1139	36.7412	40.1133	43.1945	46.9628	49.6450	55.4751	57.8556
28	12.4613	13.5647	15.3079	16.9279	18.9392	37.9159	41.3372	44.4608	48.2782	50.9936	56.8918	59.2990
29	13.1211	14.2564	16.0471	17.7084	19.7677	39.0875	42.5569	45.7223	49.5878	52.3355	58.3006	60.7342
30	13.7867	14.9535	16.7908	18.4927	20.5992	40.2560	43.7730	46.9792	50.8922	53.6719	59.7022	62.1600
40	20.7066	22.1642	24.4331	26.5093	29.0505	51.8050	55.7585	59.3417	63.6908	66.7660	73.4029	76.0963
50	27.9908	29.7067	32.3574	34.7642	37.6886	63.1671	67.5048	71.4202	76.1538	79.4898	86.6603	89.5597
60	35.5344	37.4848	40.4817	43.1880	46.4589	74.3970	79.0820	83.2977	88.3794	91.9518	99.6078	102.6971
70	43.2753	45.4417	48.7575	51.7393	55.3289	85.5270	90.5313	95.0231	100.4251	104.2148	112.3167	115.5766
80	51.1719	53.5400	57.1532	60.3915	64.2778	96.5782	101.8795	106.6285	112.3288	116.3209	124.8389	128.2636
90	59.1963	61.7540	65.6466	69.1260	73.2911	107.5650	113.1452	118.1359	124.1162	128.2987	137.2082	140.7804
100	67.3275	70.0650	74.2219	77.9294	82.3581	118.4980	124.3421	129.5613	135.8069	140.1697	149.4488	153.1638

常態分配與卡方分配

df = 24

(a)$P(9.8862 \le \chi^2 \le 39.3641)$

(b)$P(t \le \chi^2 \le 45.7222) = 0.95$

(c)$P(10.8564 \le \chi^2 \le t)$

(d)$P(\chi^2 < t) = 0.005$

利用Excel求卡方分配機率

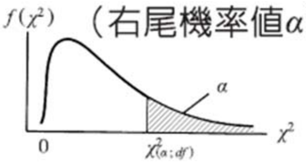
CHIDIST(x,degrees_freedom)

X 是要評估其分配的數值。

Degrees_freedom 為自由度。

備註 ▪ CHIDIST = P(χ^2>x) 方式來計算，

（右尾機率值α）



$Ex:$

$\mathrm{df}=24 \quad \mathrm{P}(9.8862 \le \chi^2 \le 39.3641)$

$$= P(\chi^2 > 9.8862) - P(\chi^2 > 39.3641)$$

$$= 0.995 - 0.025 = 0.97$$

常態分配與卡方分配

(1)若 $X \sim N(0,1)$ 且 $Y = X^2$，則 $Y \sim \chi^2(1)$。

(2)設 $X_1, X_2, \cdots, X_n$ 爲獨立隨機變數，若 $X_i \sim N(0,1), i=1,2,\cdots,n$

則 $Y = \sum_{i=1}^{m} X_i^2 \sim \chi^2(n)$。

(3)設 $X_1, X_2, \cdots, X_n$ 爲獨立隨機變數，若 $X_i \sim N(\mu_i, \sigma_i^2), i=1,2,\cdots,n$

則 $Y = \sum_{i=1}^{n} (\frac{X_i - \mu_i}{\sigma_i})^2 \sim \chi^2(n)$。

例 題

若 $X \sim N(5,4)$，求 $P\{15.364 < (X-5)^2 < 20.096\}$