Sample Distributions

關於“平均數”與“比例”的抽樣分配

抽樣分配

統計對象	型態	分配類型	
母體分配 樣本分配	離散型 機率分配	均勻分配 多項分配 超幾何分配 伯努利分配	二項分配 負二項分配 幾何分配 波松分配
	連續型 機率分配	連續均勻分配 標準常態分配 Gamma分配	常態分配 指數分配 卡方分配
抽樣分配	機率分配	樣本比例差（$\lvert\hat{p}-p\rvert$） 平均數～ Z分配	 t分配
		變異數～ 單母體變異數 兩母體變異數比	 卡方分配 F分配

Sampling Distribution of the Proportion

樣本比率的 $\hat{p}$ 抽樣分配

函數定義（分配形式）

在抽樣調查實務中，經常對比率的問題感到興趣，例如偶像劇的收視率、政黨支持的比率、學生打工比率、產品生產瑕疵品比率等問題。若是母體的比率未知，就必須透過抽樣方式來估計，假定：

母體比率與樣本比率

$$p = \frac{K}{N}$$

，其中N：母體個數；
K：母體中某類別的總個數

$$\hat{p} = \frac{k}{n}$$

，其中n：樣本個數；
k：樣本中抽出某類別的個數

抽樣誤差

抽樣誤差 =$\left|\text{母體參數} - \text{樣本統計量}\right| = \left|\, p - \hat{p} \,\right|$

函數定義（分配形式）

樣本比率的抽樣分配 ~ 採取出放回或**無限母體**

$\hat{p} = \frac{x}{n} \quad \Rightarrow \quad x = n\hat{p}$ ，帶入**二項分配**的機率函數，可得：

$$f(\hat{p}) = C^{n}_{n\hat{p}}\, p^{n\hat{p}}\, q^{n-n\hat{p}}, \quad \hat{p} = 0, \frac{1}{n}, \frac{2}{n} .., 1$$

期望值 $\mathrm{E}(\hat{p}) = p$

變異數 $\mathrm{V}(\hat{p}) = \frac{p(1-p)}{n}$

$$\mathrm{E}(\hat{p}) = \mathrm{E}\left(\frac{x}{n}\right) = \frac{1}{n}E(x) = \frac{1}{n}(np) = p$$

$$\mathrm{V}(\hat{p}) = V\left(\frac{x}{n}\right) = \frac{1}{n^2}V(x) = \frac{1}{n^2}(npq)$$

$$= \frac{pq}{n} = \frac{p(1-p)}{n}$$

函數定義（分配形式）

樣本比率的抽樣分配 ~ 採取出不放回且**有限母體**

$\hat{p} = \frac{x}{n} \quad \Rightarrow \quad x = n\hat{p}$ ，帶入**超幾何分配**的機率函數：

$$f(\hat{p}) = \frac{C^{S}_{n\hat{p}}\, C^{N-S}_{n-n\hat{p}}}{C^{N}_{n}},\ \hat{p}=0, \frac{1}{n}, \frac{2}{n} .., 1$$

有限母體修正因子
（超幾何分配）

期望值 $\mathrm{E}(\hat{p}) = p$

變異數 $\mathrm{V}(\hat{p}) = \frac{p(1-p)}{n} \times \frac{N-n}{N-1}$

大樣本時的樣本比率抽樣分配

如果取出的樣本數很大時，np ≧ 5 且 n(1-p) ≧ 5，
根據中央極限定理：

採取出放回或**無限母體**

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right) \Rightarrow Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

採取出不放回且**有限母體**

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n} \times \frac{N-n}{N-1}\right)$$

案例

假設台北市國小學生有10%的比例有體重過重的現象，現在隨機抽取400位台北市國小學生，其體重過重的比例之期望值及變異數為多少？又體重過重之比率大於12%之機率為多少？

「現代統計學」p123，吳柏林 著，五南圖書

案例解說

10%比例 → p = 0.1 ；抽取400人 → n=400 ；→ 符合二項分配

(1) E($\hat{p}$) = p = $\underline{0.1}$, $V(\hat{p}) = \frac{p(1-p)}{n} = \frac{0.1\times 0.9}{400} = \underline{0.000225}$

(2)P($\hat{p} > 0.12$)= $p\left(\frac{\hat{p}-0.1}{\sqrt{0.000225}} > \frac{0.12-0.1}{\sqrt{0.000225}}\right)$

= P(Z > 1.33) [查表]

= 1-0.9082 ＝ $\underline{0.0918}$

Sampling Distribution of the Mean

樣本平均數 $\bar{x}$ 的抽樣分配

函數定義（分配形式）

利用樣本平均數 $\bar{x}$ 來推論出母體的平均數μ，來瞭解母體集中趨勢。假定現在從母體中隨機抽取 n 個樣本，根據「中央極限定理」，當 n 很大時，樣本平均數 $\bar{x}_n$ 將趨近於平均數 μ ，變異數 σ^2/n 的常態分配：

$$\mathrm{Z}_{\bar{x}_n} = \frac{\bar{x}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \longrightarrow N(0,1)$$

轉化為標準常態分配後，就可以利用查表方式將結果查出。

樣本變異數 S^2

我們利用樣本平均數 $\bar{x}$ 來推論出母體的平均數μ，同樣地，根據變異數σ^2的原始定義，我們可以使用樣本變異數 S^2 來推論母體的變異數：

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

「n-1」稱為〝自由度〞（degree of freedom, df）

S^2 也稱為σ^2的「不偏估計量」，因為：

$$E(S^2) = E\left(\frac{\sum (x_i - \bar{x})^2}{n-1}\right) = \sigma^2$$

Student's t distribution

t 分配

函數定義（分配形式）

t分配由英格蘭化學家W.S.Gosset在1908年所提出，他將Z分配與卡方分配結合創造了t分配。其適用在當我們實務上不知道母體變異數的狀況下，可以利用t分配來取代Z分配。也就是說，只要母體為常態分配且母體變異數未知，就可以用t分配來估計母體平均數。

$$t = \underbrace{\frac{\bar{x}_n - \mu}{\sqrt{\frac{S^2}{n}}}} \longrightarrow t(\underline{\mathrm{n}-1})$$

可藉由查表而得

t分配特性

若隨機變數X為具自由度ν之t分配

期望值　$E(x) = 0$，$\nu > 1$

變異數　$V(x) = \frac{\nu}{\nu-2}$，$\nu > 2$

偏態係數　$\beta_1 = 0$

峰度係數　$\beta_2 = \frac{3(\nu-2)}{\nu-4}$，$\nu > 4$

$n \rightarrow \infty$時，t 分配趨近N(0,1)

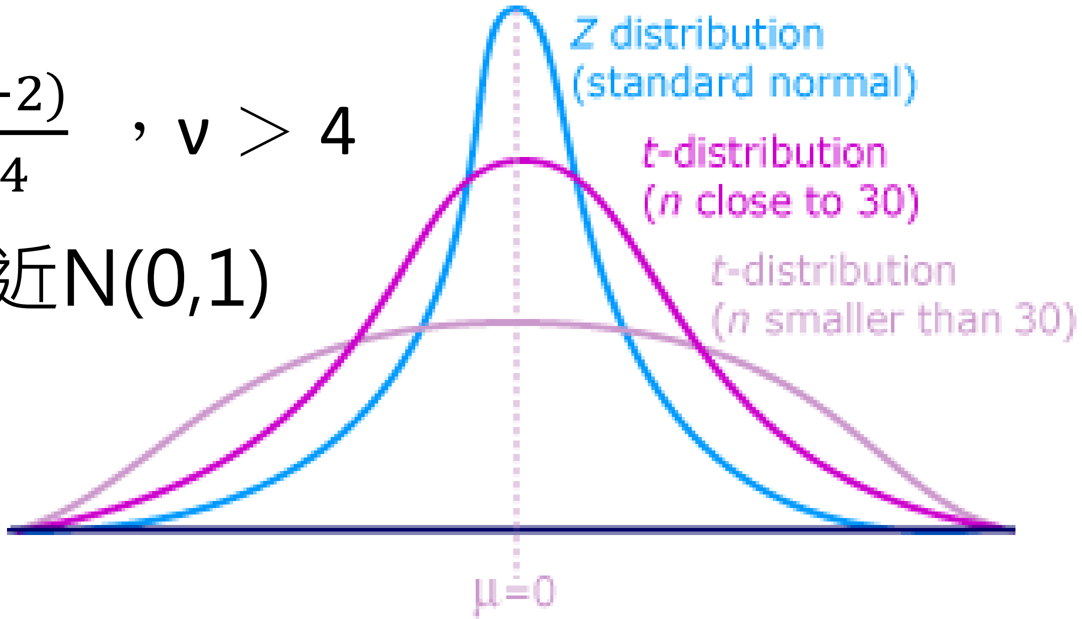
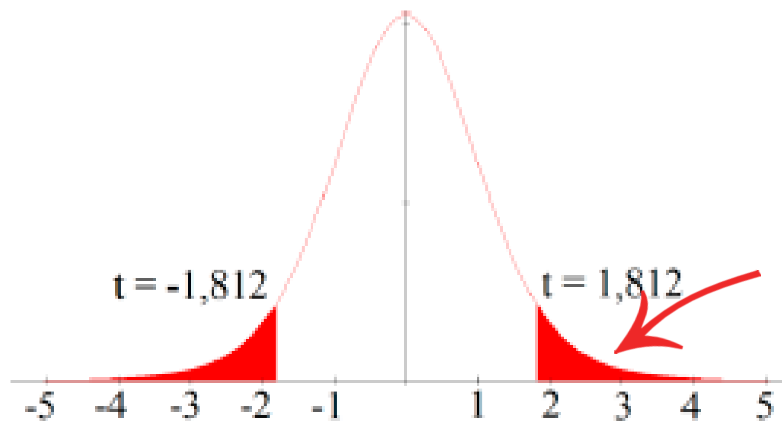


TABLA
DISTRIBUCIÓN t DE STUDENT



Ejemplos:

Para n-1 = 10 grados de libertad

$$\boldsymbol{P(t > 1,812) = 0,05}$$

$$\boldsymbol{P(t < -1,812) = 0,05}$$

α / n-1	0,25	0,2	0,15	0,1	**0,05**	0,025	0,01	0,005	0,0005
1	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	636,6192
2	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
8	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,6998	0,8791	1,0931	1,3722	**1,8125**	2,2281	2,7638	3,1693	4,5869
11	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370

t分配使用時機

樣本抽自母體，n ＜ 30，母體變異數未知時，可用t 分配估計母體平均數。

兩個常態母體變異數未知，且樣本皆為小樣本情況下，兩母體平均差 μ_1-μ_2 的推論。

可用於變異數分析的事後檢定，以及相關分析、迴歸係數之推論。

案例

假設X代表某班級統計學成績，已知其為常態分配，X ~ N(80, σ2)，現自該班級抽出10位學生，算出其標準差等於8.63，求此10位學生之平均成績在85分以上的機率是多少？

「應用統計學 二版」p215，李德治、童惠玲 著，博碩文化

案例解說

X ~ N(80, σ^2) → 常態分配，小樣本(10)且變異數「未知」→ 考慮 t分配

根據題目所提供資訊： μ = 80， S^2=8.63 ，自由度 = 9

$$P(\bar{x} > 85) = p\left(t_9 > \frac{85-80}{\sqrt{\frac{8.63^2}{10}}} \right)$$

$$= P(t_9 > 1.83) \quad [查表]$$

$$= \underline{0.05}$$

The End