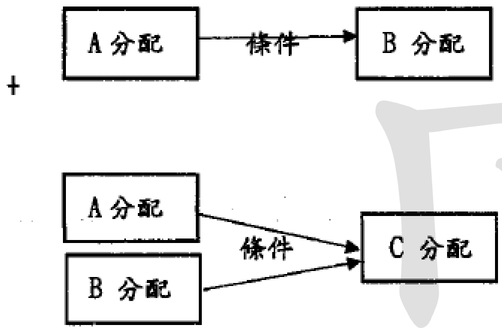
國立臺灣大學109學年度轉學生招生考試試題

題號：23　　　　　　　　　　　　題號：23
科目：心理及教育統計學　　　　　　共 1 頁之第 1 頁

1. 思考一下你所知道的<u>所有</u>隨機變數的分配 (distributions)，然後來回答下面的問題：
假設有 A 分配、B 分配，若 A 分配在某些條件之下會是 B 分配，請你用下面圖示的方式來表達出來，並請記得清楚寫出各分配的名稱及該些條件。有可能箭頭的左邊不只一個分配，也就是說可能有超過一個分配的組合 (如 A、B 分配)在某些條件下而變成另一個分配 (如 C 分配)。　[20%]



講義 P81~P82 分配間關係

※ F 分配的特性

1. 由於 F 為兩平方量之比值，故 $0 \leq F \leq \infty$，即 F 均為正值。
2. 每一對 df_1，df_2 就有一條分配曲線，如圖 9-5。
3. 與 χ^2 分配一樣，F 分配一為正偏分配，惟當 df_1，df_2 增加時，偏斜程度降低。
4. $df_2 \rightarrow \infty$時，F=1；當 $df_1=1$，$df_2=\infty$時，$\sqrt{F}$ 會趨近於 Z 分配。
5. 曲線下面積為 1，故 F 分配為機率分配，其機率值可查 F 分配表。
6. F 分配常用來檢定兩常態母體之變異數是否相同。

$$\text{假設 } \sigma_1^2=\sigma_2^2 \text{，則 } F=\frac{S_1^2}{S_2^2}$$

II. 各種分配的歸納與比較

(1) χ^2 分配、F 分配、t 分配的共同特性

i. 三者皆為小樣本分配。
ii. 三者皆為連續分配。
iii. 三者皆為最重要的抽樣分配。
iv. 三者所來自的母體皆為常態分配。
v. 三者的母體都是自由度運算而來。
vi. 三者皆能在某些條件下轉換成標準常態分配：

$df \rightarrow \infty,\ \sqrt{2\chi^2} \sim N.D.(\sqrt{2df-1}, 1)$

$df \rightarrow \infty,\ t \sim N.D.(0,1)$

$df_1=1,\ df_2 \rightarrow \infty,\ \sqrt{F} \sim N.D.(0,1)$

(2) t 分配與 Z 分配的比較

由 t 分配與標準常態分配(Z 分配)之重要表徵數的比較得知：

※ 相同處

1. 範圍在- ∞與+∞之間。
2. 以 0 為中心的左右對稱分配。

※ 相異處

1. t 分配的變異程度較大。
2. t 分配的峰態為高狹峰，Z 分配的峰態為常態峰。
3. 只有 t 分配會受自由度影響

(3) χ^2、F、t 統計量的定義與實用公式：

χ^2、F、t 統計量所來自之母體分配皆為常態分配，當母體之平均數 μ，變異數 σ^2 已知，則定義公式得以成立；但母數 μ、 σ^2 常是統計推論的對象，此時就該使用實用公式。

(4) Z、χ^2、F、t 統計量間的關係

在自由度不同下，Z、χ^2、F、t 統計量間的關係為：

$$1.F_{1-\alpha(1,df)} = t^2_{1-\frac{\alpha}{2}(df)}$$
$$2.F_{1-\alpha(1,\infty)} = Z^2_{1-\frac{\alpha}{2}}$$
$$3.dfF_{1-\alpha(df,\infty)} = \chi^2_{1-\alpha(df)}$$
$$4.df/F_{1-\alpha(\infty,df)} = \chi^2_{\alpha(df)}$$
$$5.\chi^2_{1-\alpha(1)} = Z^2_{1-\frac{\alpha}{2}}$$
$$6.t_{1-\alpha(\infty)} = Z_{(1-\alpha)}$$

3. 我們會用樣本平均數（$\bar{X}$ ）來估計母群的平均數（μ），請從<u>五個角度</u>來討論樣本平均數（$\bar{X}$ ）是不是母群的平均數（μ）的一個好的估計值的？ [4% x 5=20%]

講義 P88~89 優良點估計值條件

二、良好點估計的判斷標準：

1. 不偏性(unbiasedness)

樣本統計量 $\hat{\theta}$ 的期望值等於母群參數 θ，即 E($\hat{\theta}$)=θ，則稱 $\hat{\theta}$ 為 θ 的不偏估計值(unbiased estimator)。例如：E($\bar{X}$)=μ，

$E(\frac{\sum(X-\bar{X})^2}{n-1}) = \sigma^2, E(\frac{\sum(Y-\hat{Y})^2}{n-2}) = \sigma^2_{Y\cdot X}$, 所以$\bar{X}, \hat{S}^2, \hat{S}^2_{Y\cdot X}$分別是母群$\mu, \sigma^2, \sigma^2_{Y\cdot X}$的「不偏估計值」。

證明：$E(\hat{S}^2_X) = \sigma^2_X$

4. 甚麼是自由度(degree of freedom)? [5%] 就你所學過的統計（如 標準差、卡方檢定、t 檢定、迴歸、變異數分析等），在計算時怎麼定出自由度? （請把條件寫清楚） [15%]

講義 P77 自由度觀念

f. 自由度

自由度(degree of freedom,以 df 表示)(或 v 表示)
指樣本中能夠自由變動數值的個數。若樣本中有 n 個數值，則自由度 df=n，每增加一個限制(利用一個樣本估計值去推論一個母群參數)，則原來能自由變動數值就少掉一個，故自由度變為 df=n-1。

例: $(X_1+X_2+X_3+X_4+X_5)/n=\overline{X}$

$(X_1,\ldots\ldots X_5$，都可以自由指定，故 df=5)
當$(6+7+8+4+X_5)/5=6$

(當 $\overline{X}$ 指定為 6 時(因為母群平均數為 6)，任意指定 $X_1=6,X_2=7,X_3=8$；

$X_4=4$ 此時 X_5 就不能再自由指定了，而必須強迫為 5，因此自由度少掉一個，df=5-1)

5. 甚麼是中央極限定理 (Central Limit Theorem)? [10%] 它在推論統計的假設檢定(hypothesis testing)中扮演甚麼角色? [10%]

講義 P76 中央極限定理

d. 中央極限定理(Central limit theorem, CLT)以 $\overline{X}$ 分配為例
自平均數 μ，標準差為 σ 的母群中隨機抽取樣本大小為 n(n>30)的一組樣本

$(X_1,X_2,X_3,\ldots X_n)$，並求其平均數，稱為樣本平均數 $\overline{X}$，如此重複無限多次，即可得到無限多個 $\overline{X}$，則此無限多個 $\overline{X}$ 所形成的分配就稱為「$\overline{X}$ 分配」。
該分配為常態分配，期望值(或平均數)為 $\mu(E(\overline{X})=\mu)$，變異數為

$\frac{\sigma_x^2}{n}(V(\overline{X})=\frac{\sigma_x^2}{n})$，開根號為 $\frac{\sigma_x}{\sqrt{n}}$ 即標準誤。抽樣分配形成所依據的理論稱為「中央極限定理」。

國立臺灣大學107學年度轉學生招生考試試題

題號： 23
科目：心理及教育統計學

題號： 23
共 / 頁之第 / 頁

3. 通常我們用什麼樣的統計指標評估一估計值（estimator）的「有效性」（efficiency）？請列出式子並舉例說明之。【10分】

命中講義 P.89

2. 有效性(efficiency)
   (1) 在其他條件不變(如 n 固定)，母群參數的所有不偏估計量中，具有最小變異數者，稱為「最有效的估計量」。
   (2) 若 $E(\hat{\theta}_i)=\theta$，$i=1, 2, 3, \ldots, k$，則 $\min Var(\hat{\theta}_i)$就是「最有效估計量」(most efficient estimator)。
   (3) 例如，$\overline{X}$ 和 Md 皆為母群平均數 μ 的不偏估計量，但因為 R.E(相對有效性)=$\dfrac{Var(Md)}{Var(X)}=157\%$，所以 $Var(Md)>Var(\overline{X})$，亦即$\overline{X}$比 Md 的變異數更小，所以$\overline{X}$比 Md 的變異數更小，所以$\overline{X}$比 Md 更合乎有效性。
3. 一致性(consistency)
   所謂「一致性」是指樣本容量無限增大時，估計值應越來越接近他的母群參數。即 $n\to\infty$時，$\overline{X}\to\mu$，$\hat{S}^2\to\sigma^2, \hat{S}^2_{Y\cdot X}\to\sigma^2_{Y\cdot X}$。
4. 充份性(sufficiency)
   指一個容量為 n 的樣本統計量，是否充分地代表 n 個數據，去反映母群的性質。例如$\overline{X}$較 Mo，Md 能充分地代表所有數據，去反映母群性質；同樣 S^2 比 AD(平均差)和 Q(四分差)更具充分性。
5. 抗拒性(Resistence):
   指的是樣本估計量是否能夠不受極端值影響，例如：中位數比平均數就不受極端值影響。

8-2 區間估計

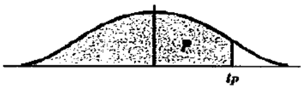
一、根據點估計值、抽樣分配及機率原理去推估母群參數在某一區間或範圍的統計方法，即稱為「區間估計」。

二、原理：

母數=點估計值±抽樣誤差

=點估計值±(臨界值×點估計值的標準誤)

4. 以下為 *t* test 的部份臨界值（critical values）表格。請據此算出 *F* test 在分子自由度為 1 及分母自由度為 4，位於 .05 顯著水準（significance level）的臨界值（四捨五入至小數點後第三位）。【10 分】



df	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	.325	.727	1.376	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032

命中講義 P.79

3. 每一種自由度就有一種 t 分配曲線，如圖 9-3。
4. 曲線下的總面積為 1，故為機率分配，其機率值可查 t 分配表。
5. $t^2(df)=F_{(1,df)}$，即$t_{1-\frac{\alpha}{2}(df)} = F_{1-\alpha(1,df)}$

證明：

$$t^2_{(df)} = \frac{Z^2}{\chi^2/df} = \frac{\chi^2/1}{\chi^2/df}(df=1\text{時，}\chi^2 = Z^2) = F_{(1,df)}$$

h. 卡方(χ^2)分配

自常態母群中，每次隨機抽取 n 個變量$X_1, X_2, X_3, \ldots, X_n$，並將其轉換為$Z_i$，再加以平方，即$Z_i^2 = (\frac{X_i - \mu}{\sigma})^2$，然後將這 n 個$Z_i^2$加起來，求其總和$\sum Z_i^2$，則命名$\sum Z_i^2$為卡方($\chi^2$，讀做 chi square)。如此進行抽樣及轉換無線多次，即可得到無限多個χ^2所形成的分配，稱為自由度 df 的χ^2分配，以χ^2_{df}表示。

※χ^2分配的特性

1. 因為χ^2為 Z 的平方和，所以 $0<\chi^2<\infty$，即χ^2均為正值。
2. 每一種 df，就有一條卡方分配曲線。
3. χ^2分配為正偏態分配，但當 df 增加，χ^2分配逐漸接近常態分配，可轉換成 Z 值，再查常態分配表。
4. 自由度為 df 的卡方分配，其平均數為 df，標準差為$\sqrt{2df}$
5. $\boxed{\text{曲線下的面積為 1}}$，故為機率分配，其機率值可查χ^2分配表。

5. **考慮 n 對觀察值（x_i , y_i）（i = 1, 2, ⋯, n）的簡單回歸（simple regression）。我們可將最小平方法（method of least squares）用於 regression of Y on X 以及 regression of X on Y，而分別得到其回歸線的斜率估計值。請問如何藉此二斜率估計值來推估 X 與 Y 的相關係數（correlation coefficient）？請詳述理由。【15 分】**

命中講義 P.183

十、相關比(曲線相關)(η/eta/):

當 X、Y 變項之間的相關，不是直線相關時，則積差相關便不適用。例如:焦慮和成績通常呈二次函數關係，則必需使用相關比。

$$\eta=\sqrt{\frac{SS_b}{SS_t}}$$

12-4 淨相關、部分相關及複相關

一、淨相關(partial correlation)(偏相關)：

兩個變項在它們與一個或多個變項的共同解釋力被排除之外，所剩下的純相關程度以 $r_{12.3}$ 表示。

例如:國語成績和數學成績的相關極高，但是國語、數學都會受到智力的影響，因此將智力自果與、數學成績中分別排除之後所剩下的純相關即是「淨相關」。

$$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

二、部分相關(part correlation)(半淨相關)：

指「第一變項」和「排除第三變項共同解釋力後的第二變項」之間的相關程度。

以 $r_{1(2.3)}$ 表示。

例如：國語成績和排除致力因素後的數學成績間的相關。

$$r_{1(2.3)}==\frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{23}^2}}$$

三、複相關(multiple correlation)(多元相關)：

指某依變項和數個自變項間的共變關係。

例如:數學成績(Y)可能和智商(X_1)練習次數，作答習慣(X_2)都有相關。

$$R=\sqrt{\beta_1 r_{Y1}+\beta_2 r_{Y2}}$$