

110年專門職業及技術人員高等考試建築師、
24類科技師（含第二次食品技師）、大地工程技師
考試分階段考試（第二階段考試）、公共衛生師
考試暨普通考試不動產經紀人、記帳士考試試題

等 別：高等考試
類 科：公共衛生師
科 目：生物統計學
考試時間：2 小時

座號：_____

※注意：可以使用電子計算器。

甲、申論題部分：（50 分）

- (一)不必抄題，作答時請將試題題號及答案依照順序寫在申論試卷上，於本試題上作答者，不予計分。
(二)請以藍、黑色鋼筆或原子筆在申論試卷上作答。
(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

- 一、某藥廠品管部門針對該藥廠每一批所生產的藥品查驗合格率的原理是：在 24 顆抽驗藥品中，只要不合格藥品數目是 0 或 1 顆（意即，不合格率 $\leq 4.17\%$ ），則該批藥品的查驗結果就算合格。假設在某批數千顆藥品中實際的不合格率為 4%，請計算該批藥品通過品管查驗被認定為合格的機率是多少？（註： $e^{(-0.96)} = 0.3829$ ，其中 $e = 2.7183$ ）（10 分）
- 二、某橫斷性研究探討經常（每週飲用 3 次以上）飲用市售含糖飲料與有無罹患第二型糖尿病之間的相關性。該研究的自變項除了是否經常飲用市售含糖飲料外，也包括年齡與吸菸兩個變項。下表呈現 6 個不同邏輯式迴歸模式（logistic regression model）的迴歸係數，例如，第（2）欄的邏輯式迴歸方程式即為： $\ln(p/q) = X_0 + \beta_1 X_1$ ， $\beta_1 = 0.5211$ ；而第（3）欄的方程式則是 $\ln(p/q) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ ， $\beta_1 = 1.3364$ ， $\beta_2 = 0.1418$ 。請計算經常飲用市售含糖飲料與第二型糖尿病之粗勝算比（crude odds ratio），並列舉理由說明本研究中年齡是否是一個干擾因子？另外，請計算控制年齡與吸菸後，經常飲用含糖飲料與罹患第二型糖尿病之調整後勝算比之 95% 信賴區間。（20 分）

變項	邏輯式迴歸模式			
	(1)	(2)	(3)	(4)
X ₀ intercept	-2.0075	-2.0591	-7.7640	-9.2834 (0.6360)
X ₁ 經常飲用市售含糖飲料		0.5211	1.3364	1.1883 (0.5784)
X ₂ 年齡（歲）			0.1418	0.1521 (0.0142)
X ₃ 吸菸（每日 1-24 支）				1.1246 (0.2182)
X ₄ 吸菸（每日 ≥ 25 支）				2.1371 (0.2211)

表格中的數字為迴歸係數，模式（4）同時呈現迴歸係數與迴歸係數標準誤（括弧內的數字）

三、某地區 1 年期間總共發生 144 例的溺水死亡事件，這些溺水死亡事件在各個月份分布的情況如下表所示：在 $\alpha = 0.05$ 的顯著水準下，請檢定該地區溺水死亡事件發生是否有月份聚集的現象？意即，溺水死亡事件是否在某些月份發生率較高？並請於檢定過程中寫出：(一)檢定的虛無假設與對立假設；(二)各月份溺水死亡事件發生的期望值，以及檢定統計量；(三)檢定統計量相對應的 p -value 範圍與臨界值、自由度；以及(四)檢定的決策與結論。(註： $p(\chi^2 > 17.28, df = 11) = 0.1$ ； $p(\chi^2 > 19.68, df = 11) = 0.05$ ； $p(\chi^2 > 21.92, df = 11) = 0.025$ ； $p(\chi^2 > 18.55, df = 12) = 0.1$ ； $p(\chi^2 > 21.03, df = 12) = 0.05$ ； $p(\chi^2 > 23.34, df = 12) = 0.025$) (20 分)

一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
1	3	2	7	14	20	37	33	16	6	2	3

乙、測驗題部分：(50 分)

代號：2401

(一)本測驗試題為單一選擇題，請選出一個正確或最適當的答案，複選作答者，該題不予計分。

(二)共40題，每題1.25分，須用2B鉛筆在試卡上依題號清楚劃記，於本試題或申論試卷上作答者，不予計分。

- 1 若有一種新型篩檢工具可以在 1 天內即測出是否有 D 病，該工具敏感度 (sensitivity) 為 85%，特異度 (specificity) 為 95%。張先生使用此工具篩檢得知結果為陽性，則張先生實際確診為 D 病的機率為何？
(A) 0.3% (B) 1.7% (C) 85% (D) 99%
- 2 新冠肺炎從 2020 年開始肆虐至今，已經至少出現 7 種變異病毒，愈多人沒打疫苗，病毒變異的機會愈大。某一機構之員工 30% 為外勤人員，70% 為內勤人員。已知 60% 的外勤人員和 40% 的內勤人員皆已打疫苗，若隨機抽取一沒打疫苗者，請問其為內勤人員之機率為何？
(A) 1/2 (B) 3/5 (C) 7/9 (D) 7/10
- 3 下列有關描述性統計的敘述，何者正確？
(A) 臺灣受僱勞工的平均月收入是一個左偏 (left skewed) 的分布，因為有一些高收入所得者
(B) 左偏 (left skewed) 的分布，中位數 (median) 通常會比平均數 (mean) 來得小
(C) 中位數比平均數較容易受到極端值 (outlier) 影響
(D) 莖葉圖 (stem and leaf plot) 可以用來描述資料的分布情況
- 4 卡方分布 (chi-square distribution) 的平均值與下列那一個數值相等？
(A) 變異數 (B) 中位數 (C) 自由度 (D) 眾數
- 5 為了解市面上販賣的飲料中某物質 X 的平均值，隨機抽取 40 杯飲料的樣本，這 40 杯飲料 X 的平均值為 10。假設 X 在母群體及此樣本的標準差剛好皆為 4。下列敘述何者錯誤？
(A) 標準誤為 0.632
(B) 標準誤為量化 X 分散程度的敘述性統計指標
(C) X 的平均值近似常態分布
(D) 如果只抽取 20 杯，X 的平均值分布未知
- 6 一群過重的成年人被隨機分派成兩組，一組進入減重計畫，一組作為對照組，半年之後用 t 檢定評估兩組受試者體重改變的差異，參加減重計畫的受試者體重平均減少 5 公斤，對照組平均減少 1 公斤，二組體重改變平均值差異的 95% 信賴區間是 1 公斤到 7 公斤。下列敘述何者正確？
(A) 雖然兩組體重減少的差異未達到統計顯著，但仍應拒絕虛無假設
(B) t 檢定的 P 值大於 0.05，所以不能拒絕虛無假設
(C) 因為是前後測量，所以應該用成對樣本的 t 檢定
(D) 從二組體重改變平均值差異的 95% 信賴區間，可以拒絕虛無假設

- 7 下列關於樣本數與統計檢定力 (statistical power) 的敘述, 何者錯誤?
 (A) 統計檢定力受到型一錯誤 (type-I error) 和型二錯誤 (type-II error) 大小的影響
 (B) 如果希望統計檢定力提高, 在其他條件不變的情況下, 所需樣本數就應隨著提高
 (C) 如果樣本數太小, 容易犯下型一錯誤
 (D) 減少測量誤差, 可以讓統計檢定力增加
- 8 為了估計臺灣 30 歲至 60 歲男性的平均血壓的收縮壓, 研究者挑選一家健身俱樂部 50 位 30 歲至 60 歲的男性會員測量其血壓。根據中央極限定理, 下列敘述何者正確?
 (A) 會參加健身俱樂部的人可能比較健康, 收縮壓平均數可以代表母群體平均數
 (B) 由這個樣本所估計的平均數, 其準確度 (accuracy) 會受到母群體平均數的影響
 (C) 由這個樣本所估計的平均數, 其準確度 (accuracy) 會受到母群體變異數的影響
 (D) 由這個樣本所估計的平均數, 其準確度 (accuracy) 會受到母群體總人數的影響
- 9 關於平均數之信賴區間的敘述, 下列何者正確?
 (A) 母群體的平均數會落在 95% 信賴區間的機率是 0.95
 (B) 95% 信賴區間會比 99% 信賴區間來得更寬
 (C) 從隨機樣本計算平均數的 95% 信賴區間, 當樣本數增加, 95% 信賴區間也會隨著變寬
 (D) 如果從母群體隨機抽樣相同的大小樣本 100 次, 從這些樣本所計算的 95% 信賴區間, 大約會有 95 個包含母群體的平均數
- 10 根據調查, 肺癌患者治療後復發的機率為 15%。現有一醫生想了解這樣的機率是否正確, 該醫生希望 95% 的信賴水準下, 估計誤差不超過 1%, 最少應該收集多少樣本才足夠?
 (A) 1520 (B) 4899 (C) 9604 (D) 32654
- 11 在一臨床醫學研究中, 有 N 位病況相當之病人被隨機安排在 K 種治療方式, 若欲了解是否至少有一治療方式, 治療後其病人之平均病情分數與其他治療方式有統計上顯著差異, 研究者得到下列變異數分析表:

變異來源	平方和	自由度	平均平方和	F 值
治療方式	1090	2	(丙)	(戊)
誤差	(甲)	(乙)	(丁)	
總和	1477	23		

請問根據上表, (戊) 大約為何?

- (A) 5 (B) 15 (C) 25 (D) 30
- 12 當抽樣的樣本數增加時, 樣本平均值的標準誤 (standard error) 會發生何種變化?
 (A) 變成零 (B) 不變 (C) 變大 (D) 變小
- 13 某調查想了解民眾對某政策是否贊同, 100 位受訪者中, 有 55 位回答「贊同」, 45 位回答「不贊同」, 同意比例的點估計為 55%。有關估算同意比例的 95% 信賴區間 (95% Confidence interval; 95% CI) 的方法, 下列敘述何者錯誤?
 (A) 計算 95% CI 會用到 Z 分布的第 95% 百分位數 (the 95th percentile)
 (B) 同意比例的標準誤約為 0.05
 (C) 若同意比例的點估計仍為 55%, 但總樣本數改為 80 位, 95% CI 會變寬
 (D) 同樣的樣本若估算 99% CI 會比 95% CI 寬

- 14 某抽樣調查甲地區高中生嘗試吸菸狀況，調查結果樣本中 30% 曾吸菸，估計嘗試吸菸率的 95% 信賴區間（下界，上界）得 (27%, 33%)。已知全國高中吸菸率為 25%，試檢定甲地區高中生嘗試吸菸率是否與全國高中生有顯著差異（顯著水準設 0.05）。下列敘述何者正確？
 (A) $p \text{ value} > 0.5$ (B) $0.25 < p \text{ value} < 0.5$
 (C) $0.05 < p \text{ value} < 0.25$ (D) $p \text{ value} < 0.05$
- 15 承上題，若顯著水準（significance level）設定為 0.05，下列敘述何者正確？
 (A) 甲地區高中生嘗試吸菸率與全國高中生有差異，具統計上顯著意義
 (B) 甲地區高中生嘗試吸菸率與全國高中生沒有差異，具統計上顯著意義
 (C) 甲地區高中生嘗試吸菸率與全國高中生有差異，沒有統計上顯著意義
 (D) 甲地區高中生嘗試吸菸率與全國高中生沒有差異，沒有統計上顯著意義
- 16 某機構針對甲乙兩個部門進行服務滿意度調查，收集甲乙各 100 位民眾（總共 200 位民眾）分別對服務人員勾選的滿意度程度，非常不滿意勾 1、不滿意勾 2、普通勾 3、滿意勾 4、非常滿意勾 5。若要檢定甲乙兩部門滿意程度中位數是否有統計上顯著差異，下列何種方法最適當？
 (A) Spearman correlation coefficient (B) Wilcoxon rank sum test
 (C) McNemar chi-square test (D) Sign test
- 17 進行全國性大規模健康行為與狀況調查時，完成隨機抽樣後，若要檢定樣本與全國母體在年齡及性別上的分布是否一致時，下列統計方法何者最適當？
 (A) Chi-square goodness-of-fit test (B) Logistic regression
 (C) McNemar test (D) Linear regression
- 18 下列為 10 對雙胞胎成年後身體質量指數（BMI）數據：

配對號	1	2	3	4	5	6	7	8	9	10
雙胞胎A	20	23	30	22	19	28	29	30	25	33
雙胞胎B	21	21	28	24	25	31	21	33	28	34

若要量化雙胞胎 BMI 的相關程度，下列何種方法最適當？

- (A) 勝算比（Odds Ratio） (B) 變異係數（Coefficient of variation）
 (C) 斯皮爾曼（Spearman）等級相關係數 (D) Kappa 係數
- 19 下表為 A、B 兩個地區發生疾病 D 的分布：

	疾病D		
	是	否	
A地區	8	1992	2000
B地區	16	984	1000

若要估算及檢定地區與疾病 D 的關聯，下列那一種統計方法最不適當？

- (A) Two-sample t test (B) Fisher exact test
 (C) Pearson's chi-square test (D) Logistic regression
- 20 某社區健康服務中心辦理健康促進活動，100 位民眾報名參加，其中一項活動為調查參加者對愛滋病的傳染途徑是否了解。每個人在活動開始及結束各回答同一個問題一次，1 表示答對，0 表示答錯，以下為民眾的資料示意表（僅列出前 10 位）。若要檢定參加前後民眾對愛滋病傳染途徑了解是否有顯著差異，下列那一個統計方法最適當？

ID	1	2	3	4	5	6	7	8	9	10
前	1	1	1	0	0	0	0	0	0	1
後	0	1	1	0	1	0	1	1	1	0

- (A) Paired t test (B) Two-sample t test
 (C) Linear regression (D) McNemar chi-square test

- 21 以下是一個統計車禍發生時，機車騎士是否有戴安全帽和是否發生頭部外傷之間關係的二乘二列聯表：

	沒有戴安全帽	有戴安全帽	總人數
沒有頭部外傷	10	70	80
有頭部外傷	20	40	60
總人數	30	110	140

下列敘述何者正確？

- (A) 如果使用卡方檢定 (Chi-squared test) 檢查戴安全帽和頭部外傷關係，則卡方檢定的自由度為 4
- (B) 如果戴安全帽與發生頭部外傷是兩獨立事件，沒有戴安全帽也沒有發生頭部外傷的預期人數為 16.1 人
- (C) 如果戴安全帽與發生頭部外傷是兩獨立事件，有戴安全帽仍發生頭部外傷的預期人數為 47.1 人
- (D) 卡方分布是間斷分布 (discrete distribution)，可以用來檢定列聯表中兩類別變數的關係
- 22 下列關於單因子變異數分析 (one-way analysis of variance) 的敘述，何者錯誤？
- (A) 單因子變異數分析用來比較三組的母群體平均數 μ_1 ， μ_2 和 μ_3 是否相等時，虛無假設是 $\mu_1 = \mu_2 = \mu_3$
- (B) 對立假設為母群體平均數不相等，亦即 $\mu_1 \neq \mu_2$ 且 $\mu_1 \neq \mu_3$ 且 $\mu_2 \neq \mu_3$
- (C) 變異數分析的假設，其中之一為母群體變異數相等
- (D) 如果用變異數分析比較兩組平均數相等時，其結果會等於兩獨立樣本 t 檢定的結果
- 23 下列關於兩獨立樣本 t 檢定 (two-sample independent t test) 的敘述，何者正確？
- (A) 兩組獨立樣本的大小必須相同 (B) 可以用來檢定左右眼近視度數的差別
- (C) 兩組獨立樣本的變異數大小必須接近 (D) 兩組樣本的平均值必須要相近
- 24 臨床心理師將 18 位病情相當的輕度失智症患者完全隨機分派到下表三種治療方式，每一種治療方式有 6 人，持續進行 3 個月的治療，之後接受憂鬱量表的測量。治療後的憂鬱量表測量分數描述統計量如下，分數越高表示憂鬱傾向越高。臨床心理師執行變異數分析，請問此變異數分析表中之組間變異平方和 (Sum of Squares Between Groups) 大約為何？

	個數	平均數	標準差	標準誤
有氧運動	6	8.00	1.414	.577
音樂療法	6	12.00	3.347	1.366
認知療法	6	25.00	2.828	1.155
總和	18	15.00	7.874	1.856

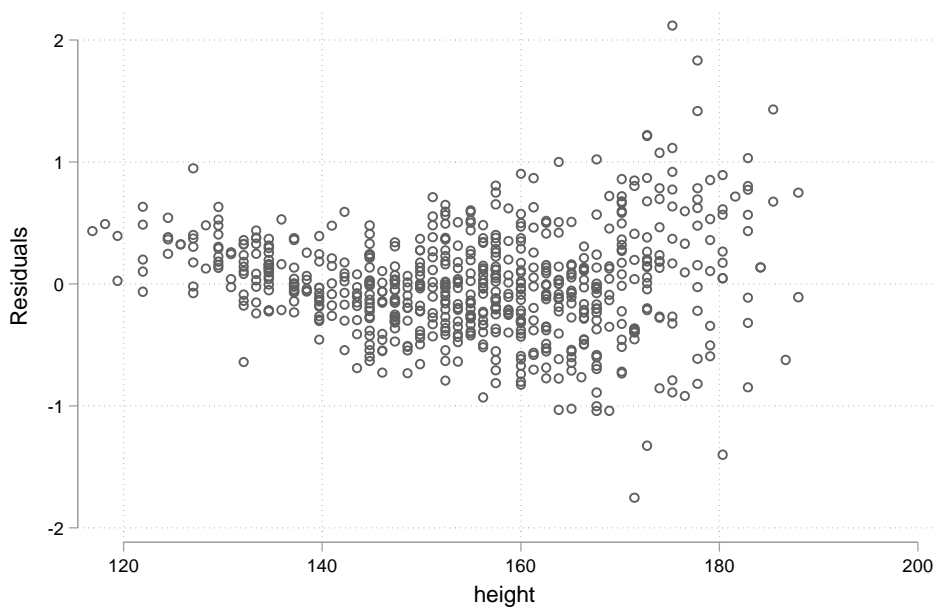
- (A) 1422 (B) 948 (C) 474 (D) 106
- 25 下列何者不是無母數方法？
- (A) Kruskal-Wallis test (B) Mann-Whitney U test
- (C) Paired t test (D) Sign test
- 26 下列關於解釋變數之間的共線性 (collinearity) 何者錯誤？
- (A) 共線性是指解釋變數之間的相關係數很高
- (B) 共線性只有當解釋變數的數目超過 10 個的時候才是個問題
- (C) 當一個解釋變數的迴歸係數方向改變，例如從正向變成負向關係，代表共線性可能是個問題
- (D) 當模型的判定係數 (coefficient of determination) 很大，但是模型裡面的解釋變數都不顯著，代表共線性可能是個問題

27 189 位足部雞眼 (foot corns) 的患者，94 位接受手術治療，另外 95 位接受非手術性治療。下列是進行邏輯式迴歸分析治療成功與否 (成功設為 1, 失敗設為 0) 的結果：

變項	迴歸係數	標準誤	勝算比	勝算比信賴區間	
				下界	上界
治療組別 (非手術=0, 手術=1)	0.278	0.485	1.32	0.51	3.414
性別 (男性=0, 女性=1)	-0.697	0.509	0.498	0.184	1.352
截距	-0.934	0.356			

下列敘述何者正確？

- (A) 治療組別的迴歸係數虛無假設勝算比為 1
 - (B) 不管使用那種方式治療，男性和女性治療成功機會的差異達到統計上顯著
 - (C) 治療組別之間治療成功機會的差異達到統計上顯著
 - (D) 在調整性別之後，手術治療的成功機率 (probability) 是非手術治療的 1.32 倍，但是沒有達到統計上的顯著
- 28 下列關於迴歸分析變數選擇的敘述，何者正確？
- (A) 使用逐步迴歸 (stepwise regression) 總是能夠得到最正確的模型
 - (B) 統計不顯著的解釋變數，必須要從模型裡移除掉，否則會影響其他變數的解釋
 - (C) 當解釋變數很多的時候，可以完全交給逐步迴歸來決定那些變數應該納入模型中
 - (D) 當解釋變數很多的時候，經由逐步迴歸所挑選出來的變數，其 P 值可能比實際上小
- 29 隨機抽樣兒童醫院 654 位 18 歲以下兒童和青少年，測量他們的身高 (單位：公分) 及每分鐘最大通氣量 (單位：公升)。以最大通氣量為依變項，身高作為解釋變項，進行簡單線性迴歸分析。結果發現，身高的迴歸係數為 0.05 公升/公分，標準誤為 0.001，判定係數為 0.75。下圖為模型的殘差圖 (residual plot)，橫座標是身高，縱座標是殘差。



下列敘述何者錯誤？

- (A) 身高和最大通氣量的相關係數為 0.375
- (B) 殘差圖顯示變異數同質性 (homoscedasticity) 的假設不成立
- (C) 身高的迴歸係數達到統計顯著
- (D) 殘差圖顯示身高和最大通氣量之間可能存在非線性的關係

- 34 某調查比較兩個地區某心血管風險指標分數（連續變項）的平均值是否不同，但兩個地區的年齡及性別組成不同，需要估算校正年齡及性別後之兩組平均值的差，下列那一個方法最適當？
 (A) Multiple linear regression (B) Multiple logistic regression
 (C) Two-sample t test (D) McNemar chi-square test
- 35 某研究想探討社區長者身體活動情況（Group）與某心血管類疾病（D 病）的關係。令 D 為二分類變項，D=1 表示有 D 病，D=0 表示無 D 病；Group=1 表示有規律運動習慣，Group=0 表示無運動習慣。以下為依變項為 D 的迴歸模式分析結果，B 代表迴歸係數，OR 代表勝算比（Odds Ratio），95%CI OR 代表勝算比的 95%信賴區間。

	B	OR	95%CI OR	P value
截距	1.39			<0.001*
Group	-0.73	0.48	(0.26, 0.89)	

上表最可能使用那一個統計方法？

- (A)斯皮爾曼（Spearman）相關係數 (B)皮爾森（Pearson）相關係數
 (C)邏輯式迴歸模式（Logistic regression） (D)線性迴歸模式（Linear regression）
- 36 61 位晚期大腸直腸癌病人，34 位接受合併化學治療，27 位接受單一化學治療，利用 Cox 風險比例模型，針對死亡風險分析得到以下結果：

	風險比 (Hazard Ratio)	標準誤	95%信賴區間下界	95%信賴區間上界
治療方式 (合併治療設為1，單一治療設為0)	0.40	0.16	0.19	0.87
性別（男性設為1，女性設為0）	1.08	0.42	0.50	2.31
年齡	1.00	0.02	0.97	1.04

根據上方表格，下列敘述何者正確？

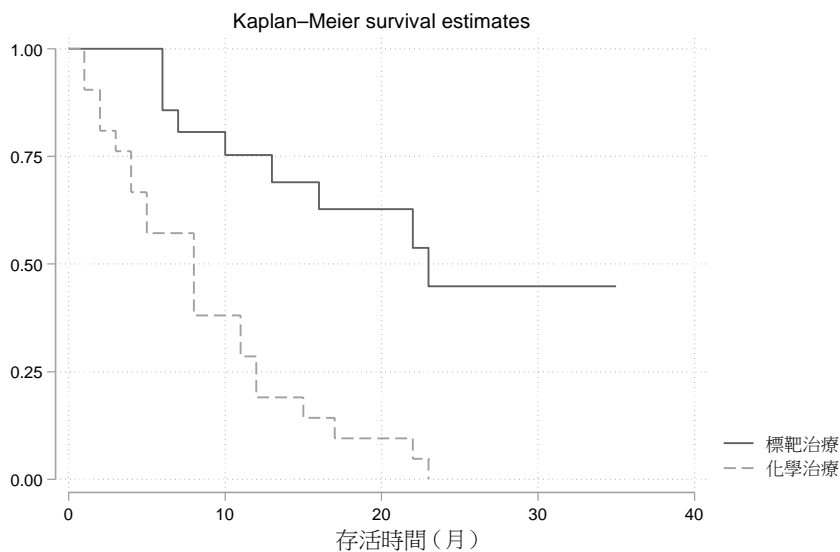
- (A)女性接受治療之後死亡的風險比男性小，但沒有達到統計上顯著
 (B)病人接受合併治療比接受單一治療的死亡風險較小，但沒有達到統計上顯著
 (C)病人的年齡對於治療之後死亡的風險有統計上顯著的影響
 (D)分析結果可顯示男性接受合併治療死亡的風險，比女性接受合併治療死亡的風險來得低
- 37 某研究調查大學生身體活動（PA）與 BMI 的關係。以下為線性迴歸模式的係數估計，BMI 為依變項（dependent variable），PA 為自變項（independent variable）。其中 PA 為身體活動的類別，1 代表經常身體活動（“活動組”），0 代表很少身體活動（“少動組”）。

Variable	DF	Coefficients	Standard error	t value	P value
Intercept	1	22.9	1.0	23.8	<.0001
PA	1	-0.3	1.1	-0.3	0.81

有關分析結果之判讀，下列敘述何者正確？（顯著水準=0.05）

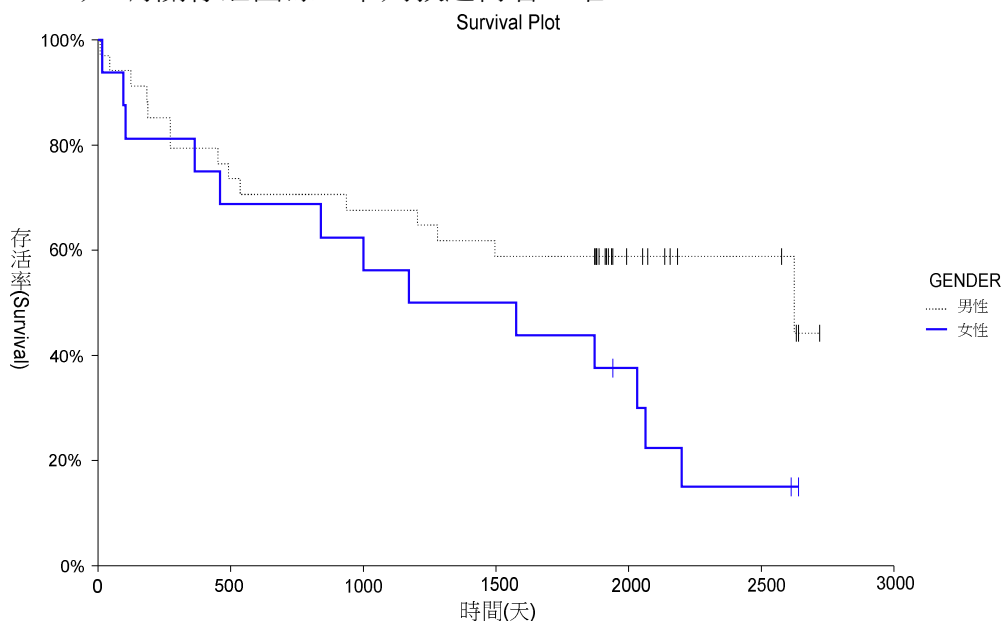
- (A)活動組平均 BMI 為 22.9
 (B)活動組 BMI 樣本平均值比少動組 BMI 樣本平均值略低
 (C)根據統計檢定，活動組與少動組 BMI 的平均值有統計上顯著差異
 (D) BMI 有 81%的變異可被 PA 所解釋

38 42 位急性白血病的病人隨機分派接受標靶治療或化學治療，觀察事件是病人死亡，下圖是 Kaplan-Meier 存活分析圖：



在本研究中，沒有病人中途退出，所有病人不是在觀察過程中死亡，就是在研究結束時仍存活。根據上圖，下列敘述何者正確？

- (A) 接受標靶治療的病人，Kaplan-Meier 存活分析所估計的中位數存活時間比實際觀察的中位數存活時間來得長
 - (B) 接受標靶治療的病人，所有病人在研究結束前都死亡
 - (C) 接受化學治療的病人，第一例死亡發生在 6 個月的時候
 - (D) 接受化學治療的病人，最長活了 30 個月
- 39 下圖為總共 50 位病人的存活曲線，虛線為男性，實線為女性，跨在每條線上的小細直線標示為設限 (censor)。有關存活曲線，下列敘述何者正確？



- (A) 若病人在 2000 天時設限，表示此病人在第 2000 天時死亡
 - (B) 女性存活時間的中位數約落在 1000 天到 1700 天之間
 - (C) 女性存活時間的中位數比男性長
 - (D) 男性組無法估算存活時間的中位數
- 40 病人的 Kaplan-Meier 存活分析圖形裡，存活函數是一個遞減的階梯函數 (step function)。如果觀察的事件是病人的死亡，關於階梯函數每往下一個階梯代表的意義，下列敘述何者正確？
- (A) 至少有一個病人失去追蹤
 - (B) 至少有一個病人主動退出試驗
 - (C) 至少有一個病人失去追蹤或死亡
 - (D) 至少有一個病人死亡