



XUNTA DE GALICIA
CONSELLERÍA DE FACENDA
Dirección Xeral da Función Pública



**Proceso selectivo para o ingreso no corpo superior da
Administración Xeral da Comunidade Autónoma de
Galicia, subgrupo A1, escala superior de estatísticos/as**

Segundo exercicio
Acceso: acceso libre

NON abra o exame ata que o tribunal llo indique

SUPOSTO 1 (I. ESTADÍSTICA TEÓRICA)

I. Na poboación dun concello de Galicia de 10.000 persoas existe unha proporción do 5% de parados.

- Se seleccionamos ao chou e sen substitución 20 persoas, determina a probabilidade de que cando menos, 1 sexa parada.
- Das 20 persoas seleccionadas no apartado anterior 4 son paradas, determina a probabilidade de que, se seleccionamos 5 persoas sen substitución de entre as 20, soamente 1 sexa parada.
- Se seleccionamos aleatoriamente 32 persoas da poboación, con substitución, cal é a probabilidade de que como máximo 2 sexan paradas?
- Se escollemos ás persoas da poboación con substitución, cal é probabilidade de que teñamos que escoller a máis de 4 antes de encontrar o terceiro non parado?.

II. Nun concello de Galicia realizouse un estudo para comparar o nivel de colesterol en persoas novas que practican deporte habitualmente co das que non practican deporte. Os resultados mostrais obtidos son os recollidos na seguinte táboa. Admitindo normalidade nos datos, responde aos apartados traballando cun nivel de confianza do 95%.

	Practican deporte (X)	Non practican deporte (Y)
Tamaño da mostra	n=11	m=16
Media mostral	$\bar{x} = 150$ mg/dl.	$\bar{y} = 183$ mg/dl.
Cuasidesviación mostral	$\hat{S}_x = 40$ mg/dl.	$\hat{S}_y = 20$ mg/dl.

- Calcula o intervalo de confianza para o cociente de variancias poboacionais σ_Y^2/σ_X^2 . Indica que hipóteses se necesitan para asegurar a validez do intervalo determinado.
- Determina o intervalo de confianza para a diferenza dos niveis medios de colesterol poboacionais $\mu_X - \mu_Y$. Indica que hipóteses se necesitan para asegurar a validez do intervalo determinado. De ser necesario, supón que os grados de liberdade para a aproximación de Welch son 13.
- Podemos afirmar que, en media, as persoas novas que practican deporte habitualmente teñen niveis máis baixos de colesterol? Presenta o contraste axeitado e conclúe.

III. Unha enquisa dirixida a consumidores estudiou os motivos (nivel de satisfacción nunha escala de 0 a 100 puntos) polos que se realiza a compra nun establecemento comercial. Os ítems que valoraron os clientes foron:

Existencia de marca propia	Calidade dos produtos ofertados
Disponibilidade de estacionamento	Limpeza e orde do local
Prezo dos artigos	Tempo de atención ao público nas caixas
Situación do establecemento	Variedade de marcas e artigos

Os datos medios da clientela das 130 cadeas comerciais participantes no estudo foron obxecto de diferentes análises. Entre eles realizouse unha análise de compoñentes principais con base na matriz de correlacións das 130 cadeas. Responde ás seguintes cuestións de maneira xustificada.

- Con base na seguinte táboa, indica que número mínimo de compoñentes con maior poder explicativo deben ser seleccionadas para garantir que se

explique cando menos un 70% da inercia total.

	Autovalores iniciais
1	4,445
2	1,505
3	0,695
4	0,509
5	0,429
6	0,232
7	0,137
8	0,048

- b) Coincide a selección anterior do número de compoñentes cunha selección baseada no criterio da media ou Kaiser?
- c) O estatístico KMO de Kaiser-Meyer-Olkin da adecuación da mostraxe presenta un resultado de 0,802. Indica a importancia que ten a medida KMO na realización dunha análise de compoñentes principais e conclúe a partir do dato.
- d) A táboa seguinte presenta os resultados da proba de esfericidade de Bartlett. Indica a hipótese nula e alternativa da proba de esfericidade de Bartlett. Determina os graos de liberdade da aproximación chi-cadrado para o caso a estudo. Conclúe en función do resultado recollido na táboa e sinala a importancia que ten nunha análise de compoñentes principais a realización desta proba.

Aprox. Chi-cadrado	810,853
Graos de liberdade	
Signif.	0,000

- e) Para facilitar a interpretación das 2 primeiras compoñentes realizouse unha rotación Varimax da solución inicial. A seguinte táboa mostra a matriz das cargas factoriais da solución rotada. Para a primeira variable, determine e analice a porcentaxe da variancia explicada polas compoñentes retidas na solución rotada.

Variables	Compoñentes	
	1	2
Tempo de atención ao público nas caixas	0,917	-0,225
Situación do establecemento	0,894	-0,162
Prezo dos artigos	0,904	-0,234
Disponibilidade de estacionamento	0,886	-0,130
Existencia de marca propia	0,773	0,167
Variedade de marcas e artigos	-0,281	0,750
Calidade dos produtos ofertados	0,347	0,742
Limpeza e orde do local	-0,423	0,664

- f) É factible obter a porcentaxe de variancia explicada pola segunda compoñente retida rotada a partir da táboa coa solución rotada anterior?
- g) Responde ás seguintes cuestións: debe coincidir a porcentaxe da variancia explicada de cada variable por cada compoñente da solución rotada presentada coa da solución inicial? E a porcentaxe total da variancia explicada de cada variable polas dúas compoñentes rotadas? E a porcentaxe total da inercia explicada polas dúas compoñentes rotadas?
- h) Interprete as dúas compoñentes retidas.
- i) Unha teoría sobre o comportamento e interese das persoas consumidoras indica a existencia de 2 factores comúns na tipoloxía das cadeas comerciais. Se esta metodoloxía de análise de compoñentes principais

tivera sido empregada como método de obtención dos 2 factores comúns na realización dunha análise factorial, como se denomina e interpreta a porcentaxe da variancia non explicada de cada variable polos 2 factores comúns retidos? Canto vale no caso da primeira variable cos datos dispoñibles?

IV. Un estudo do perfil socioeconómico da clientela permitiu clasificar 130 cadeas comerciais en 4 grupos en función da seus clientes maioritarios. Os 4 grupos de cadeas definidos en función do nivel económico da súa clientela foron: Baixo, Medio Baixo, Medio Alto e Alto. Realizouse unha análise discriminante obtida de forma simultánea cos datos de 8 variables das cadeas comerciais obtidos por mostraxe (as súas denominacións recóllense nalgunha das táboas posteriores).

- a) Cantas funcións discriminantes poden obterse, como máximo, neste contexto de traballo? Responde xustificadamente.
- b) A táboa seguinte inclúe resultados das funcións discriminantes da análise realizada. Analiza e interpreta xustificadamente a información correspondente á segunda fila.

Función	Autovalor	% de variancia	% variancia acumulada	Correlación canónica
1	5,086	94,7	94,7	0,914
2	0,224	4,2	98,9	0,428
3	0,060	1,1	100,0	0,237

- c) Na seguinte táboa preséntanse unhas probas sobre o modelo calculado. Analiza os resultados incluídos e indica o interese que teñen.

Lambda de Wilks				
Proba de funcións	Lambda de Wilks	Chi-cadrado	Graos de liberdade	Signif.
1 a 3	0,127	254,167	24	0,000
2 a 3	0,771	32,025	14	0,004
3	0,944	7,122	6	0,310

- d) A seguinte táboa amosa os coeficientes da función de clasificación lineal de Fisher. Que utilidade teñen e como se poden empregar? En que contexto de traballo da análise discriminante se poden empregar?

Coeficientes da función de clasificación lineal de Fisher				
	Baixo	Medio Baixo	Medio Alto	Alto
Tempo de atención ao público nas caixas	0,169	0,198	0,301	0,320
Situación do establecemento	0,067	0,037	0,112	0,236
Prezo dos artigos	-0,089	-0,034	0,016	0,171
Dispoñibilidade de estacionamento	-0,077	-0,099	-0,204	-0,173
Existencia de marca propia	0,105	0,124	0,082	-0,065
Variedade de marcas e artigos	0,081	0,092	0,267	0,306
Calidade dos produtos ofertados	-0,073	0,014	0,000	0,115
Limpeza e orde do local	0,290	0,210	0,190	0,150
(Constante)	-8,412	-7,610	-12,151	-24,999

- e) A seguinte táboa inclúe os resultados directos da clasificación das 130 cadeas comerciais utilizando como probabilidades a priori a probabilidade de pertenza a cada grupo representada pola porcentaxe de cadeas comerciais en cada grupo da mostra, baseándose nas funcións de clasificación lineal de Fisher e na regra de Bayes. Analiza a información e conclúe adecuadamente.

		Grupo prognosticado				
		Baixo	Medio Baixo	Medio Alto	Alto	
Grupo orixinal	Baixo	18	5	0	0	23
	Medio Baixo	6	17	7	0	30
	Medio Alto	1	7	26	2	36
	Alto	0	0	6	35	41
		25	29	39	37	130

- f) Indica, dende o contexto da aprendizaxe automática, se a análise discriminante pode ser considerada unha técnica útil dentro da aprendizaxe supervisada ou non supervisada.

SUPOSTO 2 (II. MOSTRAXE)

As persoas responsables do medio ambiente de tres concellos galegos (A, B e C) están interesadas en coñecer o consumo diario en electricidade –medido en kWh- efectuado nas vivendas dos seus concellos (X). As tres optan por efectuar a investigación mediante unha enquisa:

I. No primeiro concello (A) deciden seleccionar unha mostra aleatoria simple (sen reposición) de vivendas.

- Se o número total de vivendas (N) é 15.000, determine o tamaño de mostra necesario para estimar o consumo medio diario en electricidade por vivenda (\bar{X}) cun nivel de confianza do 95% e que non difira máis do 5% do verdadeiro valor. Enquisas previas indican que o coeficiente de variación ($CV = \sigma/\bar{X}$) toma o valor 1,0. (Admita normalidade).
- Baixo as mesmas condicións do punto a), cal sería o tamaño de mostra se a mostraxe se realizase con reposición?. Relacione teoricamente ambos resultados.
- Seleccionada unha mostra aleatoria simple (sen reposición) de tamaño $n=1.000$ vivendas, estime o consumo medio diario en electricidade por vivenda (\bar{X}) e estime o seu erro de mostraxe cos seguintes datos: $\sum_{i=1}^{1000} x_i = 7.828$; $\sum_{i=1}^{1000} x_i^2 = 400.000$

II. No concello B dispoñen dun rexistro de todas as vivendas sitas no concello coa súa superficie útil. Deciden empregar unha mostraxe estratificada segundo os metros cadrados útiles das vivendas, con mostraxe aleatoria simple (sen reposición) dentro de cada estrato. Estableceron tres estratos cos seguintes datos:

Estrato (h)	Número de vivendas (N_h)	Cuasidesviación poboacional ($S_{h,x}$)
1	8.000	5,0
2	5.000	6,0
3	1.000	12,0
Total	14.000	

- Determine o tamaño de mostra mínimo para estimar o consumo medio diario en electricidade por vivenda cunha variancia de 0,1 ($V(\hat{X})=0,1$) -asignación óptima-. Cal é a súa repartición por estratos?.
- Seleccionouse unha mostra estratificada de tamaño $n=350$ vivendas, con mostraxe aleatoria simple (sen reposición) dentro de cada estrato. Os resultados mostrais son os recollidos na seguinte táboa,

Estrato (h)	Tamaño da mostra (n_h)	Media mostral de X (\bar{x}_h)
1	180	7,0
2	120	9,0
3	50	12,0
Total	350	

Calcule:

- Unha estimación inesgada do consumo medio diario en electricidade por vivenda (\bar{X}) e o seu erro de mostraxe.
- Unha estimación inesgada do consumo diario total en electricidade no concello (X) e o seu erro de mostraxe.

As seguintes sumas poden ser de axuda:

$$\sum_{h=1}^3 N_h S_{h,x} = 82.000; \sum_{h=1}^3 N_h S_{h,x}^2 = 524.000; \sum_{h=1}^3 N_h^2 S_{h,x}^2 / n_h = 19.268.889$$

III. No concello C agrupan as 12.000 vivendas en 20 distritos e seleccionan 5 (distritos) con probabilidades iguais e sen reposición. Posteriormente, en cada un deles, seleccionan o 20% das súas vivendas con probabilidades iguais e sen reposición,

- a) Calcule unha estimación inesgada do consumo medio diario en electricidade por vivenda (\bar{X}) e unha estimación inesgada do consumo diario total en electricidade (X) no concello C a partir dos datos da seguinte mostra:

Conglomerado (i)	Tamaño conglomerado (M_i)	Total mostral de X no conglomerado (x_i)
1	600	720
2	300	492
3	800	1.120
4	500	810
5	600	948
Suma	2.800	4.090

- b) Indique desde un punto de vista teórico unha expresión para estimar a variancia do consumo diario total de electricidade no concello C aplicando o Theorema II de Durbin.

SUPOSTO 3 (III. DEMOGRAFÍA E ESTATÍSTICA APLICADA AO SECTOR PÚBLICO)

I. Dados os seguintes datos de nacementos e de poboación residente no ano 2020 en Ourense:

Grupos de idade	Poboación a 1 de xullo de 2020			Nacementos por idade da nai en 2020		
	Total	Homes	Mulleres	Total	Homes	Mulleres
0-14	29.653	15.309	14.344			
15-19	10.908	5.624	5.283	25	12	13
20-24	11.441	5.898	5.542	86	40	46
25-29	13.094	6.698	6.396	248	135	113
30-34	14.704	7.406	7.298	446	237	209
35-39	17.945	9.077	8.868	445	240	205
40-44	21.998	10.969	11.029	183	107	76
45-49	22.101	11.044	11.057	23	14	9
50 e máis	163.852	75.254	88.598			
Total	305.696	147.279	158.415	1.456	785	671

Calcular para o ano 2020:

- A taxa bruta de natalidade
- A taxa xeral ou global de fecundidade
- O ratio (ou relación) de masculinidade ao nacemento
- Sabemos que o índice sintético de fecundidade (ou indicador conxuntural de fecundidade) toma o valor 0,9. Se outra provincia ten unha taxa bruta de natalidade superior, terá un índice sintético de fecundidade maior?. Razoe a resposta.

II. Coñecendo a seguinte información demográfica para un territorio:

Poboación residente a 1 de xaneiro de 2019	1.119.596
Taxa de crecemento anual acumulativa (ou acumulada) en 2019 (tanto por cen)	0,23
Taxa bruta de mortalidade en 2019 (tanto por mil)	11,0
Taxa de migración neta en 2019 (tanto por mil)	7,5

Calcule:

- A poboación residente no territorio a 1 de xaneiro de 2020.
- O número de defuncións e de nacementos, o saldo vexetativo e o saldo migratorio no territorio durante o ano 2019.

Empregando a taxa de crecemento anual da poboación que figura na táboa anterior, estime a poboación a 1 de xaneiro de 2025 supoñendo:

- Un crecemento xeométrico
- Un crecemento loxístico cun limiar máximo de 1.200.000 de habitantes, e que acadou un tamaño de 600.000 de habitantes en 2000.

III. Dado o seguinte extracto dunha táboa de mortalidade:

	m_x	a_x	q_x	l_x	d_x	L_x	T_x	e_x
Idade	Taxa de mortalidade (‰)	Promedio de anos vividos no último ano de vida	Risco de morte (‰)	Supervivientes	Defuncións teóricas	Poboación estacionaria	Tempo vivido	Esperanza de vida
0	0,00228	0,1300	0,00228	100.000	228	99.801	8.283.851	82,84
1	0,00012	0,4800						
2	0,00012	0,4700	0,00012	99.760	12	99.754	8.084.284	81,04
(...)								

a) Calcule as células baleiras q_1 , l_1 , d_1 , L_1 , T_1 e e_1 . Interpretar os resultados.

Para analizar a morbilidade dunha enfermidade:

b) Que expresan a taxa de prevalencia e a taxa de incidencia da enfermidade?.

SUPUESTO 1 (I. ESTADÍSTICA TEÓRICA)

I. En la población de un ayuntamiento de Galicia de 10.000 personas existe una proporción del 5% de parados.

- Si seleccionamos al azar y sin reemplazamiento 20 personas, determina la probabilidad de que por lo menos 1 sea parada.
- De las 20 personas seleccionadas en el apartado anterior 4 son paradas, determina la probabilidad de que, si seleccionamos 5 personas sin reemplazamiento de entre las 20, solamente 1 sea parada.
- Si seleccionamos aleatoriamente 32 personas de la población, con reemplazamiento, ¿cuál es la probabilidad de que como máximo 2 sean paradas?
- Si elegimos las personas de la población con reemplazamiento, ¿cuál es la probabilidad de que tengamos que escoger a más de 4 antes de encontrar el tercer no parado?

II. En un ayuntamiento de Galicia se realizó un estudio para comparar el nivel de colesterol en personas jóvenes que practican deporte habitualmente con el de las que no practican deporte. Los resultados muestrales obtenidos son los recogidos en la siguiente tabla. Admitiendo normalidad en los datos, responde a los apartados trabajando con un nivel de confianza del 95%.

	Practican deporte (X)	No practican deporte (Y)
Tamaño de la muestra	n=11	m=16
Media muestral	$\bar{x} = 150$ mg/dl.	$\bar{y} = 183$ mg/dl.
Cuasidesviación muestral	$\hat{S}_x = 40$ mg/dl.	$\hat{S}_y = 20$ mg/dl.

- Calcula el intervalo de confianza para el cociente de varianzas poblacionales σ_Y^2/σ_X^2 . Indica qué hipótesis se necesitan para asegurar la validez del intervalo determinado.
- Determina el intervalo de confianza para la diferencia de los niveles medios de colesterol poblacionales $\mu_X - \mu_Y$. Indica que hipótesis se necesitan para asegurar la validez del intervalo determinado. De ser necesario, supón que los grados de libertad para la aproximación de Welch son 13.
- ¿Podemos afirmar que, en promedio, las personas jóvenes que practican deporte habitualmente tienen niveles más bajos de colesterol? Plantea el contraste adecuado y concluye.

III. Una encuesta dirigida a consumidores estudió los motivos (nivel de satisfacción en una escala de 0 a 100 puntos) por los que se realiza la compra en un establecimiento comercial. Los ítems que valoraron los clientes fueron:

Existencia de marca propia	Calidad de los productos ofertados
Disponibilidad de estacionamiento	Limpieza y orden del local
Precio de artículos	Tiempo de atención al público en cajas
Ubicación del establecimiento	Variedad de marcas y artículos

Los datos medios de la clientela de las 130 cadenas comerciales participantes en el estudio fueron objeto de diferentes análisis. Entre ellos se realizó un análisis de componentes principales a partir de la matriz de correlaciones de las 130 cadenas. Responde a las siguientes cuestiones de manera justificada.

- En base a la siguiente tabla, indica qué número mínimo de componentes con mayor poder explicativo deben ser seleccionadas para garantizar que

se explique al menos un 70% de la inercia total.

	Autovalores iniciales
1	4,445
2	1,505
3	0,695
4	0,509
5	0,429
6	0,232
7	0,137
8	0,048

- b) ¿Coincide la selección anterior del número de componentes con una selección basada en el criterio de la media o Kaiser?
- c) El estadístico KMO de Kaiser-Meyer-Olkin de adecuación de muestreo presenta un resultado de 0,802. Indica la importancia que tiene la medida KMO en la realización de un análisis de componentes principales y concluye a partir del dato.
- d) La tabla siguiente presenta los resultados de la prueba de esfericidad de Bartlett. Indica la hipótesis nula y alternativa de la prueba de esfericidad de Bartlett. Determina los grados de libertad de la aproximación chi-cuadrado para el caso a estudio. Concluye en función del resultado recogido en la tabla y señala la importancia que tiene en un análisis de componentes principales la realización de esta prueba.

Aprox. Chi-cuadrado	810,853
Grados de libertad	
Signif.	0,000

- e) Para facilitar la interpretación de las 2 primeras componentes se realizó una rotación Varimax de la solución inicial. La siguiente tabla muestra la matriz de cargas factoriales de la solución rotada. Para la primera variable, determine y analice el porcentaje de varianza explicada por las componentes retenidas en la solución rotada.

Variables	Componentes	
	1	2
Tiempo de atención al público en cajas	0,917	-0,225
Ubicación del establecimiento	0,894	-0,162
Precio de artículos	0,904	-0,234
Disponibilidad de estacionamiento	0,886	-0,130
Existencia de marca propia	0,773	0,167
Variedad de marcas y artículos	-0,281	0,750
Calidad de los productos ofertados	0,347	0,742
Limpieza y orden del local	-0,423	0,664

- f) ¿Es factible obtener el porcentaje de varianza explicada por la segunda componente retenida rotada a partir de la tabla con la solución rotada anterior?
- g) Responde a las siguientes cuestiones: ¿debe coincidir el porcentaje de varianza explicado de cada variable por cada componente de la solución rotada presentada con el de la solución inicial? ¿y el porcentaje total de varianza explicado de cada variable por las dos componentes rotadas? ¿y el porcentaje total de la inercia explicado por las dos componentes rotadas?
- h) Interprete las dos componentes retenidas.
- i) Una teoría sobre el comportamiento e interés de las personas

consumidoras indica la existencia de 2 factores comunes en la tipología de las cadenas comerciales. Si esta metodología de análisis de componentes principales hubiera sido empleada como método de obtención de los 2 factores comunes en la realización de un análisis factorial, ¿cómo se denomina e interpreta el porcentaje de varianza no explicada de cada variable por los 2 factores comunes retenidos? ¿cuánto vale en el caso de la primera variable con los datos disponibles?

IV. Un estudio del perfil socioeconómico de la clientela permitió clasificar 130 cadenas comerciales en 4 grupos en función de sus clientes mayoritarios. Los 4 grupos de cadenas definidos en función del nivel económico de su clientela fueron: Bajo, Medio Bajo, Medio Alto y Alto. Se realizó un análisis discriminante obtenido de forma simultánea con los datos de 8 variables de las cadenas comerciales obtenidos por muestreo (sus denominaciones se recogen en alguna de las tablas posteriores).

- a) ¿Cuántas funciones discriminantes pueden obtenerse, como máximo, en este contexto de trabajo? Responde justificadamente.
- b) La tabla siguiente incluye resultados de las funciones discriminantes del análisis realizado. Analiza e interpreta justificadamente la información correspondiente a la segunda fila.

Función	Autovalor	% de varianza	% varianza acumulada	Correlación canónica
1	5,086	94,7	94,7	0,914
2	0,224	4,2	98,9	0,428
3	0,060	1,1	100,0	0,237

- c) En la siguiente tabla se presentan unas pruebas sobre el modelo calculado. Analiza los resultados incluidos e indica el interés que tienen.

Lambda de Wilks				
Prueba de funciones	Lambda de Wilks	Chi-cuadrado	Grados de libertad	Signif.
1 a 3	0,127	254,167	24	0,000
2 a 3	0,771	32,025	14	0,004
3	0,944	7,122	6	0,310

- d) La siguiente tabla presenta los coeficientes de la función de clasificación lineal de Fisher. ¿Qué utilidad tienen y cómo se pueden emplear? ¿En qué contexto de trabajo del análisis discriminante se pueden emplear?

Coeficientes de la función de clasificación lineal de Fisher				
	Bajo	Medio Bajo	Medio Alto	Alto
Tiempo de atención al público en cajas	0,169	0,198	0,301	0,320
Ubicación del establecimiento	0,067	0,037	0,112	0,236
Precio de artículos	-0,089	-0,034	0,016	0,171
Disponibilidad de estacionamiento	-0,077	-0,099	-0,204	-0,173
Existencia de marca propia	0,105	0,124	0,082	-0,065
Variedad de marcas y artículos	0,081	0,092	0,267	0,306
Calidad de los productos ofertados	-0,073	0,014	0,000	0,115
Limpieza y orden del local	0,290	0,210	0,190	0,150
(Constante)	-8,412	-7,610	-12,151	-24,999

- e) La siguiente tabla incluye los resultados directos de la clasificación de las 130 cadenas comerciales utilizando como probabilidades a priori la probabilidad de pertenencia a cada grupo representada por el porcentaje de cadenas comerciales en cada grupo de la muestra, basándose en las

funciones de clasificación lineal de Fisher y la regla de Bayes. Analiza la información y concluye adecuadamente.

		Grupo pronosticado				
		Bajo	Medio Bajo	Medio Alto	Alto	
Grupo original	Bajo	18	5	0	0	23
	Medio Bajo	6	17	7	0	30
	Medio Alto	1	7	26	2	36
	Alto	0	0	6	35	41
		25	29	39	37	130

- f) Indica, desde el contexto del aprendizaje automático, si el análisis discriminante puede ser considerado una técnica útil dentro del aprendizaje supervisado o no supervisado.

SUPUESTO 2 (I. MUESTREO)

Las personas responsables de medio ambiente de tres municipios gallegos (A, B y C) están interesadas en conocer el consumo diario en electricidad –medido en kWh- efectuado en las viviendas de su municipio (X). Las tres optan por efectuar la investigación mediante una encuesta:

I. En el primer municipio (A) deciden seleccionar una muestra aleatoria simple (sin reposición) de viviendas.

- Si el número total de viviendas (N) es 15.000, determine el tamaño de muestra necesario para estimar el consumo medio diario en electricidad por vivienda (\bar{X}) con un nivel de confianza del 95% y que no difiera más del 5% del verdadero valor. Encuestas previas indican que el coeficiente de variación ($CV = \sigma/\bar{X}$) toma el valor 1,0. (Admita normalidad).
- Bajo las mismas condiciones del punto a), ¿cuál sería el tamaño de muestra si el muestreo se realizase con reposición? Relacione teóricamente ambos resultados.
- Seleccionada una muestra aleatoria simple (sin reposición) de tamaño $n=1.000$ viviendas, estime el consumo medio diario en electricidad por vivienda (\bar{X}) y estime su error de muestreo con los siguientes datos: $\sum_{i=1}^{1000} x_i = 7.828$; $\sum_{i=1}^{1000} x_i^2 = 400.000$

II. En el municipio B disponen de un registro de todas las viviendas situadas en el municipio con su superficie útil. Deciden emplear un muestreo estratificado según los metros cuadrados útiles de las viviendas, con muestreo aleatorio simple (sin reposición) dentro de cada estrato. Establecieron tres estratos con los siguientes datos:

Estrato (h)	Número de viviendas (N_h)	Cuasi desviación poblacional ($S_{h,x}$)
1	8.000	5,0
2	5.000	6,0
3	1.000	12,0
Total	14.000	

- Determine el tamaño de muestra mínimo para estimar el consumo medio diario en electricidad por vivienda con una varianza de 0,1 ($V(\hat{X})=0,1$) -asignación óptima-. ¿Cuál es su repartición por estratos?
- Se seleccionó una muestra estratificada de tamaño $n=350$ viviendas, con muestreo aleatorio simple (sin reposición) dentro de cada estrato. Los resultados muestrales son los recogidos en la siguiente tabla,

Estrato (h)	Tamaño de la muestra (n_h)	Media muestral de X (\bar{x}_h)
1	180	7,0
2	120	9,0
3	50	12,0
Total	350	

Calcule:

- Una estimación insesgada del consumo medio diario en electricidad por vivienda (\bar{X}) y su error de muestreo.
- Una estimación insesgada del consumo diario total en electricidad en el municipio (X) y su error de muestreo.

Las siguientes sumas pueden ser de ayuda:

$$\sum_{h=1}^3 N_h S_{h,x} = 82.000; \sum_{h=1}^3 N_h S_{h,x}^2 = 524.000; \sum_{h=1}^3 N_h^2 S_{h,x}^2 / n_h = 19.268.889$$

III. En el municipio C agrupan las 12.000 viviendas en 20 distritos y seleccionan 5 (distritos) con probabilidades iguales y sin reposición. Posteriormente, en cada uno de ellos, seleccionan el 20% de sus viviendas con probabilidades iguales y sin reposición,

- a) Calcule una estimación insesgada del consumo medio diario en electricidad por vivienda (\bar{X}) y una estimación insesgada del consumo diario total en electricidad (X) en el municipio C a partir de los datos de la siguiente muestra:

Conglomerado (i)	Tamaño conglomerado (M_i)	Total muestral de X en el conglomerado (x_i)
1	600	720
2	300	492
3	800	1.120
4	500	810
5	600	948
Suma	2.800	4.090

- b) Indique desde un punto de vista teórico una expresión para estimar la varianza del consumo diario total de electricidad en el municipio C aplicando el Theorema II de Durbin.

SUPUESTO 3 (III. DEMOGRAFÍA Y ESTADÍSTICA APLICADA AL SECTOR PÚBLICO)

I. Dados los siguientes datos de nacimientos y de población residente en el año 2020 en Ourense:

Grupos de edad	Población a 1 de julio de 2020			Nacimientos por edad de la madre en 2020		
	Total	Hombres	Mujeres	Total	Hombres	Mujeres
0-14	29.653	15.309	14.344			
15-19	10.908	5.624	5.283	25	12	13
20-24	11.441	5.898	5.542	86	40	46
25-29	13.094	6.698	6.396	248	135	113
30-34	14.704	7.406	7.298	446	237	209
35-39	17.945	9.077	8.868	445	240	205
40-44	21.998	10.969	11.029	183	107	76
45-49	22.101	11.044	11.057	23	14	9
50 y más	163.852	75.254	88.598			
Total	305.696	147.279	158.415	1.456	785	671

Calcular para el año 2020:

- La tasa bruta de natalidad
- La tasa general o global de fecundidad
- El ratio (o relación) de masculinidad al nacimiento
- Sabemos que el índice sintético de fecundidad (o indicador coyuntural de fecundidad) toma el valor 0,9. Si otra provincia tiene una tasa bruta de natalidad superior, ¿tendrá un índice sintético de fecundidad mayor?. Razone la respuesta.

II. Conociendo la siguiente información demográfica para un territorio:

Población residente a 1 de enero de 2019	1.119.596
Tasa de crecimiento anual acumulativa (o acumulada) en 2019 (tanto por cien)	0,23
Tasa bruta de mortalidad en 2019 (tanto por mil)	11,0
Tasa de migración neta en 2019 (tanto por mil)	7,5

Calcule:

- La población residente en el territorio a 1 de enero de 2020.
- El número de defunciones y de nacimientos, el saldo vegetativo y el saldo migratorio en el territorio durante el año 2019.

Empleando la tasa de crecimiento anual de la población que figura en la tabla anterior, estime la población a 1 de enero de 2025 suponiendo:

- Un crecimiento geométrico
- Un crecimiento logístico con un umbral máximo de 1.200.000 de habitantes, y que alcanzó un tamaño de 600.000 de habitantes en 2000.

III. Dado el siguiente extracto de una tabla de mortalidad:

	m_x	a_x	q_x	l_x	d_x	L_x	T_x	e_x
Edad	Tasa de mortalidad (‰)	Promedio de años vividos en el último año de vida	Riesgo de muerte (‰)	Supervivientes	Defunciones teóricas	Población estacionaria	Tiempo vivido	Esperanza de vida
0	0,00228	0,1300	0,00228	100.000	228	99.801	8.283.851	82,84
1	0,00012	0,4800						
2	0,00012	0,4700	0,00012	99.760	12	99.754	8.084.284	81,04
(...)								

a) Calcule las celdas vacías q_1 , l_1 , d_1 , L_1 , T_1 e e_1 . Interpretar los resultados.

Para analizar la morbilidad de una enfermedad:

b) ¿Qué expresan la tasa de prevalencia y la tasa de incidencia de la enfermedad?

