



XUNTA DE GALICIA
CONSELLERÍA DE FACENDA
Dirección Xeral da Función Pública



empregarse
procesos de ingreso

**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

NON abra o exame ata que o tribunal llo indique

EXERCICIO 1

Nun país querease facer unha enquisa dirixida a empresas. Por dispoñibilidade de recursos humanos decidiuse realizar un esquema de mostraxe estratificada (con mostras independentes e mostraxe aleatoria simple sen reposición en cada estrato) con tamaño mostral $n=15.000$. A partir dun directorio existente estratificáronse as empresas de acordo ao seu número de empregados.

- I. Por datos administrativos coñecemos a cuasivarianza da variable "cotizacións sociais por empresa" que utilizaremos para determinadas afixacións.
 - a. A partir dos datos da táboa I, calcule as afixacións igual, proporcional e óptima respecto á varianza (afixación de Neyman).

TABOA I

ESTRATO	POBOACION	CUASIVARIANZA POBOACIONAL
1	10.000	16.107
2	18.000	28.364
3	100.000	35.555
4	400.000	74.510
TOTAL	528.000	

- b. Dende o punto de vista teórico, compara a eficiencia (medida en termos da varianza) da afixación óptima (Neyman) e a mostraxe aleatoria simple sen reposición con igual tamaño de mostra.
- II. Realizouse a enquisa coa afixación que ven na táboa II. Investigouse a variable Y ="valor de consumos intermedios".
 - a. Calcule os estimadores separado e combinado de razón para o total da variable Y utilizando os valores da variable auxiliar X "tamaño da empresa".
 - b. Neste caso, en termos de eficiencia medida en termos de varianza, ¿cal dos dous estimadores sería máis adecuado e por que?

TABOA II

ESTRATO	TAMAÑO MOSTRAL	TOTAL MOSTRAL VARIABLE X	TOTAL POBOACIONAL VARIABLE X	TOTAL MOSTRAL VARIABLE Y
1	2.000	421.260	2.201.200	200.208.028
2	2.500	263.125	1.801.980	84.863.075
3	3.000	174.540	4.517.000	44.298.252
4	7.500	173.400	6.044.000	39.067.020

- III. No estrato 4 (tamaño poboacional 400.000) decidíronse realizar a posteriori dúas enquisas.
 - a. Primeiramente querease estudar utilizando unha esquema de mostraxe aleatoria simple sen reposición a proporción de empresas que dispoñen de páxina web. Partindo dunha enquisa piloto realizada anteriormente, a proporción de ditas empresas é do 30%. ¿Cal sería o tamaño mostral necesario supoñendo que o erro de mostraxe relativo máximo admisible sería do 3%?
 - b. Querease estudar a variable y ="consumo de auga". Decidiuse facer mostraxe sistemática con arranque aleatorio. ¿cales son os problemas e posibles solucións para lograr unha estimación da varianza do estimador?

EXERCICIO 2

Resolva os seguintes apartados:

- I. Dados os seguintes datos de nacementos e de cifras de poboación para o ano 2015:

Grupos de idade	Poboación a 1 de xullo de 2015	Nacementos por idade da nai en 2015		
		Total	Homes	Mulleres
15-19	5.250	37	22	15
20-24	6.302	123	61	62
25-29	7.463	290	153	137
30-34	9.011	644	338	306
35-39	10.996	579	288	291
40-44	10.932	152	72	80
45-49	11.371	10	7	3
15-49	61.323	1.835	941	894

1 de xaneiro de 2015 1 de xaneiro de 2016

Poboación total	318.391	314.853
-----------------	---------	---------

Calcular para o ano 2015:

- A taxa bruta de natalidade
- A taxa xeral ou global de fecundidade
- As taxas específicas de fecundidade por idade das nais e sexo do nacido
- O índice sintético de fecundidade (ou indicador conxuntural de fecundidade), tanto polo procedemento máis simplificado como pola forma máis precisa posible.
- A idade media á maternidade
- A idade media das nais que tiveron un fillo durante o ano 2015. ¿En que se diferencia da idade media á maternidade?
- A taxa bruta de reprodución, da forma máis precisa posible.

- II. A seguinte táboa contén a evolución no período 2012-2015 do movemento natural e migratorio nunha provincia:

	2012	2013	2014	2015
Nacementos	2.243	2.083	2.150	2.174
Defuncións	5.058	4.941	4.884	5.126
Inmigracións	10.315	9.688	9.630	9.130
Emigracións	10.052	10.230	10.471	11.022

Se a poboación a 1 de xaneiro do ano 2012 era de 348.902 persoas, calcular:

- O crecemento natural (ou vexetativo), o saldo migratorio e o crecemento total da poboación durante o conxunto do período 2012-2015.
- O tamaño da poboación a 1 de xaneiro de 2013, 2014, 2015 e 2016
- Para cada ano, a taxa bruta de inmigración (ou índice de atracción), a taxa bruta de emigración, a taxa de migración neta e a taxa de migración bruta. Interpreta os resultados.
- A taxa de crecemento anual acumulativa (ou taxa de crecemento xeométrico composto) para o conxunto do período 2012-2016
- A taxa de crecemento composto continuo (ou instantáneo) para o conxunto do período

- f. Utilizando o crecemento do último ano, ¿cál sería o tamaño da poboación a 1 de xaneiro de 2020 supoñendo un crecemento loxístico cun limiar máximo de 500.000 de habitantes, e que acadou un tamaño de 250.000 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
Idade	Taxa de mortalidade (‰)	Promedio de anos vividos no último ano de vida	Risco de morte (‰)	Superviventes	Defuncións teóricas	Poboación estacionaria
0	0,0029	0,1200	0,0031	100.000	313	99.724
1	0,0001	0,4800	0,0001	99.687	11	99.681
2	0,0002	0,3600	0,0002	99.676	21	99.662
3	0,0000	0,5500	0,0000	99.655	0	99.655
4	0,0000	0,3400	0,0000	99.655	0	99.655
5	0,0001	0,4600	0,0001	99.655	10	99.649
6	0,0002	0,6800	0,0002	99.644	21	99.638
7	0,0000	0,3900	0,0000	99.623	0	99.623
8	0,0001	0,5900	0,0001	99.623	11	99.619
9	0,0001	0,6100	0,0001	99.613	11	99.608
10	0,0000	0,6200	0,0000	99.602	0	99.602
11	0,0000	0,5500	0,0000	99.602	0	99.602
12	0,0002	0,5600	0,0002	99.602	23	99.592
13	0,0004	0,5800	0,0004	99.579	35	99.565
14	0,0000	0,4800	0,0000	99.545	0	99.545
15	0,0000	0,4400	0,0000	99.545	0	99.545
16	0,0002	0,4800	0,0002	99.545	24	99.532
17	0,0002	0,5600	0,0002	99.521	24	99.510
18	0,0006	0,5300	0,0006	99.497	59	99.469
19	0,0006	0,5400	0,0006	99.438	55	99.413
20	0,0003	0,4800	0,0003	99.383	31	99.367

- a. Calcular as celas baleiras q_{13} , l_7 , d_{17} e L_{10} . Interpretar os resultados.
 b. Sabendo que a función de tempo por vivir (T_x) á idade 10 é 7.238.143, calcule a esperanza de vida ao nacemento.

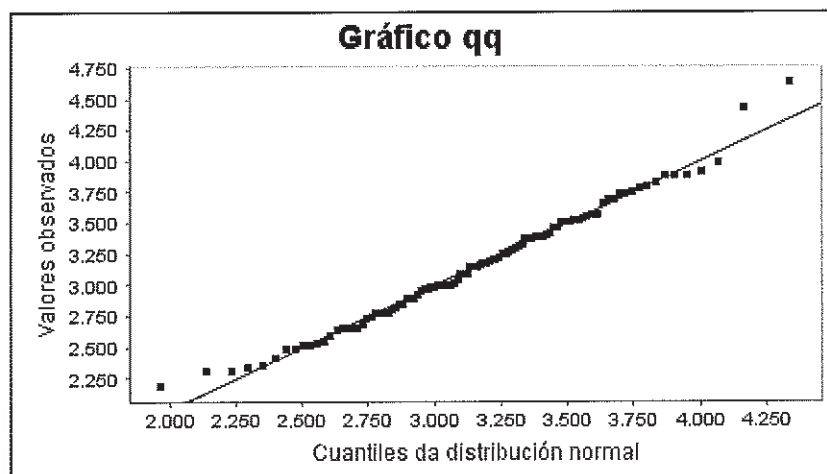
EXERCICIO 3

Resolva os seguintes apartados:

- I. O peso dos bebés que acaban de nacer en Galicia segue unha distribución normal de media 3,2 Kg e desviación típica 0,5 Kg:
 - a. Cal é a probabilidade de que un neno nacido pese máis de 1,5 Kg?
 - b. A partir de que peso se encontra o 75% dos nenos nados en Galicia? Interpreta o resultado.
 - c. No mes de xullo está previsto que nazan 1700 nenos, cal é a probabilidade de que 10 nenos pesen menos de 1,5 kg?
 - d. Nun estudio específico realizado na provincia de A Coruña para determinar o peso dos bebés extraeuse unha mostra aleatoria simple de 100 bebés que acaban de nacer. Pódese aceptar que esta mostra segue unha distribución normal cos seguintes resultados achegados polo test de Shapiro-Wilks (cunha confianza do 95%) e o gráfico qq-plot? Razoe a súa resposta.

Test de Shapiro-Wilks

n	Estatístico W	Valor crítico
100	1	0,83



- II. Extraemos unha mostra aleatoria simple de tamaño 1.100 persoas en Galicia. Con base nesta mostra e na táboa de continxencia seguinte resultante da súa explotación, deséxase estudar se existe asociación entre o nivel de estudos e a dificultade ou facilidade que teñen para chegar a fin de mes. Especifique claramente a hipótese nula e alternativa neste escenario de traballo, o criterio de decisión e a decisión final. Realizar o contraste apropiado cun nivel de significación do 5%.

Nivel de estudos	Chega con facilidade a fin de mes	Chega con dificultade a fin de mes	TOTAL
Educación primaria	79	111	190
Educación secundaria 1ª etapa	174	265	439
Educación secundaria 2ª etapa e postsecundaria	88	98	186
Educación superior	180	105	285
TOTAL	521	579	1.100

- III. Nun estudo de incidencia da pobreza en Galicia sábese que na zona rural a porcentaxe de pobres é do 15%. Dunha mostra aleatoria simple de 400 persoas na zona urbana obtívose que o 13% eran pobres.
- Pódese afirmar que a porcentaxe de pobres na zona urbana é igual que na zona rural cun nivel de significación do 5%? Responda xustificadamente.
 - Utilizando os datos anteriores, interprete a seguinte saída que nos achega un paquete estatístico:

Contraste	Valor p
Bilateral	0,29

- IV. Un grupo de investigadores/ras conseguiron establecer un modelo teórico polo que o aforro e endebedamento da poboación ven explicado por dous factores comúns ou latentes en toda a poboación. Nun instituto de análise económica decidiron comprobar a existencia destes dous factores facendo unha enquisa en 130 países. Os datos empregados foron as porcentaxes de persoas en cada país que durante o pasado ano:
- Posuían unha conta bancaria e posuían tarxeta de crédito
 - Aforraron en entidades financeiras e aforraron en entidades non financeiras
 - Pediron cartos prestados ao banco, pediron a un prestamista, e pediron cartos prestados á familia e aos amigos

Realizouse unha análise factorial empregando o método de compoñentes principais con base na matriz de correlacións decidíndose seleccionar os dous primeiros factores. A seguinte táboa amosa a matriz de cargas factoriais da solución rotada empregando unha rotación Varimax.

Variables	Factores	
	1	2
Posuír unha conta bancaria	,859	-,315
Posuír unha tarxeta de crédito	,726	-,390
Aforrar en entidades financeiras	,890	,110
Aforrar en entidades non financeiras	-,138	,806
Pedir cartos prestados ao banco	,775	-,129
Pedir cartos prestados a un prestamista	-,066	,689
Pedir cartos prestados á familia e aos amigos	-,186	,847

- Para a primeira variable, determine a comunalidade reproducida polos dous factores e a súa especificidade. Interpreteas.
- Determine a porcentaxe da varianza explicada polo primeiro factor.
- Interprete os dous factores.
- O estatístico de Kaiser-Meyer-Olkin amosou un resultado de 0,733. Analice e interprete este dato dentro do contexto da análise factorial.
- As puntuacións factoriais nos dous primeiros factores e outras variables empregáronse para facer unha análise clúster que deu lugar a 4 grupos de países que foron denominados: ricos bancarizados, ricos familistas, emerxentes e pobres. Coa agrupación realizouse unha análise da varianza das puntuacións factoriais no primeiro factor fronte ao tipo de país supoñendo efectos fixos. A táboa resultante foi:

Fonte de variación	Media cadrática
Inter-grupos	14,750
Intra-grupos	,575

Formule o modelo ANOVA correspondente a esta situación con todas as súas hipóteses e o contraste de interese principal. Admitindo a validez das hipóteses do modelo formulado, empregue a información da táboa para resolver o contraste de interese xeral cunha significación do 95%.

EJERCICIO 1

En un país se quiere hacer una encuesta dirigida a empresas. Por disponibilidad de recursos humanos se decidió realizar un esquema de muestreo estratificado (con muestras independientes y muestreo aleatorio simple sin reposición en cada estrato) con tamaño muestral $n=15.000$. A partir de un directorio existente se estratificaron las empresas de acuerdo a su número de empleados.

- I. Por datos administrativos conocemos la cuasivarianza de la variable "cotizaciones sociales por empresa" que utilizaremos para determinadas afijaciones.
 - a. A partir de los datos de la tabla I, calcule las afijaciones igual, proporcional y óptima respecto a la varianza (afijación de Neyman).

TABLA I

ESTRATO	POBLACION	CUASIVARIANZA POBLACIONAL
1	10.000	16.107
2	18.000	28.364
3	100.000	35.555
4	400.000	74.510
TOTAL	528.000	

- b. Desde el punto de visto teórico, compara la eficiencia (medida en términos de la varianza) de la afijación óptima (Neyman) y el muestreo aleatorio simple sin reposición con igual tamaño de muestra.
- II. Se realizó la encuesta con la afijación que viene en la tabla II. Se investigó la variable Y ='valor de consumos intermedios'
 - a. Calcule los estimadores separado y combinado de razón para el total de la variable Y utilizando los valores de la variable auxiliar X 'tamaño de la empresa'.
 - b. En este caso, en términos de eficiencia medida en términos de varianza, ¿cuál de los dos estimadores sería más adecuado y por qué?

TABLA II

ESTRATO	TAMAÑO MUESTRAL	TOTAL MUESTRAL VARIABLE X	TOTAL POBLACIONAL VARIABLE X	TOTAL MUESTRAL VARIABLE Y
1	2.000	421.260	2.201.200	200.208.028
2	2.500	263.125	1.801.980	84.863.075
3	3.000	174.540	4.517.000	44.298.252
4	7.500	173.400	6.044.000	39.067.020

- III. En el estrato 4 (tamaño poblacional 400.000) se decide realizar a posteriori dos encuestas.
 - a. Primeramente se quiere estudiar utilizando un esquema de muestreo aleatorio simple sin reposición la proporción de empresas que dispones de página web. Partiendo de una encuesta piloto realizada anteriormente, la proporción de dichas empresas es del 30%. ¿cuál sería el tamaño muestral necesario suponiendo que el error de muestreo relativo máximo admisible sería del 3%?
 - b. Se quiere estudiar la variable y ="consumo de agua". Se decidió hacer un muestreo sistemático con arranque aleatorio. ¿cuáles son los problemas y posibles soluciones para obtener una estimación de la varianza del estimador?



EJERCICIO 2

Resuelva los siguientes apartados:

- I. Dados los siguientes datos de nacimientos y de cifras de población para el año 2015::

Grupos de edad	Población a 1 de junio de 2015	Nacimientos por edad de la madre en 2015		
		Total	Hombres	Mujeres
15-19	5.250	37	22	15
20-24	6.302	123	61	62
25-29	7.463	290	153	137
30-34	9.011	644	338	306
35-39	10.996	579	288	291
40-44	10.932	152	72	80
45-49	11.371	10	7	3
15-49	61.323	1.835	941	894

	1 de enero de 2015	1 de enero de 2016
Población total	318.391	314.853

Calcular para el año 2015:

- La tasa bruta de natalidad
- La tasa general o global de fecundidad
- Las tasas específicas de fecundidad por edad de las madres y sexo del nacido
- El índice sintético de fecundidad (o indicador coyuntural de fecundidad), tanto por el procedimiento más simplificado como por la forma más precisa posible.
- La edad media a la maternidad
- La edad media de las madres que tuvieron un hijo durante el año 2015. ¿En qué se diferencia de la edad media a la maternidad?
- La tasa bruta de reproducción, de la forma más precisa posible.

- II. La siguiente tabla contiene la evolución en el período 2012-2015 del movimiento natural y migratorio en una provincia::

	2012	2013	2014	2015
Nacimientos	2.243	2.083	2.150	2.174
Defunciones	5.058	4.941	4.884	5.126
Inmigraciones	10.315	9.688	9.630	9.130
Emigraciones	10.052	10.230	10.471	11.022

Si la población a 1 de enero de año 2012 era de 348.902 personas, calcular:

- El crecimiento natural (o vegetativo), el saldo migratorio y el crecimiento total de la población durante el conjunto del período 2012-2015.
- El tamaño de la población a 1 de enero de 2013, 2014, 2015 y 2016
- Para cada año, la tasa bruta de inmigración (o índice de atracción), la tasa bruta de emigración, la tasa de migración neta y la tasa de migración bruta. Interpreta los resultados.
- La tasa de crecimiento anual acumulativa (o tasa de crecimiento geométrico compuesto) para el conjunto del período 2012-2016
- La tasa de crecimiento compuesto continuo (o instantáneo) para el conjunto del período

- f. Utilizando el crecimiento del último año, ¿cuál sería el tamaño de la población a 1 de enero de 2020 suponiendo un crecimiento logístico con un límite máximo de 500.000 de habitantes, y que logró un tamaño de 250.000 de habitantes en 2000?

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

Edad	m_x Tasa de mortalidad (‰)	a_x Promedio de años vividos en el último año de vida	q_x Riesgo de muerte (‰)	l_x Supervivientes	d_x Defunciones teóricas	L_x Población estacionaria
0	0,0029	0,1200	0,0031	100.000	313	99.724
1	0,0001	0,4800	0,0001	99.687	11	99.681
2	0,0002	0,3600	0,0002	99.676	21	99.662
3	0,0000	0,5500	0,0000	99.655	0	99.655
4	0,0000	0,3400	0,0000	99.655	0	99.655
5	0,0001	0,4600	0,0001	99.655	10	99.649
6	0,0002	0,6800	0,0002	99.644	21	99.638
7	0,0000	0,3900	0,0000		0	99.623
8	0,0001	0,5900	0,0001	99.623	11	99.619
9	0,0001	0,6100	0,0001	99.613	11	99.608
10	0,0000	0,6200	0,0000	99.602	0	
11	0,0000	0,5500	0,0000	99.602	0	99.602
12	0,0002	0,5600	0,0002	99.602	23	99.592
13	0,0004	0,5800		99.579	35	99.565
14	0,0000	0,4800	0,0000	99.545	0	99.545
15	0,0000	0,4400	0,0000	99.545	0	99.545
16	0,0002	0,4800	0,0002	99.545	24	99.532
17	0,0002	0,5600	0,0002	99.521		99.510
18	0,0006	0,5300	0,0006	99.497	59	99.469
19	0,0006	0,5400	0,0006	99.438	55	99.413
20	0,0003	0,4800	0,0003	99.383	31	99.367

- Calcular las celdas vacías q_{13} , l_7 , d_{17} y L_{10} . Interpretar los resultados.
- Sabiendo que la función de tiempo por vivir (T_x) a la edad 10 es 7.238.143, calcule la esperanza de vida al nacimiento.

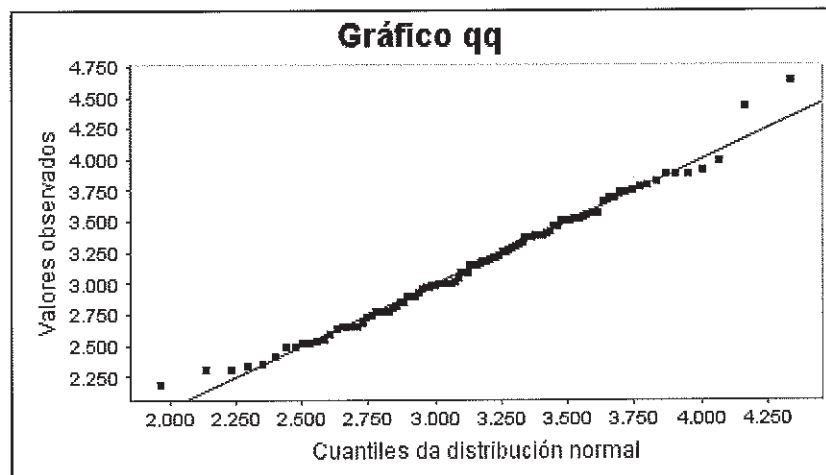
EJERCICIO 3

Resuelva los siguientes apartados:

- I. El peso de los bebés que acaban de nacer en Galicia sigue una distribución normal de media 3,2 Kg y desviación típica 0,5 Kg:
 - a. Cuál es la probabilidad de que un niño nacido pese más de 1,5 Kg?
 - b. ¿A partir de qué peso se encuentra el 75% de los niños nacidos en Galicia? Interprete el resultado.
 - c. En el mes de julio está previsto que nazcan 1700 niños, ¿cuál es la probabilidad de que 10 niños pesen menos de 1,5 kg?
 - d. En un estudio específico realizado en la provincia de A Coruña para determinar el peso de los bebés se extrajo una muestra aleatoria simple de 100 bebés que acaban de nacer. ¿Se puede aceptar que esta muestra sigue una distribución normal con los siguientes resultados presentados del test de Shapiro-Wilks (con una confianza del 95%) y el gráfico qq-plot? Razone su respuesta

Test de Shapiro-Wilks

n	Estadístico W	Valor crítico
100	1	0,83



- II. Extraemos una muestra aleatoria simple de tamaño 1.100 personas en Galicia. A partir de esta muestra y de la tabla de contingencia siguiente resultante de su explotación, se desea estudiar si existe asociación entre el nivel de estudios y la dificultad o facilidad que tienen para llegar a fin de mes. Especifique claramente la hipótesis nula y alternativa en este escenario de trabajo, el criterio de decisión y la decisión final. Realizar el contraste apropiado con un nivel de significación del 5%. %.

Nivel de estudios	Llega con facilidad a fin de mes	Llega con dificultad a fin de mes	TOTAL
Educación primaria	79	111	190
Educación secundaria 1ª etapa	174	265	439
Educación secundaria 2ª etapa e postsecundaria	88	98	186
Educación superior	180	105	285
TOTAL	521	579	1.100

- III. De un estudio de incidencia de la pobreza en Galicia se sabe que en la zona rural el porcentaje de pobres es del 15%. De una muestra aleatoria simple de 400 personas en la zona urbana se obtuvo que el 13% eran pobres.
- ¿Se puede afirmar que el porcentaje de pobres en la zona urbana es igual que en la zona rural con un nivel de significación del 5%? Responda justificadamente.
 - Utilizando los datos anteriores, interprete la siguiente salida que nos presenta un paquete estadístico:

Contraste	Valor p
Bilateral	0,29

- IV. Un grupo de investigadores/ras consiguió establecer un modelo teórico por el que ahorro y endeudamiento de la población viene explicado por dos factores comunes o latentes en toda la población. En un instituto de análisis económico decidieron comprobar la existencia de estos dos factores haciendo una encuesta en 130 países. Los datos empleados fueron los porcentajes de personas en cada país que durante el pasado año:
- Poseían una cuenta bancaria y poseían tarjeta de crédito
 - Ahorraron en entidades financieras y ahorraron en entidades no financieras
 - Pidieron dinero prestado al banco, pidieron a un prestamista, y pidieron dinero prestado a la familia y amigos

Se realizó un análisis factorial empleando el método de componentes principales a partir de la matriz de correlaciones decidiéndose seleccionar los dos primeros factores. La siguiente tabla muestra la matriz de cargas factoriales de la solución rotada empleando una rotación Varimax.

Variables	Factores	
	1	2
Poseer una cuenta bancaria	,859	-,315
Poseer una tarjeta de crédito	,726	-,390
Ahorrar en entidades financieras	,890	,110
Ahorrar en entidades no financieras	-,138	,806
Pedir dinero prestado al banco	,775	-,129
Pedir dinero prestado a un prestamista	-,066	,689
Pedir dinero prestado a la familia y amigos	-,186	,847

- Para la primera variable, determine la comunalidad reproducida por los dos factores y su especificidad. Interpretélas.
- Determine el porcentaje de la varianza explicada por el primer factor.
- Interprete los dos factores.
- El estadístico de Kaiser-Meyer-Olkin presentó un resultado de 0,733. Analice e interprete este dato dentro del contexto del análisis factorial.
- Las puntuaciones factoriales en los dos primeros factores y otras variables se emplearon para hacer un análisis clúster que dio lugar a 4 grupos de países que fueron denominados: ricos bancarizados, ricos familistas, emergentes y pobres. Con la agrupación se realizó un análisis de la varianza de las puntuaciones factoriales en el primer factor frente al tipo de país suponiendo efectos fijos. La tabla resultante fue:

Fuente de variación	Media cuadrática
Inter-grupos	14,750
Intra-grupos	,575

Formule el modelo ANOVA correspondiente a esta situación con todas sus hipótesis y el contraste de interés principal. Admitiendo la validez de las hipótesis del modelo formulado, utilice la información de la tabla para resolver el contraste de interés general con una significación del 95%

