



Universidad Nacional Abierta  
Vicerrectorado Académico  
Área de Matemática

Análisis de Datos (Cód. 778)  
Cód. Carrera: 126  
Fecha: Antes del 15-05-2020

**PRUEBA DE DESARROLLO / CORRECCIÓN MANUAL**  
**Tiempo de Prueba: 4 horas**

**INSTRUCCIONES**

1. El presente instructivo tiene como finalidad orientar el desarrollo de las actividades de estrategias sustitutivas, transitorias y finitas a realizarse solo por este lapso, para el logro del 60% o más de los objetivos establecidos para la aprobación de la asignatura **Análisis de Datos (Cód. 778)**, de acuerdo con los lineamientos que presentan el: Comunicado del Rector y demás Autoridades de la Universidad Nacional Abierta (UNA). En el Marco de la Universidad en Casa Abril 2020 y de los Subprogramas de Diseño Académico y Áreas Académicas y Carreras.
2. Con este trabajo se evalúan los objetivos 1 al 9 de la asignatura **Análisis de Datos (Cód. 778)**. En él se evidenciará las competencias matemáticas y destrezas adquiridas por el estudiante.
3. Los Trabajos prácticos son estrictamente individuales y una producción inédita del estudiante, cualquier indicio que ponga en duda su originalidad, será motivo para su anulación. Queda a discreción del asesor o profesor corrector, solicitar una verificación de los objetivos contemplados en el mismo, mediante una video conferencia o cualquier otra estrategia. El trabajo debe ser enviado al  
**Correo electrónico del asesor: ricorichard27@gmail.com o a su whatsapp: 0414-9328341 / 0426-9970190 (Avisa el envío del trabajo)**
4. Se permite el uso de la tabla de la **DISTRIBUCIÓN JI CUADRADO**.
5. **Debe justificar completamente sus respuestas para que éstas tengan validez.**
6. **Requerimientos exigidos para desarrollar, presentar y aprobar los trabajos**
7. Debes entregar por escrito el trabajo práctico a más tardar el **15 de Mayo 2020, SIN PRÓRROGA**, de acuerdo a los lineamientos emanados. Es necesario que para la entrega de estas actividades se sigan las orientaciones que presentamos a continuación:
  - 7.1 Responde de manera clara, ordenada, secuencial y argumentada el proceso seguido y las soluciones obtenidas al resolver el problema.
  - 7.2 Si usas un procesador de palabras debes usar como mínimo una letra tamaño 11 puntos y máximo 12 puntos, usa tipos de letra Arial o Times New Roman, emplea el editor de ecuaciones.
  - 7.3 Si vas a realizar el trabajo a mano, para ser enviado mediante un capture de imagen o Foto, usa letra legible y clara, preferiblemente hazlo en bolígrafo o marcador para facilitar su lectura, usar los símbolos matemáticos correspondientes y claramente escritos.
  - 7.4 El trabajo debe estar limpio y legible. Con un uso adecuado de la ortografía, los signos de puntuación

8. **LOS OBJETIVOS DEL TRABAJO SE EVALÚAN DE FORMA SUMATIVA UNA SOLA VEZ.** No existe la recuperación de los mismos.
9. **ESTÁ ATENTO A LOS CRITERIOS DE DOMINIO PARA EL LOGRO DEL OBJETIVO.** Recuerda el punto 7.1.
10. Los aspectos para la presentación del trabajo práctico son: (10.1) **portada** , la cual debes elaborar como se te indica a continuación (usa tu pc o a mano)

REPÚBLICA BOLIVARIANA DE VENEZUELA

UNIVERSIDAD NACIONAL ABIERTA

Centro Local \_\_\_\_\_ Oficina de Apoyo \_\_\_\_\_

**Asignatura:** \_\_\_\_\_ **Cód.** \_\_\_\_\_

Nombre Completo:

Número de cédula de identidad:

Fecha completa en la que entregó el trabajo:

Correo electrónico del estudiante:

**Resultados de Corrección´**

<b>OBJ N°</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
<b>1 = L.</b> <b>0 = N.L.</b>										

**CUERPO DEL TRABAJO**

(10.2) **Cuerpo del trabajo**, el cual contiene todas las respuestas a las actividades propuestas. Debe identificarlas con claridad con un título cada sección del trabajo e indicar el objetivo al que se corresponde cada una de ellas.

### PREGUNTAS

**P: 1, O: 1**

- a) Hallar la matriz de inercia de los siguientes puntos de  $R^3$ :  $x_1=(2,6,2)$ ;  $x_2=(0,2,2)$ ;  $x_3=(4,0,2)$  con sus masas respectivas 3,2,1 unidades.
- b) Hallar el momento de inercia de este conjunto de puntos con respecto al plano que pasa por el origen y tiene normal  $n_1 = (1,0,1)$ .

**Criterio de Dominio:** Para el logro de este objetivo debes responder correctamente todas las partes de la pregunta. Sus respuestas deben estar justificadas y mostrar todos los cálculos involucrados.

**P: 2, O: 2** Considere la siguiente tabla asociada a las medidas de las variables X, Y, Z sobre 8 individuos diferentes:

	<b>X</b>	<b>Y</b>	<b>Z</b>
1	2	-1	-2
2	3	0	-1
3	2	-1	-3
4	2	1	-2
5	3	1	-2
6	1	0	-1
7	2	-1	-3
8	1	-1	-2

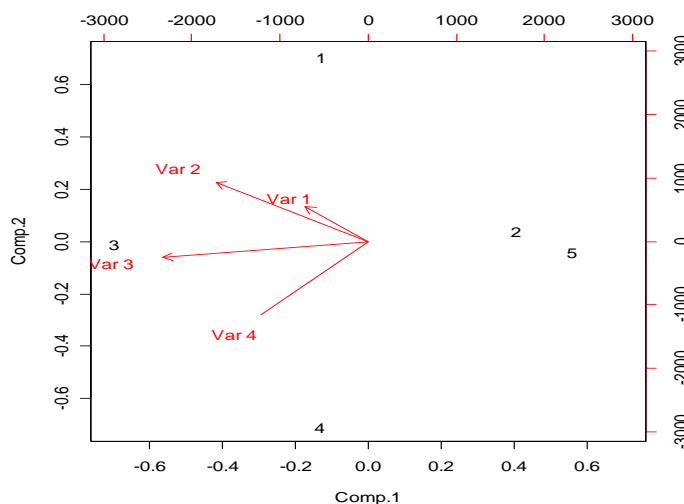
Suponiendo que todos los individuos tienen la misma importancia, realizar un análisis de componentes principales de los datos.

**Criterio de dominio:** Para el logro del objetivo debe llegar a la matriz  $F^o$  de las nuevas coordenadas de los puntos dados, partiendo de su matriz centrada en  $X^o$ . Sus respuestas deben estar justificadas y mostrar todos los cálculos involucrados.

**P: 3, O: 3** Los datos que reproducimos a continuación corresponden al color del pelo y de los ojos para una muestra de niños en una ciudad de Escocia.

<b>PELO</b> \ <b>OJOS</b>	<b>1-Azules</b>	<b>2- Claros</b>	<b>3- Castaño</b>	<b>4- Negros</b>
<b>1-Rubio</b>	1368	2577	1390	454
<b>2-Rojo</b>	170	474	420	255
<b>3-Castaño claro</b>	1041	2703	3826	1848
<b>4-Castaño oscuro</b>	398	932	1842	2506
<b>5-Negro</b>	1	11	33	112

Tomando en cuenta sus **valores propios: 1 0,3806 0,1393 0,02208** y basados en su gráfica. Hacer un análisis de correspondencia para estos datos.



**NOTA:** En este grafico los ejes para ambos factores tienen por intercepción el punto de partida de las flechas (estas flechas son las columnas de la tabla).

Especialista: Richard Rico

**P: 4, O: 4** Basándose en los datos de la **Pregunta 3** y en el **valor de su estadístico: 3683,87**, realice una prueba  $\chi^2$  para decidir si hay o no dependencia entre el color del pelo y de los ojos para una muestra de niños en una ciudad de Escocia.

Responda a la pregunta: ¿Existe alguna relación entre el color del pelo y de los ojos para una muestra de niños en una ciudad de Escocia?

**P: 5, O: 5** Señala con una **V** si las siguientes afirmaciones son verdaderas o con una **F** si son falsas:

**Justifica tu Respuesta**

- a. Una consecuencia importante de la invarianza del **análisis canónico** es que el valor de la correlación es máxima \_\_\_\_\_
- b. El problema a resolver en el **análisis de las correlaciones canónicas** es encontrar las combinaciones lineales de las observaciones en cada grupo, cuya correlación es máxima. \_\_\_\_\_

c. Si se tiene la matriz de covariancias de un par de variables Y, Z dada por  $Var \begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} \sum_{YY} & \sum_{YZ} \\ \sum_{ZY} & \sum_{ZZ} \end{pmatrix}$ , donde  $Var(Y) = \sum_{YY}$ , y se desea calcular  $Var(Y|Z)$  de la descomposición de:  $Y - \eta = \sum_{YZ} \sum_{ZZ}^{-1} (Z - \zeta) + Y - \eta - \sum_{YZ} \sum_{ZZ}^{-1} (Z - \zeta)$

Suponga que tienen correlación nula y que provienen de una distribución conjunta gaussiana e independiente de Z, entonces:

$$Var(Y - \eta) = Var\left(\sum_{YZ} \sum_{ZZ}^{-1} (Z - \zeta)\right) + Var\left(Y - \eta - \sum_{YZ} \sum_{ZZ}^{-1} (Z - \zeta)\right) \\ = \sum_{YY} = Var(Y).$$

**Criterio de Dominio:** Para el logro de este objetivo debes responder correctamente todas las partes de la pregunta. Sus respuestas deben estar justificadas y mostrar todos los cálculos involucrados.

**P: 6, O: 6** Señala con una **V** si las siguientes afirmaciones son verdaderas o con una **F** si son falsas:

**Justifica tu Respuesta**

- a. En la práctica al hacer la discriminación en un conjunto de observaciones clasificadas en dos grupos mediante la Discriminación Lineal de Fisher, lo que se hace es proyectar la observación a discriminar (6.7, 2) y los promedios de cada grupo, para el **Grupo 1** son (4.268, 1.345) y para el **Grupo 2** son (5.623, 2.055) en la dirección de **a**, y se decide atribuir la nueva observación **(6.7, 2)** al **Grupo 2** \_\_\_\_\_
- b. En el problema de discriminación los conjuntos se separan cuando la inercia dentro de los grupos es grande frente a la inercia entre grupos. \_\_\_\_\_
- c. En el problema de discriminación es posible que sea conocida la distribución de probabilidades de Y para individuos del grupo j, llamada  $P_j$ . \_\_\_\_\_

**Criterio de Dominio:** Para el logro de este objetivo debes responder correctamente todas las partes de la pregunta. Sus respuestas deben estar correctamente justificadas.

**Sugerencia:** En la parte (a) tome en cuenta el cálculo:  $(0.007758, 0.03123)(6.7, 2)^T = 0.1144$

**P: 7, O: 7** En cierta región del país habían arrestado a una persona por robar los huevos de las codornices del patio de su vecino. El presunto ladrón afirma que los huevos de codornices que había en su congelador eran los comprados para uso doméstico en el supermercado. Se hicieron mediciones en milímetros de los huevos de codornices confiscados al sospechoso y el Estado necesita probar (en un juicio en la corte) que los huevos de codornices encontrados en el congelador del sospechoso eran comprados para el uso **doméstico** y no los criados de manera **silvestre** por su vecino. Se sabe que en promedio las mediciones de estos huevos pueden variar de acuerdo con la procedencia de la codorniz (Domestica o Silvestre) y estas se resumen en la siguiente tabla:

Tipo de procedencia	Domestico		Silvestre	
Mediciones de sus ejes de simetría	Eje X	Eje Y	Eje X	Eje Y
<b>PROMEDIOS</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>3</b>

Los estimadores de las medias para las cuatro columnas de la tabla son: **2, 2, 1, 3**. Se quiere clasificar un nuevo par de datos **2, 3** que sabemos corresponde a una de esta dos categorías (Domestico ó Silvestre). Use la regla de máxima verosimilitud para ubicar una nueva observación  $y = (y_o, y_l) = (2, 3)$  en la categoría (Domestico ó Silvestre) que resulte más verosímil, suponiendo que **NO** hay igualdad de varianza ni covarianza y que estas matrices estimadas son:

$$\hat{\Sigma}_1 = \begin{pmatrix} 0.45 & 0.04 \\ 0.04 & 0.02 \end{pmatrix} \quad \hat{\Sigma}_2 = \begin{pmatrix} 0.35 & 0.04 \\ 0.04 & 0.03 \end{pmatrix}$$

$$\hat{\Sigma}_1^{-1} = \begin{pmatrix} 2.702 & -5.405 \\ -5.405 & 60.810 \end{pmatrix} \quad \hat{\Sigma}_2^{-1} = \begin{pmatrix} 3.370 & -4.494 \\ -4.494 & 39.325 \end{pmatrix}$$

La verosimilitud:  $(2\pi)^{-1} (\det \hat{\Sigma})^{-1/2} \exp(-\frac{1}{2}(y - \mu_h)^T \hat{\Sigma}^{-1} (y - \mu_h))$ ,  $h$  es la categoría.

**P: 8, O: 8** A continuación se te presenta una serie de enunciados, indica con una (V) si es verdadero o una (F) en caso de ser falso. **Justifica tu respuesta**

Si existen puntos  $X_1, X_2, \dots, X_n$  tales que  $\sum_i \mu_i X_i = 0$ , y denotamos:

$$d^2_{i,j} = \| X_i - X_j \|^2, \quad d_{..}^2 = \sum_i \sum_j \mu_i \mu_j d_{i,j}^2,$$

$$d_{i..}^2 = \sum_j \mu_j d_{i,j}^2, \quad d_{.,j}^2 = \sum_i \mu_i d_{i,j}^2,$$

$$D_{i,j} = (1/2)(d_{i..}^2 + d_{.,j}^2 - d_{i,j}^2 - d_{..}^2)$$

Entonces, podemos afirmar que:

A. Se cumple que  $(X_i, X_j) = D_{ij}$  ( )

B. Si la matriz  $D = (( D_{ij} ))$  no es semidefinida positiva, entonces existen puntos  $X_1, X_2, \dots, X_n$  que satisfacen las condiciones anteriores ( )

- C. Si la matriz  $D = \left( ( D_{ij} ) \right)$  no es semidefinida positiva y los autovalores en la diagonal no son negativos, entonces no se pueden calcular las primeras componentes principales (\_\_\_\_\_)

**Criterio de Dominio:** Para el logro del objetivo debes responder correctamente dos de los enunciados. Sus respuestas deben estar justificadas y mostrar todos los cálculos involucrados.

**P: 9, O: 9** Explique brevemente: Si se desea hacer pruebas de hipótesis con el estimador  $\psi = (\lambda_1 + \lambda_2 + \dots + \lambda_n) / (\lambda_1 + \lambda_2 + \dots + \lambda_p)$ , el cual está definido con los autovalores  $\lambda_1, \lambda_2, \dots, \lambda_n$ , y  $\lambda_1, \lambda_2, \dots, \lambda_p$  de las componentes principales tomados a partir de una distribución de  $X$  gaussiana que presenta una muestra aleatoria simple con distribución normal ¿Qué significa que el valor de  $\psi = 1$  (o bien próximo a 1)?

**FIN DEL TRABAJO**

<b>Tabla 4. Distribución de ji-cuadrado</b>					
	<b>Probabilidad de un valor superior</b>				
<b>Grados de libertad</b>	<b>0,1</b>	<b>0,05</b>	<b>0,025</b>	<b>0,01</b>	<b>0,005</b>
1	2,71	3,84	5,02	6,63	7,88
2	4,61	5,99	7,38	9,21	10,60
3	6,25	7,81	9,35	11,34	12,84
4	7,78	9,49	11,14	13,28	14,86
5	9,24	11,07	12,83	15,09	16,75
6	10,64	12,59	14,45	16,81	18,55
7	12,02	14,07	16,01	18,48	20,28
8	13,36	15,51	17,53	20,09	21,95
9	14,68	16,92	19,02	21,67	23,59
10	15,99	18,31	20,48	23,21	25,19
11	17,28	19,68	21,92	24,73	26,76
12	18,55	21,03	23,34	26,22	28,30
13	19,81	22,36	24,74	27,69	29,82
14	21,06	23,68	26,12	29,14	31,32
15	22,31	25,00	27,49	30,58	32,80
16	23,54	26,30	28,85	32,00	34,27
17	24,77	27,59	30,19	33,41	35,72
18	25,99	28,87	31,53	34,81	37,16
19	27,20	30,14	32,85	36,19	38,58
20	28,41	31,41	34,17	37,57	40,00
21	29,62	32,67	35,48	38,93	41,40
22	30,81	33,92	36,78	40,29	42,80

