**CARGO**: OFICIAL ESTADÍSTICO

**ÁREA**: INTELIGENCIA DE NEGOCIO

**TEMA**: EVALUACIÓN TEÓRICO –PRÁCTICO

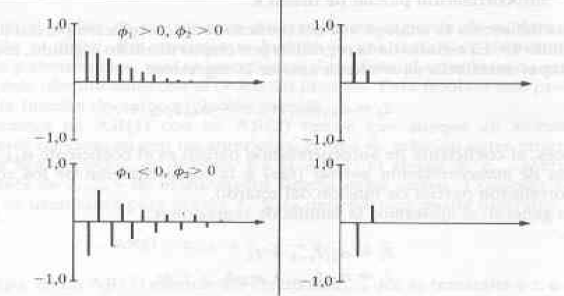
**TIEMPO REQUERIDO**: 2 HORAS

SECCIÓN A. LAS SIGUIENTES PREGUNTAS CORRESPONDE A PREGUNTAS ABIERTAS, FAVOR RESPONDA BAJO SU MEJOR CRITERIO

EL VALOR DE TODA PREGUNTA BIEN CONTESTADA ES UN PUNTO Y MAL CONTESTADA ES DE -0.5

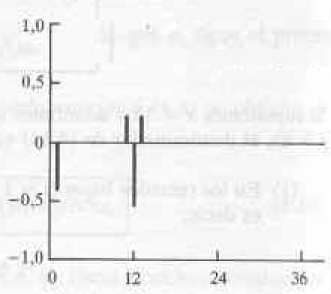
1. Una alta covarianza necesariamente indica:
   1. Alta volatilidad para cada una de las variables
   2. Un bajo valor de correlación entre las variables
   3. Un alto grado de asociación entre las variables
   4. Ninguna de las anteriores
2. ¿Qué tipo de matriz es requerida para calcular la varianza de un portafolio?
   1. Matriz cuadrada
   2. Matriz nula
   3. Matriz inversa
   4. Matriz identidad
3. La inclusión del término del error en el modelo econométrico se justifica por
   1. Recoge el efecto neto de una larga lista de variables omitidas que influye en la variable independiente, la mayoría de las cuales no son observables
   2. Captura los errores en la especificación de la forma funcional
   3. Incluye errores en la medición de las variables del modelo
   4. Todos los anteriores
   5. Ninguna de las anteriores
4. En el término de regresión lineal general, el significado lineal hace énfasis a:
   1. En término de los parámetros
   2. En término de los parámetros y error
   3. En término del error
5. La estimación de mínimo cuadrados de los parámetros de un modelo de regresión lineal general de manera matricial es:
6. Indique cuál supuesto no es parte de los supuestos de un modelo de regresión lineal general
   1. Las variables explicativas o exógenas son determinísticas
   2. No existe multicolinealidad entre las variables regresoras
   3. Los errores son homoscedasticos y no correlacionados
   4. Los errores no siguen la distribución normal
7. Indique cuál impacto se tiene por el evento de multicolinealidad entre variables explicativas en un modelo de regresión lineal
   1. Las varianzas de los estimadores son muy grandes
   2. Los ratios t son pequeñas están sesgadas a la aceptación de la hipótesis nula de no significancia individual
   3. Los intervalos de confianza de los parámetros son muy amplios
   4. Todas las anteriores
   5. Ninguna de las anteriores
8. Indique cuál no es un componente de una serie temporal en el contexto de un análisis descriptivo
   1. Ciclo
   2. Estacionalidad
   3. Tendencia
   4. Error
   5. Ninguna de las anteriores
9. Indique cuál(es) modelo(s) se utiliza para corregir el limitante de estimar el componente de la tendencia determinísticas
   1. Alisado Simple
   2. Alisado de Hold
   3. Proceso de Ito
   4. Todas las anteriores
   5. Ninguna de las anteriores
10. Cuando se determina que las raíces del polinomio autorregresivo son mayores que uno, entonces se puede decir que la serie es estacionaria en:
    1. Media
    2. Varianza
    3. Covarianza
    4. Todas las anteriores
    5. Ninguna de las anteriores
11. A partir de las siguientes figuras de autocorrelación simple y parcial indique que tipo de proceso se está considerando

f.a.s f.a.p



* 1. AR(1)
  2. MA(1)
  3. ARMA(1,1)
  4. ARMA(2,2)

1. A partir de las siguientes figuras de autocorrelación simple indique que tipo de proceso se está considerando



* 1. AR(12)
  2. AR(1)
  3. MA(1), MA(12)
  4. AR(1), AR(12)

1. Cuál de los siguiente elementos no son parte de la estimación del valor en riesgo
   1. Grado de sensibilidad del valor de la cartera con los factores de riesgo
   2. Forma de la distribución de probabilidad del cambio en los factores de riesgo
   3. Horizonte de tiempo
   4. Nivel de confianza
   5. Todos los anteriores
   6. Ninguno de los anteriores
2. El principal factor de riesgo del riesgo de mercado es
   1. Riesgo de inflación
   2. Tasa de descuento
   3. Fraudes
   4. Desempleo
3. El principal factor de riesgo del riesgo operativo es
   1. Riesgo de inflación
   2. Tasa de descuento
   3. Fraudes
   4. Desempleo
4. Indique la(s) razón(es) por la que se debe utilizar el logaritmo de la razón de los factores de riesgo
   1. Su interpretación es equivalente al rendimiento compuesto continuamente
   2. Las series de los rendimientos son estacionarias
   3. Es fácil extender el plazo del rendimiento (Ejemplo, el rendimiento de un activo a dos días es igual a la suma de los dos rendimientos de un día)
   4. Todas las anteriores
   5. Ninguna de las anteriores
5. Indique qué modelo de volatilidad de un instrumento financiero permite considerar las ponderaciones pasadas como fijas
   1. ARMA
   2. GARCH
   3. Volatilidad histórica
6. La principal característica del riesgo de crédito es
   1. La distribución de pérdidas y ganancias es sesgada a la derecha
   2. La media de la distribución de pérdida y ganancia es diferente de cero
   3. La exposición es igual al saldo del portafolio de análisis
7. Indique cuál de las siguiente políticas se utilizan para mitigar el riesgo de crédito
   1. Provisión y capital económico
   2. Provisión, capital económico y límites de crédito
   3. Capital económico y límites de crédito
   4. Ninguna de las anteriores
8. Cuál de los siguientes motivos son dirimentes para no considerar un sistema experto como una herramienta adecuada para la administración de riesgo
   1. Juicio subjetivo para considerar los pesos en las variables
   2. Experiencia en valorar los factores de riesgo
   3. Los dos anteriores
   4. Ninguna de la anteriores
9. Una matriz de transición considera que:
   1. Los créditos que pertenece a una misma calificación se consideran homogéneos
   2. La probabilidad de transición de los créditos depende solamente del estado inicial de calificación
   3. La probabilidad de transición es independiente del tiempo
   4. Ninguna de las anteriores
   5. Todas las anteriores
10. Se ha ajustado modelo de regresión logística con la siguiente fórmula: Y=a+β1X+ β2Z+ β3XZ ¿Cuál de las siguientes opciones es cierta?
    1. log(Y/(1-Y)) sigue una distribución binomial
    2. Si β3 es mayor que cero, el riesgo asociado a tener X y Z siempre es mayor que el producto del riesgo de sólo X por el riesgo de sólo Z
    3. Si β3 es menor que cero, el riesgo asociado a tener X y Z siempre es menor que la suma del riesgo de sólo X más el riesgo de sólo Z
    4. En este tipo de modelos debe cumplirse que β3 < |β1+ β2|
    5. Ninguna de las anteriores

SECCIÓN B. SELECCIONE LA PALABRA QUE MÁS SE AJUSTA A LAS SIGUIENTE DEFINICIONES

EL VALOR DE TODA PREGUNTA BIEN CONTESTADA ES UN PUNTO Y MAL CONTESTADA ES DE -0.5

1. Un \_\_\_\_\_\_\_\_\_\_\_\_\_\_ es un conjunto de variables aleatorias {zt} donde el índice t toma valores en un cierto conjunto C
2. proceso estocástico b. vector aleatorio c. proceso de ito d. proceso estacionario
3. Un \_\_\_\_\_\_\_\_\_\_\_\_\_ es estacionario en sentido estricto si: las distribuciones marginales de todas las variables son idénticas; las distribuciones finito-dimensionales de cualquier conjunto de variables sólo dependen de los retardos entre ellas
   1. proceso estocástico b. vector aleatorio c. proceso de ito d. proceso estacionario
4. Un proceso es estacionario en \_\_\_\_\_\_\_\_ si para todo t la media es constante, la varianza es constante y la covarianza solo depende de la variación entre dos variable y no del tiempo.
   1. los reales b.en los complejos c. sentido débil d. sentido fuerte
5. Un proceso \_\_\_\_\_\_\_ es aquel que para todo t la media es cero, la varianza es constante y la covarianza es cero.

a. estacionario b.estocástico c. ruido blanco d. estacionario en el sentido débil

1. Los principales factores de riesgo de crédito individual corresponde a la probabilidad de incumplimiento, \_\_\_\_\_\_\_ y severidad
   1. exposición b. concentración c. dependencia entre incumplimientos
2. En la medida de correlación de \_\_\_\_\_\_\_\_\_\_ se puede considerar que un valor de cero se interpreta como independencia entre las variable
   1. Pearson b. Tao de Kendall c. Spearman

SECCIÓN C. MARQUE CON X SI ES VERDADERO (V) O FALSO (F) EL SIGUIENTE PARRAFO

EL VALOR DE TODA PREGUNTA BIEN CONTESTADA ES UN PUNTO Y MAL CONTESTADA ES DE -0.5

1. El coeficiente de determinación es invariante a transformaciones lineales de las variables del modelo

Verdadero ( ) Falso( )

1. Los proceso MA finitos son siempre estacionarios y los AR finitos sólo lo son si las raíces están fuera del circulo unidad.

Verdadero ( ) Falso( )

1. Diremos que una serie es estacional cuando su valor esperado no es constante, pero varía con una pauta cíclica

Verdadero ( ) Falso( )

1. La transformación de Box Cox permite realizar transformaciones a los datos para obtener series estacionarias

Verdadero ( ) Falso( )

1. La diferencia o diferencia logarítmica permite realizar transformaciones a los datos para obtener series estacionarias

Verdadero ( ) Falso( )

1. El valor en riesgo (VaR) es la máxima pérdida que una institución podría observar en el caso de presentarse cambios en los factores de riesgo durante un horizonte de tiempo definido con un nivel de probabilidad determinado

Verdadero ( ) Falso( )

1. El modelo Delta VaR (la derivada del VaR de un portafolio) permite analizar la contribución marginal de un instrumento o factor de riesgo al riesgo total del portafolio

Verdadero ( ) Falso( )

1. La mayor preocupación que tiene los administradores de riesgo de crédito es la ocurrencia desproporcional de incumplimientos conjuntos que puede darse en un portafolio de préstamos ya que tiene un impacto directo en el nivel de pérdidas

Verdadero ( ) Falso( )

SECCIÓN D. EJERCICIO PRÁCTICO

EL VALOR DE TODA PREGUNTA BIEN CONTESTADA ES DOS PUNTOS Y MAL CONTESTADA ES DE -0.5

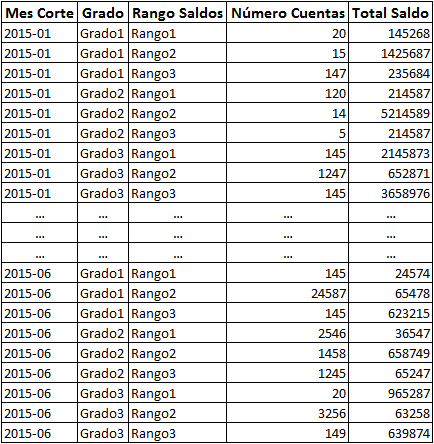
**1**. El archivo **muestra\_saldos.txt** dentro de la carpeta **Datos** es un reporte muestral sobre los saldos en las cuentas a cierre de la semana desde enero 2015 hasta junio 2015. Éste es un archivo de texto plano separado por tabulador y que usa como separador decimal el punto; además consta de 2,464,489 observaciones y 5 campos detallados a continuación:

|  |  |  |
| --- | --- | --- |
| **VARIABLE** | **TIPO** | **DESCRIPCIÓN** |
| sr\_fecha\_proceso | dd/mm/yyyy | Fecha de corte |
| sr\_cod\_estado | Varchar | Estado de la cuenta: activa (ACT), cancelada (CAN), cerrada (CER), inactiva (INA) |
| sr\_cod\_tipo\_cuenta | Varchar | Tipo de cuenta: de ahorros (AH) o corriente (CC) |
| sr\_md\_saldo\_contable | Numérica | Saldo contable en la cuenta |
| cr\_des\_grado | Varchar | Grado militar del titular de la cuenta |

# TAREA

Iniciar una sesión en R o RStudio y crear un nuevo script de R llamado **resumen.R**. En este script realizar las tareas siguientes:

* Cargar el reporte muestra\_saldos.txt con el nombre **saldos**
* Crear una tabla de doble entrada entre el grado militar (filas) y la fecha de corte (columnas) que describa la razón del número de cuentas corrientes canceladas versus el número de cuentas corrientes activas, y guardar el resultado en la carpeta **Resultados** con el nombre **Porcentaje**. Para guardar este resultado puede hacerlo tanto en un libro de Excel como en un archivo de texto plano separado por tabulador.
* Crear un nuevo campo dentro del data.frame “saldos” (llamado sr\_rango\_saldos) para discretizar la variable saldo contable en cuatro categorías ordenadas: <=1000, (1000,2000], (2000,5000] y >5000. Luego, debe crear el siguiente reporte (data drame): determinar el número de cuentas y saldo contable total para cada uno de los rangos de saldos creados, por el grado militar y por mes de corte, como, a modo de ejemplo, se muestra a continuación:



Grabar el resultado dentro de la carpeta **Resultados** con el nombre **Resumen**. Para guardar este resultado puede hacerlo tanto en un libro de Excel como en un archivo de texto plano separado por tabulador.

**2.** Cierto Hospital tiene como objetivo diseñar un plan que provisione el equipamiento de emergencia neonatal hacia las salas de parto. Los recién nacidos son evaluados entre uno y cinco minutos después del nacimiento mediante el uso de la prueba de *Apgar*, la cual está diseñada para determinar si un bebé necesita de cuidado de emergencia o atención médica extra. Un bebé cuyo score esté por debajo de 7 (en una escala de 0 a 10) en la escala de Apgar necesita atención extra. Tales bebés en riesgo son casos muy raros, de modo que el hospital no desea provisionar equipamiento de emergencia extra para cada parto. Por otro lado, los bebés en riesgo podrían necesitar atención inmediata, así que el aprovisionamiento de recursos de manera proactiva a los partos apropiados pueden salvar vidas.

La meta de los personeros del hospital es identificar con debida antelación situaciones con una *alta probabilidad de riesgo*, de modo que los recursos puedan ser asignados adecuadamente.

Para ello, se usará una muestra del archivo de datos de uso público sobre natalidad de los Estados Unidos para la construcción de un modelo que prediga la probabilidad que el recién nacido necesite atención extra. La siguiente tabla lista las columnas del conjunto de datos a usar (**SData. RData** dentro de la carpeta Datos).

|  |  |  |
| --- | --- | --- |
| **VARIABLE** | **TIPO** | **DESCRIPCIÓN** |
| atRisk | Lógica | TRUE si score Apgar 5-min < 7; FALSE caso contrario |
| PWGT | Numérico | Peso de la madre embarazada |
| UPREVIS | Numérico (entero) | Número de visitas médicas prenatales |
| CIG\_REC | Lógica | TRUE si fuma; FALSE caso contrario |
| GESTREC3 | Categórica | Dos categorías: < 37 weeks (prematuro) y >= 37 weeks |
| DPLURAL | Categórica | Nacimiento Múltiple, tres categorías: single/twin/triplet+ |
| ULD\_MECO | Lógica | TRUE si se tiene una moderada/alta tinción del líquido amniótico |
| ULD\_PRECIP | Lógica | TRUE para labor inusualmente corta (< tres horas) |
| ULD\_BREECH | Lógica | TRUE para una posición de nacimiento pélvica |
| URF\_DIAB | Lógica | TRUE si la madre es diabética |
| URF\_CHYPER | Lógica | TRUE si la madre tiene hipertensión crónica |
| URF\_PHYPER | Lógica | TRUE si la madre tiene hipertensión relacionada al embarazo |
| URF\_ECLAM | Lógica | TRUE si la madre experimenta eclampsia |

# TAREA

Iniciar una sesión en R o RStudio y crear un nuevo script de R llamado regresion.R. En este script detallar (código y comentarios) de las tareas siguientes:

* Cargar el conjunto de datos descrito en la problemática (data frame SData. RData)
* Dividir la muestra (aleatoriamente) en dos conjuntos: entrenamiento (70%) y prueba (30%). Fijar la semilla.
* Ajustar un modelo de regresión logística con la muestra de entrenamiento.
* Realizar predicciones sobre las muestras de prueba y de entrenamiento con el modelo ajustado.
* Determinar el umbral adecuado para usar el modelo encontrado como clasificador. Justifique su respuesta.
* Evalúe el modelo ajustado como clasificador en términos de precisión y exactitud. ¿Recomendaría el modelo encontrado para su puesta en producción?
* Si para un bebé no prematuro con ciertas características la probabilidad de estar en riesgo es del 1%, ¿A cuánto ascendería dicha probabilidad para un bebé prematuro con las mismas características? Suponiendo que la madre de éste bebé prematuro no ha realizado ninguna visita prenatal; un bebé en la misma situación cuya madre ha tenido tres visitas prenatales, ¿Cuál es la probabilidad de encontrarse en riesgo?

**3.** El archivo LoanStats.RData contienen datos de préstamo emitidos a través del período de tiempo de 2007 al 2016, incluyendo el estado actual del préstamo (actual, Abierto, totalmente pagada, etc.) y el archivo LCDataDictionary.xlsx corresponde al catálogo de las variables de la base indicada.

# TAREA

1. A partir de la base anterior Identifique a su mejor criterio el principal problema u oportunidad de negocio.

2. Realice un slide (lámina) donde se identifique la importancia del problema u oportunidad de negocio a partir de la base de datos LoanStats.RData, considerando:

* Un resumen de la metodología que se utilizó para identificar el problema.
* Presente los principales hallazgos de su análisis y la importancia del punto de dolor identificado por usted (se sugiere apoyarse en gráfico(s) para presentar el hallazgo alertado).