**Trabajo Práctico en R**

**Primera Parte**

**Alumno: ……………………………**

**Registro: …………………………..**

**Ejercicio 1**

Una práctica habitual de los principales bancos centrales del mundo es realizar un **relevamiento sistemático de los principales pronósticos macroeconómicos de corto y mediano plazo** que distintos analistas realizan sobre la evolución de la economía. Además de contribuir con la política de transparencia en la comunicación, la información que proporciona este tipo de encuestas resulta de **gran relevancia para las decisiones de política monetaria y económica**. Adicionalmente, esta información es relevante para las decisiones de consumo e inversión, constituyéndose como un bien público al proveer a la comunidad la mejor información posible respecto de las estimaciones que realizan los especialistas sobre el comportamiento futuro de las principales variables económicas.

El Banco Central de la República Argentina (BCRA) publica todos los meses el **Relevamiento de Expectativas de Mercado (REM)**. Para ello, en los últimos 3 días hábiles de cada mes realiza un relevamiento sobre las expectativas de los precios minoristas, la tasa de interés, el tipo de cambio nominal, el nivel de actividad económica y el resultado primario del sector público nacional no financiero. Para mayor información sobre el REM, visitar el siguiente link: <http://www.bcra.gov.ar/Pdfs/PublicacionesEstadisticas/Consideraciones%20del%20Relevamiento%20de%20Expectativas%20de%20Mercado%20(REM).pdf>

Utilizaremos una simulación de los pronósticos que 48 agentes tenían en diciembre de 2018 sobre la inflación anual de 2019. El ejercicio consiste en analizar qué esperan estos agentes sobre la tasa de inflación de 2019.

**Ejercicio a)**

Calcular las siguientes medidas de centralidad: Media, Mediana, Moda. Calcular las siguientes medidas de dispersión: Rango, Rango intercuartílico, Varianza y Desvío Standard.

|  |  |
| --- | --- |
| Media |  |
| Mediana |  |
| Moda |  |
| Rango |  |
| Rango intercuartílico |  |
| Varianza |  |
| Desvío Standard |  |

Calcular una tabla de frecuencias. ¿Cuáles son los 5 pronósticos más frecuentes?

|  |  |
| --- | --- |
| Pronóstico Inflación 2019 | Frecuencia |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Dibujar histograma y boxplot de las expectativas de inflación.

*Insertar boxplot*

¿Identifica algún valor que pueda considerarse atípico? ¿Cuál?

*Responder aquí*

**Ejercicio b)**

Al analizar los resultados y gráficos anteriores, se sospecha de un error al ingresar los datos. En efecto, se observa que para un pronóstico de inflación se ingresó el valor de 330, cuando en realidad debería ser 33,0.

Luego de corregir el error, repita el cálculo de medidas de centralidad y dispersión.

|  |  |
| --- | --- |
| Media |  |
| Mediana |  |
| Moda |  |
| Rango |  |
| Rango intercuartílico |  |
| Varianza |  |
| Desvío Standard |  |

¿Qué se modifican y cuales quedan igual?

*Responder aquí*

También realice nuevamente el histograma y el boxplot.

*Insertar histograma*

*Insertar boxplot*

Explique las diferencias que observa con el ejercicio a)

*Responder aquí*

**Ejercicio c)**

Viendo los resultados de a) y b), explique el significado de la siguiente frase:

*'La mediana y el RIC son robustas a la ocurrencia de valores atípicos, mientras que la media y la varianza no lo son'*

*Responder aquí*

**Ejercicio d)**

Supongamos que un grupo de 15 agentes envió tarde su respuesta, por lo que no fueron originalmente incluidos en el REM. El BCRA decide actualizar su publicación incorporando estos nuevos 15 pronósticos. Sorprendentemente, los 15 pronósticos son iguales: todos esperan una inflación del 30% para 2019. Luego de incorporar estos nuevos pronósticos, repita los cálculos anteriores.

Repita el cálculo de los indicadores de centralidad y dispersión. También repita el histograma y boxplot.

|  |  |
| --- | --- |
| Media |  |
| Mediana |  |
| Moda |  |
| Rango |  |
| Rango intercuartílico |  |
| Varianza |  |
| Desvío Standard |  |

*Insertar histograma*

*Insertar boxplot*

Compare con los resultados del ejercicio b). ¿Qué efectos tiene que se hayan incorporado 15 pronósticos iguales?

*Responder aquí*

**Ejercicio 2**

El 11 de Agosto de 2019 se realizaron las elecciones primarias para las presidenciales en Argentina. El resultado, aparentemente no era lo anticipado por el mercado, por lo que el 12 de Agosto de 2019 se produjo una fuerte caída en el índice MERVAL (índice que reúne las principales empresas que cotizan en la Bolsa de Comercio local).

En este ejercicio analizaremos qué tan excepcional fue el evento.

Luego de cargar la serie histórica de precios del índice MERVAL, construya un simple gráfico de la serie. Sólo se visualizarán datos desde 2019. La línea roja indica la fecha de referencia.

***Insertar gráfico aquí***

No nos interesa trabajar con el valor diario del índice, sino con las variaciones porcentuales diarias. Si bien todos los cálculos los realizará el programa, recuerde que para calcular el rendimiento (o variación porcentual) de una variable entre los períodos t y t-1, se procede de la siguiente manera:

En este caso, como estamos calculando rendimientos diarios, corresponde a la cotización del índice en un día cualquiera y a la cotización del día anterior.

Para visualizar lo que se acaba de hacer, transcriba el valor del índice MERVAL y la variación porcentual diaria para los primeros días de julio de 2020:

|  |  |  |
| --- | --- | --- |
| **Fecha** | **MERVAL** | **Rendimientos** |
| 1-jul-2020 |  |  |
| 2-jul-2020 |  |  |
| 3-jul-2020 |  |  |
| 6-jul-2020 |  |  |
| 7-jul-2020 |  |  |

Redondeando a 2 decimales, ¿qué significa el valor de 9,01 en la columna Rendimientos para el 6 de julio de 2020? ¿y qué significa el valor de -1,94 en la columna Rendimientos para el 7 de julio de 2020?

***Insertar respuesta aquí***

Calcular indicadores de centralidad y dispersión para el rendimiento diario del índice.

|  |  |
| --- | --- |
| Media |  |
| Mediana |  |
| Rango |  |
| Rango intercuartílico |  |
| Varianza |  |
| Desvío Standard |  |

Dibujar histograma y boxplot de los rendimientos

***Insertar gráfico aquí***

***Insertar gráfico aquí***

Ahora vayamos al evento que nos interesa estudiar ¿De cuánto fue la caída que se produjo el 12 de Agosto de 2019?

***Insertar respuesta aquí***

¿Cuál es el valor z de esta caída?

***Insertar respuesta aquí***

Analicemos si se cumple con el Teorema de Chebyshev. Calcular la cantidad de observaciones que, según Chebyshev, debería haber cómo mínimo a k desvíos standards de la media. Tome valores de k entre 2 y 18 (recuerde que el valor k=1 no es informativo). Calcule el porcentaje de días que el rendimiento del MERVAL estuvo a k desvíos standards de la media (k de 2 a 18).

***Insertar respuesta aquí***

Preste atención a los resultados anteriores. Note que, a partir de cierto valor de k, la cantidad de observaciones se mantiene constante. ¿Qué significa esto?

***Insertar respuesta aquí***

Grafique los resultados anteriores

***Insertar gráfico aquí***

¿Se cumple con el Teorema de Chebyshev?

***Insertar respuesta aquí***

Luego de este análisis, ¿Qué tan raro fue el evento del 12 de Agosto?

***Insertar respuesta aquí***

**Ejercicio 3**

En este ejercicio utilizaremos datos de la base “Gapminder”, la cual contiene distintos indicadores socioeconómicos para una gran cantidad de países. (<https://www.gapminder.org/data/documentation/>)

Luego de descargar los datos, realizar una exploración inicial de los mismos. Observamos que hay datos de distintos países para distinto año. En particular, hay información sobre la expectativa de vida, el PIB per cápita y la cantidad de habitantes en cada país para cada año. También hay una variable que nos indica a qué continente pertenece cada país.

¿De cuántos países distintos hay información?

*Responder aquí*

¿Para qué años hay datos disponibles?

*Responder aquí*

Para este ejercicio, utilizaremos datos sólo del año 2007.

**Ejercicio 3. a)**

Comenzaremos a explorar la variable 'Expectativa de vida'. Indicar medidas de centralidad y de dispersión de la variable.

|  |  |
| --- | --- |
| Media |  |
| Mediana |  |
| Rango |  |
| Rango intercuartílico |  |
| Varianza |  |
| Desvío Standard |  |

**Ejercicio 3. b)**

Realizar el histograma y el boxplot de la variable 'Expectativa de vida' en 2007

*Insertar histograma*

*Insertar boxplot*

¿Cómo clasificaría a esta distribución: simétrica, sesgada a la izquierda o sesgada a la derecha?

*Responder aquí*

¿Hay valores atípicos?

*Responder aquí*

¿Entre qué valores se encuentra la mayoría de los países?

*Responder aquí*

**Ejercicio 3. c)**

Calcule el valor z de cada país, para la variable 'Expectativa de vida' en 2007

A modo de ejemplo ¿Cuál es el valor z de Angola? ¿Y el de Francia?

*Responder aquí*

¿Qué significa tener un valor z negativo? ¿Y uno positivo?

*Responder aquí*

¿Cómo se interpreta el valor z de 0.689 que tiene Argentina?

*Responder aquí*

¿Qué significaría que alguien tenga un valor z igual a 0?

*Responder aquí*

Calcule la proporción de países que tienen un valor z entre -1 y 1. Haga lo mismo para los intervalos (-2,2) y (-3,3). Represente los resultados en un gráfico (cada punto rojo representa un país

*Insertar gráfico aquí*

Observe el gráfico y analice si se cumple el Teorema de Chebyshev

*Responder aquí*

¿Es normal que no haya valores fuera del rango (-3,3)?

*Responder aquí*

**Ejercicio 3. d)**

Realizar el histograma y el boxplot de la variable 'Expectativa de vida' para Europa y África

*Insertar histograma*

*Insertar boxplot*

¿Observa alguna diferencia entre ambos continentes? ¿Cuáles?

¿En qué continente la expectativa de vida es mayor?

*Responder aquí*

¿En qué continente la expectativa de vida es más 'dispersa'?

*Responder aquí*

**Ejercicio 3. e)**

Incorporemos al análisis la variable PIB per cápita, la cual puede tomarse como un indicador del desarrollo económico.

Realizar un diagrama de dispersión de las variables 'Expectativa de vida' y 'PIB\_per\_cápita'. Calcular el coeficiente de correlación entre estas dos variables

*Insertar gráfico aquí*

¿Encuentra alguna relación entre estas variables? ¿Puede haber alguna relación de causalidad?

*Responder aquí*

Realice el mismo gráfico de dispersión, pero sólo para Europa y África

*Insertar gráfico aquí*

Viendo este gráfico, ¿Qué opina sobre la relación entre las variables Expectativa de Vida y PIB per cápita? ¿Es igual para los dos continentes?

*Responder aquí*

Calcule el coeficiente de correlación entre estas dos variables, pero diferenciando entre Europa y África.

*Responder aquí*

¿Qué puede comentar de estos resultados?

*Responder aquí*

**Ejercicio 4**

El sitio de la Reserva Federal de Sant Louis (EEUU) (https://fred.stlouisfed.org/) cuenta con una gran cantidad de datos económicos, tanto de EEUU como a nivel global.

Aquí utilizaremos uno de los precios de referencia del petróleo (WTI, https://fred.stlouisfed.org/series/DCOILWTICO) y otro del oro (https://fred.stlouisfed.org/series/GOLDAMGBD228NLBM).

Cargamos los datos posteriores a 1986 y realizamos un gráfico para conocer las series

*Insertar gráfico aquí*

En base al gráfico anterior, ¿qué serie tiene una media mayor? ¿cuál una mediana mayor? ¿Qué serie presenta más volatilidad?

*Responder aquí*

¿Qué serie tiene un mayor coeficiente de variación?

*Responder aquí*

Compare estos resultados con los de varianza y desvío standard. ¿Cuándo debería usarse el coeficiente de variación en lugar del desvío standard?

*Responder aquí*

Como habrá notado, es difícil comparar dos series con valores muy distintos en un simple gráfico como el que se realizó al principio. Una solución puede ser construir “números índices”. Para ello, es necesario elegir una “base”, la cual tomará un valor de 100. En este caso, podemos tomar como base el primer día para el cual tenemos datos: el 2 de enero de 1986.

Para construir la serie “base 2 de enero de 1986 = 100” para el oro, lo que hacemos es dividir toda la serie por el valor que tenía el oro el 2 de enero de 1986, y se multiplican los resultados por 100. Se procede de la misma forma para obtener el número índice de la serie del petróleo. Las siguientes fórmulas resumen esto:

Luego de construir los números índice para cada serie, represente los mismos en un gráfico.

*Insertar gráfico aquí*

Observe que ambos índices han alcanzado por momentos valores iguales o superiores a 400. ¿Qué significa que el índice tenga un valor de 400?

*Responder aquí*

¿Nota algo raro en la serie del petróleo a mediados de 2020?

*Responder aquí*

**Ejercicio 5**

En este ejercicio y en los siguientes, recurriremos a juegos de azar para entender cómo juegan las probabilidades.

En este caso, simularemos un mazo de naipes españolas de 40 cartas, como el que se utiliza para jugar al Truco (<https://es.wikipedia.org/wiki/Truco_argentino>).

En una mano de Truco, cada jugador recibe 3 cartas. Podemos simular una mano realizando una extracción de 3 cartas sin reposición del mazo. Recuerde que “extracción sin reposición” significa que, luego de sacar una carta, no la volvemos a introducir en el mazo, por lo que no está disponible para la siguiente extracción.

¿Qué mano obtuvo?

*Responder aquí*

¿Cuántas manos distintas se pueden obtener? Esto se puede responder de manera analítica o mediante una simulación. Empecemos por ésta última manera, simulando 1 millón de manos y contando cuántas son únicas. Obviamente, para realizar esto es imprescindible una computadora.

Luego de realizar la simulación, muestre las primeras 5 manos obtenidas:

*Responder aquí*

En el millón de manos, seguramente hay muchas conformadas por las mismas cartas ¿Cuántas manos únicas se obtuvieron en la simulación?

*Responder aquí*

Ahora intentemos responder analíticamente cuál es la cantidad de manos distintas que se pueden obtener. Esto equivale al problema de calcular cuántos subconjuntos de 3 elementos podemos conformar en base a 40 elementos. La respuesta la da el número combinatorio.

Calcule el resultado de esta fórmula:

*Responder aquí*

¿Coinciden estos resultados con los de la simulación?

*Responder aquí*

¿Todas estas manos tienen la misma probabilidad de ocurrencia?

*Responder aquí*

¿Qué cree que hubiera pasado si, en lugar de simular 1 millón de manos se hubieran simulado 10 mil?

*Responder aquí*

Haga la prueba. Realice una simulación de 10 mil manos y cuente cuántas manos únicas obtuvo.

*Responder aquí*

¿Obtuvo los mismos resultados? ¿Qué simulación es más confiable, la de 1 millón de casos o la de 10 mil? ¿Por qué?

*Responder aquí*

¿Todas estas manos tienen la misma probabilidad de ocurrencia?

*Responder aquí*

Vayamos ahora a la probabilidad de que salgan cartas específicas ¿Cuál es la probabilidad de que me toque el 1 de espadas? De nuevo, esto se puede calcular de manera analítica o mediante simulaciones.

Empecemos contando cuántas veces apareció el 1 de espadas en la simulación de 1 millón de casos:

*Responder aquí*

Hagamos ahora el cálculo analítico. ¿Cuál es el valor teórico de la probabilidad de que toque el 1 de espadas? Si de un mazo de 40 cartas sacamos una de ellas, la probabilidad de que sea el 1 de espadas es 1/40. Pero cada mano consta de 3 cartas, por lo que el 1 de espada puede salir en la primera, la segunda o la tercer carta.

¿Cuál es esta probabilidad?

*Responder aquí*

Seguramente los resultados de la simulación no resultaron exactamente iguales a la proporción calculada analíticamente. ¿Son suficientemente cercanos? ¿O debemos sospechar que calculamos mal la probabilidad de que salga el 1 de espadas en una mano?

Hay distintas formas de responder a esta pregunta. Las más rigurosas, se verán en la segunda parte de la materia. Por el momento, podemos ver qué nos dice la distribución Binomial.

Pensemos en un evento binario: sale el 1 de espadas en una mano vs. no sale el 1 de espadas en una mano. La probabilidad de estos resultados es siempre la misma, independientemente de los resultados previos. Por lo tanto, podemos definir la siguiente variable:

En este caso, n es 1 millón y p es la probabilidad calculada analíticamente.

¿Cuál es el valor esperado de esta variable aleatoria?

**INTENTE RESOLVER ANALÍTICAMENTE. USE LOS RESULTADOS DE R PARA COMPROBAR.**

*Responder aquí*

¿Cuál es el desvío standard de esta variable aleatoria?

**INTENTE RESOLVER ANALÍTICAMENTE. USE LOS RESULTADOS DE R PARA COMPROBAR.**

*Responder aquí*

Volviendo a la cantidad de veces que salió el 1 de espadas en la simulación de 1 millón de manos, ¿a cuántos desvíos standards se encuentra del valor esperado de *X*? Es decir, calcule el valor z del resultado obtenido en la simulación.

**INTENTE RESOLVER ANALÍTICAMENTE. USE LOS RESULTADOS DE R PARA COMPROBAR.**

*Responder aquí*

En base a lo anterior, ¿cree que los resultados obtenidos en la simulación son consistentes con lo anticipado con la teoría? Justifique.

*Responder aquí*

**Ejercicio 6**

Manos Calientes en el Basketball

El artículo original fue escrito por profesores y asistentes de Estadística de UCLA y adaptado para el libro **OpenIntro por Andrew Bray y Mine Cetinkaya-Rundel**. La versión en inglés del mismo se puede descargar de <http://www.openintro.org/stat/labs/02A_Probability.pdf>

La traducción para el curso de Tamara Burdisso fue realizada por Ignacio Caro Solís.

Se suele decir que los jugadores de básquet que tienen varios aciertos consecutivos tienen **“manos calientes”**. Los fanáticos y jugadores han creído por mucho tiempo en este fenómeno, que niega el supuesto de que cada tiro es independiente del anterior. Sin embargo, un trabajo de Gilovich, Vallone y Tversky de 1985 recopiló evidencia que contradecía esta creencia y mostraba que **los tiros consecutivos son eventos independientes**. Este paper inició una gran controversia que continúa hasta el día de hoy, como se puede ver buscando en Google “mano caliente basketball” (“hot hand basketball”).

No esperamos resolver esta controversia aquí. Sin embargo, en este ejercicio emplearemos un enfoque para responder preguntas como esta. Los objetivos para este ejercicio son (1) pensar acerca de los efectos de eventos independientes y dependientes, (2) aprender cómo simular rachas de aciertos en R, y (3) comparar la simulación con los datos observados para determinar si el fenómeno de las “manos calientes” parece ser real.

Nuestra investigación se concentrará en la performance de un solo jugador: **Kobe Bryant** de los Ángeles Lakers. Su desempeño contra Orlando Magic en la final de la NBA de 2009 le valió el título de “el jugador más valioso” y muchos espectadores señalaron cómo pareció mostrar una “mano caliente”.

Luego de cargar los datos, analicemos brevemente los mismos.

Cada fila de datos muestra un tiro realizado por Kobe Bryant. Observamos que hay cinco variables:

**“vs”:** indica el equipo rival. En este caso, siempre es Orlando Magic (ORL)

**“game”:** indica el número de partido de la serie vinal (valores de 1 a 5)

**“quarter”:** indica el cuarto de cada partido (de 1 a 4)

**“time”:** indica el tiempo transcurrido en el cuarto (los cuartos duran 12 minutos)

**“description”:** una descripción sobre el tiro

**“basket”:** indica si convirtió el tiro (“H”, por \*hit\* en inglés) o si falló (“M”, por \*miss\* en inglés)

Con ver a simple vista la secuencia de aciertos y desaciertos puede ser difícil evaluar si Kobe estaba teniendo una *mano caliente* o no. Una forma en la que podemos encarar esta pregunta es considerar la creencia de que los jugadores con manos calientes tienden a tener rachas en sus lanzamientos.

Para este ejercicio **vamos a definir el largo de una racha como el número de aciertos consecutivos que tiene un jugador hasta que comete un error**.

Por ejemplo, analicemos la secuencia de aciertos y errores en los nueve lanzamientos de Kobe durante el primer cuarto del juego 1. ¿Cuál fue esta secuencia?

*Responder aquí*

Dentro de los nueve intentos de tiro anteriores, hay seis rachas. Recordemos que cada racha termina con un tiro errado (“M”), por lo que la primer racha está compuesta por la secuencia (“H” “M”), la segunda racha por la secuencia (“M”), etc. **Definamos como el largo de cada racha a la cantidad de aciertos (H) en ella**. En el caso de las seis rachas anteriores, el largo de las mismas es: 1, 0, 2, 0, 0, 0 (en orden de ocurrencia).

¿Qué significa una racha de largo 1 (es decir, cuántos aciertos y errores hay en una racha de 1)? ¿Y en una racha de 0?

*Responder aquí*

Al cargar los datos, se cargó también una función que puede ser usada para calcular los largos de todas las rachas, pudiendo analizar luego su distribución. Por ejemplo, podemos calcular los largos de las rachas para los 9 lanzamientos del primer cuarto del juego 1:

*Responder aquí*

Puede verse que los resultados coinciden con los que habíamos calculado previamente.

Calcularemos los largos de las rachas para todos los juegos y realizaremos un gráfico de barras de la distribución:

*Insertar gráfico aquí*

Describir la distribución de los largos de las rachas de Kobe durante la final de la NBA de 2009. ¿Cuál fue el largo de racha más frecuente? ¿Cuál fue el largo máximo de sus rachas?

*Responder aquí*

Ya mostramos que Kobe tuvo algunas extensas rachas de aciertos, pero ¿son suficientes para sostener la creencia de que tuvo manos calientes? ¿Con qué podemos compararlas?

Para responder estas preguntas, volvamos a la idea de independencia. **Dos procesos son independientes si el resultado de uno no afecta al resultado del otro. Si cada tiro que un jugador hace es un proceso independiente, haber acertado o no su primer tiro no afecta a la probabilidad de que acierte o no el segundo.**

Un jugador con una mano caliente tendría tiros no independientes unos de otros. Específicamente, si el jugador acierta su primer tiro, el modelo de la mano caliente predice que tendrá una mayor probabilidad de encestar en su segundo tiro.

Supongamos por un momento que este modelo es válido para Kobe. Durante su carrera, el porcentaje de las veces que Kobe encesta es alrededor de 45%, o en notación de probabilidad:

**P ( Tiro 1 = H ) = 0,45**

**Si acierta su primer tiro al aro y además tiene una mano caliente (tiros no independientes entre sí), entonces la probabilidad de que acierte su segundo tiro será mayor**. Supongamos que, cuando se tiene la *mano caliente*, la probabilidad de acertar el segundo tiro habiendo acertado el primero es 60%:

**P ( Tiro 2 = H | Tiro 1 = H )con mano caliente = 0,60**

Como resultado de las mayores probabilidades, se esperaría que Kobe tuviese rachas más largas.

Comparemos esto con una **perspectiva escéptica, en la que Kobe no tiene una mano caliente y en la que cada tiro es independiente del anterior**. Si acertó al aro en su primer tiro, la probabilidad de que acierte el segundo también sigue siendo 0.45.

**P ( Tiro 2 = H | Tiro 1 = H )sin mano caliente = P ( Tiro 1 = H ) = 0,45**

En otras palabras, haber acertado el primer tiro no afectó la probabilidad de que hubiera acertado también el segundo. **Si los tiros de Kobe son independientes, entonces tendría la misma probabilidad de acertar cada tiro (sin importar sus anteriores): 45%**.

Ahora que planteamos este problema en términos de tiros independientes, volvamos a la pregunta: **¿cómo definimos si las rachas de Kobe son lo suficientemente largas como para indicar que tuvo manos calientes? Podemos comparar los largos de sus rachas con los de alguien sin manos calientes: un tirador ‘independiente’**.

**Simulaciones**

Si bien no tenemos datos de algún tirador que sepamos que tenga tiros independientes, ese tipo de datos es muy sencillo de simular en R. En una simulación se fijan las reglas de un proceso aleatorio y luego la computadora usa números al azar para generar un resultado que adhiera a esas reglas.

Para simular un jugador de básquet que tiene tiros independientes debemos usar el mismo mecanismo que usaríamos para simular los lanzamientos de moneda. Los resultados posibles de cada lanzamiento son 2: "H" (acierto, por \*hit\* en inglés), "M" (fallo, por \*miss\* en inglés):

Para comparar a este jugador independiente con Kobe Bryant, supondremos que la probabilidad de "H" es 45% y la de "M" 55%.

Simular 5 lanzamientos del **jugador independiente**

*Responder aquí*

Dado que tenemos 133 lanzamientos de Kobe Bryant, simular 133 lanzamientos del jugador independiente y graficar los resultados

*Insertar gráfico aquí*

Los resultados de una simulación no siempre son iguales. Por lo tanto, para poder sacar conclusiones mediante simulaciones puede ser necesario repetir el proceso varias veces.

Simulemos 133 tiros de otros 4 jugadores independientes y grafiquemos los resultados, comparando con los de Kobe.

*Insertar gráfico aquí*

¿Son muy diferentes?

*Responder aquí*

Comparar las rachas de los jugadores independientes con las de Kobe Bryant

*Insertar gráfico aquí*

¿Son muy diferentes?

*Responder aquí*

En base a los resultados obtenidos, ¿considera que hay evidencia para afirmar que Kobe Bryant tuvo *manos calientes* durante la final de 2009? Justificar la respuesta.

*Responder aquí*

**Ejercicio 7**

En este ejercicio, jugaremos con el lanzamiento de monedas.

**Ejercicio 7.a)**

Realizar 10 lanzamientos de una moneda **equilibrada**. Entendemos por moneda equilibrada a una moneda en donde la probabilidad de obtener "cara" es la misma a la de obtener "ceca".

¿Qué resultados obtuvo?

*Responder aquí*

¿Qué proporción de caras y cecas?

*Responder aquí*

Representar los resultados en un gráfico

*Insertar gráfico aquí*

Los resultados de una simulación no siempre son iguales. Por lo tanto, para poder sacar conclusiones mediante simulaciones puede ser necesario repetir el proceso varias veces.

Repitamos el lanzamiento de una moneda equilibrada otras 5 veces y grafiquemos la proporción de caras y cecas obtenida en cada caso.

*Insertar gráfico aquí*

**Ejercicio 7.b)**

Realizar 10 lanzamientos de una moneda donde la probabilidad de obtener cara es 0,55 y la de obtener ceca es 0,45. Es decir, **la moneda no es equilibrada sino que está cargada a favor de las caras**.

¿Qué resultados obtuvo?

*Responder aquí*

¿Qué proporción de caras y cecas?

*Responder aquí*

Representar los resultados en un gráfico

*Insertar gráfico aquí*

Nuevamente, para poder sacar conclusiones mediante simulaciones puede ser necesario repetir el proceso varias veces.

Repitamos el lanzamiento de una moneda cargada otras 5 veces y grafiquemos la proporción de caras y cecas obtenida en cada caso.

*Insertar gráfico aquí*

¿Obtuvo resultados muy distintos respecto al ejercicio 7.a)?

*Responder aquí*

Con sólo observar estos resultados, ¿estaría en condiciones de afirmar que la moneda del ejercicio 7.b) estaba cargada mientras que la del 7.a) era equilibrada?

*Responder aquí*

¿Qué tan probables considera los resultados obtenidos? Los siguientes gráficos de la distribución binomial pueden ayudarle a responder.

*Responder aquí*

**Ejercicio 7.c)**

Repetiremos el ejercicio 7.a), pero esta vez lanzando **1 millón** de veces la moneda equilibrada.

Muestre los primeros 10 resultados

*Responder aquí*

¿Qué proporción de caras y cecas obtuvo en el millón de lanzamientos?

*Responder aquí*

Representar los resultados en un gráfico

*Insertar gráfico aquí*

Viendo los resultados, ¿está en condiciones de afirmar que la moneda es equilibrada?

*Responder aquí*

Más allá de lo que se pueda responder “a ojo”, puede ser necesario recurrir a argumentos un tanto más rigurosos. Algunos de estos métodos serán vistos en la segunda parte de la materia. Por ahora, utilicemos lo que sabemos sobre la distribución Binomial.

La cantidad de caras que se obtienen al lanzar 1 millón de veces una moneda equilibrada es una variable aleatoria que sigue una distribución Binomial con .

¿Cuál es el valor esperado de esta variable?

**INTENTE RESOLVER ANALÍTICAMENTE. USE LOS RESULTADOS DE R PARA COMPROBAR.**

*Responder aquí*

¿Cuál es el desvío standard de esta variable aleatoria?

**INTENTE RESOLVER ANALÍTICAMENTE. USE LOS RESULTADOS DE R PARA COMPROBAR.**

*Responder aquí*

Volviendo a la cantidad de caras que obtuvo en el lanzamiento de 1 millón de monedas equilibradas, ¿a cuántos desvíos standards se encuentra del valor esperado? Es decir, calcule el valor z del resultado obtenido en la simulación.

**INTENTE RESOLVER ANALÍTICAMENTE. USE LOS RESULTADOS DE R PARA COMPROBAR.**

*Responder aquí*

En base a lo anterior, ¿cree que los resultados obtenidos en la simulación son consistentes con los de una moneda equilibrada? Justifique.

*Responder aquí*

**Ejercicio 7.d)**

Repetiremos el ejercicio 7.b), pero esta vez lanzando **1 millón** de veces la moneda con probabilidades 0,55 y 0,45.

Muestre los primeros 10 resultados

*Responder aquí*

¿Qué proporción de caras y cecas?

*Responder aquí*

Representar los resultados en un gráfico

*Insertar gráfico aquí*

Viendo los resultados, ¿está en condiciones de afirmar que la moneda no es equilibrada?

*Responder aquí*

Nuevamente, puede ser necesario recurrir a argumentos un tanto más rigurosos que simplemente evaluar un gráfico.

En el ejercicio anterior ya calculó el valor esperado y el desvío de la variable aleatoria “cantidad de caras en 1 millón de lanzamientos de una moneda equilibrada”.

Calcule a cuántos desvíos standards se encuentra del valor esperado el resultado obtenido con la moneda cargada. Es decir, calcule el valor z del resultado obtenido en la simulación.

**INTENTE RESOLVER ANALÍTICAMENTE. USE LOS RESULTADOS DE R PARA COMPROBAR.**

*Responder aquí*

En base a lo anterior, ¿cree que los resultados obtenidos en la simulación son consistentes con los de una moneda equilibrada? ¿o puede afirmar que la moneda estaba efectivamente cargada a favor de las caras? Justifique.

*Responder aquí*

Para terminar, ¿qué diferencias nota en la simulación con 10 lanzamientos y la de 1 millón de lanzamientos? ¿Cuál es más confiable?

*Responder aquí*

**Ejercicio 8**

En este ejercicio, jugaremos con el lanzamiento de dados de 6 caras. En particular, analizaremos qué tan probable es que salga un 1 o un 6.

**Ejercicio 8.a)**

Realizar 12 lanzamientos de un dado **equilibrado**. Llamamos dado equilibrado a aquel en donde todos los números tienen la misma probabilidad de ocurrencia.

¿Qué resultados obtuvo?

*Responder aquí*

¿Qué proporción de 1 y 6 obtuvo?

*Responder aquí*

Representar los resultados en un gráfico

*Insertar gráfico aquí*

En base a los resultados, analice la pertinencia de la siguiente frase:

*“Como el dado es equilibrado, en 12 lanzamientos cada número tiene que salir exactamente 2 veces.”*

*Responder aquí*

**Ejercicio 8.b)**

Realizar 12 lanzamientos de un dado, donde la probabilidad de obtener un 6 es 0,20 y la de obtener cada uno de los restantes números es 0,16. Es decir, **el dado no es equilibrado sino que está cargado a favor del 6.**

¿Qué resultados obtuvo?

*Responder aquí*

¿Qué proporción de 1 y 6 obtuvo?

*Responder aquí*

Representar los resultados en un gráfico

*Insertar gráfico aquí*

¿Obtuvo resultados muy distintos respecto al ejercicio 8.a)?

*Responder aquí*

En base a los resultados, analice la pertinencia de la siguiente frase:

*“Como el dado estaba cargado a favor del 6, en 12 lanzamientos tiene que salir más de dos veces el 6, mientras que los demás números tienen que salir una misma cantidad de veces.”*

*Responder aquí*

¿Qué tan probables considera los resultados obtenidos, tanto para el dado equilibrado como para el dado cargado? Los siguientes gráficos de la distribución binomial pueden ayudarle a responder.

*Insertar gráfico de la Distribución de 1 aquí*

*Insertar gráfico de la Distribución de 6 aquí*

**Ejercicio 8.c)**

Repetiremos el ejercicio 8.a), pero esta vez lanzando **1 millón** de veces el dado equilibrado.

Muestre los primeros 10 resultados

*Responder aquí*

¿Qué proporción de 1 y 6 obtuvo en el millón de lanzamientos?

*Responder aquí*

Representar los resultados en un gráfico

*Insertar gráfico aquí*

Viendo los resultados, ¿está en condiciones de afirmar que el dado es equilibrado? Por el momento, puede dar una respuesta “a ojo”. En la segunda parte de la materia haremos un “test de hipótesis para la diferencia de proporciones” para poder argumentar de manera más rigurosa.

*Responder aquí*

**Ejercicio 8.d)**

Repetiremos el ejercicio 8.b), pero esta vez lanzando **1 millón** de veces el dado con probabilidades levemente favorables para el 6.

Muestre los primeros 10 resultados

*Responder aquí*

¿Qué proporción de 1 y 6 obtuvo?

*Responder aquí*

Representar los resultados en un gráfico

*Insertar gráfico aquí*

Viendo los resultados, ¿está en condiciones de afirmar que el dado no es equilibrado? Por el momento, puede dar una respuesta “a ojo”. En la segunda parte de la materia haremos un “test de hipótesis para la diferencia de proporciones” para poder argumentar de manera más rigurosa.

*Responder aquí*

¿Qué diferencias nota entre la simulación con 12 lanzamientos y la de 1 millón de lanzamientos? ¿Cuál es más confiable para sacar conclusiones?

*Responder aquí*

**Ejercicio 9**

En este ejercicio simularemos un bolillero con 2 bolillas blancas y 5 negras.

**Ejercicio 9.a)**

Realizar 3 extracciones **sin reposición.**

¿Qué resultados obtuvo?

*Responder aquí*

¿Cuántos resultados distintos son posibles en 3 extracciones sin reposición?

*Responder aquí*

¿Cuál es la probabilidad de cada uno de estos resultados? Haga el cálculo de manera analítica. Más adelante realizaremos el cálculo mediante una simulación.

*Responder aquí*

Repetir 1 millón de veces la extracción de 3 bolillas sin reposición y contar la cantidad de bolillas blancas (las bolillas negras van a ser 3 - cantidad de bolillas blancas)

Mostrar los primeros 10 resultados

*Responder aquí*

Mostrar los resultados únicos y sus probabilidades estimadas en base a la simulación.

*Responder aquí*

Graficar los resultados en un histograma de frecuencias relativas

*Insertar gráfico aquí*

Comparar los resultados obtenidos con las probabilidades calculadas analíticamente.

*Responder aquí*

**Ejercicio 9.b)**

Repetir las 3 extracciones pero **con reposición**.

¿Qué resultados obtuvo?

*Responder aquí*

¿Cuántos resultados distintos son posibles con reposición?

*Responder aquí*

¿Cuál es la probabilidad de cada uno de estos resultados? Haga el cálculo de manera analítica. Más adelante realizaremos el cálculo mediante una simulación.

*Responder aquí*

¿Qué diferencias hay con la extracción sin reposición del ejercicio 9.a)?

*Responder aquí*

Repetir 1 millón de veces la extracción de 3 bolillas con reposición y contar la cantidad de bolillas blancas (las bolillas negras van a ser 3 - cantidad de bolillas blancas)

Mostrar los primeros 10 resultados

*Responder aquí*

Mostrar los resultados únicos y sus probabilidades estimadas mediante la simulación.

*Responder aquí*

Graficar los resultados en un histograma de frecuencias relativas

*Insertar gráfico aquí*

Comparar los resultados obtenidos con las probabilidades teóricas

*Responder aquí*

**Ejercicio 9.c)**

En base a los resultados de los ejercicios 9.a) y 9.b), explique las diferencias entre una extracción **sin reposición** y una **con reposición**.

*Responder aquí*

**Ejercicio 10**

Simularemos una ruleta con números del 0 al 36. Todos los números tienen la misma probabilidad de salir.

Pruebe de jugar 10 veces seguidas y muestre los números que salieron:

*Responder aquí*

De ahora en más, apostará siempre al mismo número. Elija el número al que quiere apostar (del 0 al 36).

**Ejercicio 10.a)**

Cada vez que juega tiene que pagar **1 peso como entrada**. En caso de ganar, recibe un **premio de 36 pesos**. Es decir, en caso de ganar, obtendrá una ganancia neta de 36-1 = 35 pesos.

Jugaremos un millón de veces al número elegido.

Calculamos los resultados (ganancia/pérdida) acumulados. ¿Cuál fue el resultado final?

*Responder aquí*

Graficamos los resultados acumulados:

*Insertar gráfico aquí*

Viendo los resultados, ¿Parece ser un juego equilibrado? ¿Conviene jugar o ser el Casino?

*Responder aquí*

**Ejercicio 10.b)**

Jugaremos nuevamente a la ruleta. Cada vez que juega tiene que pagar **1 peso como entrada**. Pero esta vez, en caso de ganar, recibe un **premio de 38 pesos** en lugar de los 36 de la ruleta anterior.

Jugaremos un millón de veces al número elegido

Calculamos los resultados acumulados. ¿Cuál fue el resultado final?

*Responder aquí*

Graficamos los resultados acumulados:

*Insertar gráfico aquí*

Viendo los resultados, ¿Parece ser un juego equilibrado? ¿Conviene jugar o ser el Casino?

*Responder aquí*

**Ejercicio 10.c)**

Jugaremos nuevamente a la ruleta, pero esta vez **el premio por ganar es 37 pesos. La entrada sigue siendo de 1 peso**.

Jugaremos un millón de veces al número elegido

Calculamos los resultados acumulados. ¿Cuál fue el resultado final?

*Responder aquí*

Graficamos comparando con los resultados anteriores:

*Insertar gráfico aquí*

Viendo los resultados con un premio de 37, ¿Parece ser un juego equilibrado? ¿Conviene jugar o ser el Casino?

*Responder aquí*

¿Qué relación encuentra entre el premio por ganar, el valor esperado del juego y la posibilidad de ganar o perder dinero en el largo plazo?

*Responder aquí*

**Ganancias esperadas del juego**

En los casos anteriores, se ha considerado una ruleta donde los resultados posibles son -1 (perder un peso) con probabilidad 36/37 o ganar un valor G con probabilidad 1/37. Recuerde que, aun ganando, tiene que pagar la entrada, por lo que la ganancia neta será G-1.

El valor esperado de jugar una vez a la ruleta es igual a:

En caso de jugar 1 millón de veces a la ruleta, como cada tirada es independiente de la otra, la ganancia esperada es igual a la ganancia esperada de jugar una vez multiplicada por un millón:

En las simulaciones de este ejercicio, se probó con G tomando valores de 36, 37 y 38. Calcule la ganancia esperada de jugar 1 millón de veces con estos premios y compare con los resultados obtenidos.

*Responder aquí*

Habrá notado que los resultados que obtuvo en las simulaciones no coinciden exactamente los valores esperados recientemente calculados. Recuerde que el valor esperado es el resultado promedio que debería obtener si repite indefinidamente el experimento. Para saber qué tan cerca o lejos pueden estar los resultados del valor esperado, hay que considerar el desvío standard de la variable aleatoria.

Comencemos calculando la varianza de jugar una vez a una ruleta con entrada -1 y premio G. Llamamos X al resultado que se puede obtener. En este caso, X puede valer -1 o G-1.

A partir de la fórmula anterior, puede calcular el desvío standard de jugar una vez a la ruleta con premios 36, 37 y 38:

*Responder aquí*

Para calcular la varianza del resultado de jugar 1 millón de veces a la ruleta, es necesario recordar que cada jugada de ruleta es independiente de la otra. De esta forma, la covarianza entre el resultado de dos tiradas de ruleta cualquiera será siempre igual a 0, y la varianza de jugar 1 millón de veces a la ruleta será igual a la varianza de jugar una sola vez multiplicada por 1 millón.

En base a la fórmula anterior, calcule el desvío standard del resultado de jugar 1 millón de veces a la ruleta, con premios 36, 37 y 38.

*Responder aquí*

Ahora que ya conoce el valor esperado y el desvío standard de la variable aleatoria “resultado de jugar 1 millón de veces a la ruleta” para cada uno de los premios considerados, puede evaluar si los resultados que obtuvo en la simulación están en línea con lo anticipado por la teoría.

Calcule el valor z de los resultados obtenidos en 1 millón de juegos de ruleta para los casos de premio igual a 36, 37 y 38.

*Responder aquí*

¿Cree que los resultados obtenidos están en línea con lo esperado?

*Responder aquí*

**Ejercicio 11**

En este ejercicio utilizaremos cotizaciones históricas de tres empresas que cotizan en la Bolsa de Comercio de Buenos Aires: Banco Galicia (GGAL), Banco Francés (FRAN) e YPF (YPFD). En particular, tenemos información sobre el rendimiento mensual que tuvo cada una de estas acciones desde enero de 2010 hasta agosto de 2020.

Para calcular el rendimiento mensual, lo que hicimos fue calcular el valor promedio del precio de cada acción para cada mes, y luego calcular las variaciones porcentuales mes a mes. Recuerde que para calcular la variación porcentual de una variable entre el período *t* y el período anterior (*t-1*), hay que realizar la siguiente cuenta:

**Ejercicio 11 a)**

Calcular indicadores de centralidad y dispersión para cada el rendimiento de cada una de las acciones.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **GGAL** | **FRAN** | **YPFD** |
| Media |  |  |  |
| Mediana |  |  |  |
| Rango |  |  |  |
| Rango intercuartílico |  |  |  |
| Varianza |  |  |  |
| Desvío Standard |  |  |  |

Dibujar histograma y boxplot de los rendimientos

*Insertar histograma*

*Insertar boxplot*

Analice los resultados. Suponiendo que la performance pasada es un indicador de lo que puede suceder a futuro, ¿Alguna acción luce más atractiva que otra? ¿Alguna acción promete mejores rendimientos? ¿Alguna acción promete menor riesgo?

*Responder aquí*

**Ejercicio 11 b)**

Calcule las covarianzas y coeficientes de correlación entre los rendimientos de cada par de acciones.

*Responder aquí*

Realice un gráfico de dispersión para el rendimiento de cada par de acciones.

*Insertar gráfico aquí*

¿Identifica alguna relación positiva o negativa entre los rendimientos? En caso afirmativo, ¿qué tan fuerte es la relación?

*Responder aquí*

**Ejercicio 11 c)**

Decide invertir en dos de estas acciones, 50% del dinero en cada una, aunque todavía no sabe qué par de acciones elegirá. Calcule el rendimiento y la varianza de cada una de las carteras posibles. Para ello, recuerde que si tiene una combinación de dos variables aleatorias: , entonces:

**INTENTE RESOLVER ANALÍTICAMENTE. USE LOS RESULTADOS DE R PARA COMPROBAR.**

*Responder aquí*

¿Qué rol juegan las covarianzas calculadas en el ejercicio 11 b)?

*Responder aquí*

¿Qué par de acciones debería elegir si prioriza el rendimiento, más allá del riesgo?

*Responder aquí*

¿Qué par de acciones debería elegir si quiere minimizar el riesgo de pérdidas?

*Responder aquí*

**Ejercicio 11 c)**

Suponga que, independientemente de los resultados anteriores, luego de leer informes de distintos analistas decidió que no era buena idea invertir en acciones del Banco Francés. Por lo tanto, ya tiene decidido que invertirá una parte de su cartera en acciones del Banco Galicia y la restante en YPF.

Resta decidir qué porcentaje de su cartera estará destinada a cada una. Para esto, realiza el cálculo del valor esperado y desvío standard de distintas carteras y anota los resultados en la siguiente tabla:

**INTENTE RESOLVER ANALÍTICAMENTE. USE LOS RESULTADOS DE R PARA COMPROBAR.**

*Responder aquí*

|  |  |  |  |
| --- | --- | --- | --- |
| **% en GGAL** | **% en YPFD** | **E(W)** | **σ (W)** |
| 0% | 100% |  |  |
| 10% | 90% |  |  |
| 20% | 80% |  |  |
| 30% | 70% |  |  |
| 40% | 60% |  |  |
| 50% | 50% |  |  |
| 60% | 40% |  |  |
| 70% | 30% |  |  |
| 80% | 20% |  |  |
| 90% | 10% |  |  |
| 100% | 0% |  |  |

Represente estos resultados en un gráfico ubicando en el eje X el riesgo de la cartera (desvío standard) y en el eje Y el retorno esperado:

*Inserte Gráfico aquí*

¿Existe algún punto de esa curva en donde un inversor racional nunca invertiría?

*Responder aquí*

¿Es posible que dos inversores racionales elijan puntos diferentes de la curva?

*Responder aquí*