

PROGRAMA (5 MINUTOS)

El programa de curso fue discutido en la primera clase quedó disponible en la web. ¿Tienen alguna pregunta?

VARIABLES ALEATORIAS (15 MINUTOS)

En la vida cotidiana muchas veces nos encontramos con situaciones que tienen resultados inciertos. Por ejemplo, si va a llover o no, la nota que obtendremos en una prueba o cuánto nos vamos a demorar desde nuestra casa a hasta la universidad. La **incertidumbre** es, por lo tanto, una parte fundamental de nuestras vidas.

Cuando enfrentamos situaciones de incertidumbre tenemos que **tomar decisiones y enfrentar sus consecuencias**. Por ejemplo, frente a la incertidumbre de si llueve o no, tenemos que decidir cómo queremos vestirnos y podría ser que llevemos más o menos abrigo del necesario. Frente a la incertidumbre de la nota que podamos sacar en una prueba, tenemos que decidir cuánto estudiar, y si optamos por salir y estudiar poco podríamos obtener una nota baja que afecte nuestra posibilidad de aprobar el ramo. De modo similar, si me demoro mucho hasta la universidad, puede ser que no alcance a entrar a una clase o llegue tarde a dar una prueba. Muchas otras áreas de la vida tienen incertidumbre y en varias de ellas las decisiones que tenemos que tomar tienen repercusiones más dramáticas que los ejemplos que hemos discutido hasta ahora. ¿Se les ocurren a ustedes algunos ejemplos?

Para enfrentar esas situaciones las **estadísticas** pueden ser de ayuda. Las estadísticas nos permiten modelar eventos para comprenderlos y anticiparnos a sus posibles resultados. Incluso podríamos intentar incidir en ellos.

El primer paso para modelar un evento es definirlo. Cuando estamos frente a un evento incierto, una forma de definirlo es como una **variable aleatoria**. Una variable es aleatoria cuando sus valores son inciertos y dependen de un evento con incertidumbre. Por ejemplo, podríamos definir una variable aleatoria X como "llueve", la cual tomará el valor 1 si efectivamente llueve y 0 en caso contrario. Sus valores dependen, lógicamente, de si llueve o no. También podemos definir una variable Y como "el tiempo que toma llegar desde casa hasta universidad", lo cual dependerá del viaje que realice el alumno.

Ejercicios Sobre Variables Aleatorias

- **Ejercicio 1.** La noticia que aparece en la siguiente página fue publicada en La Tercera del 26 de junio del 2015 y plantea que el cobre cerró con ganancias tras favorables datos económicos en China. ¿Cuál es la variable relevante de la noticia? ¿Es aleatoria o no?
- **Ejercicio 2.** La distancia que recorre un objeto que cae libremente por acción de la gravedad es proporcional al cuadrado del tiempo transcurrido. La imagen presentada en la siguiente página se tomó en medio segundo y fue capturada por medio de fotografía estroboscópica a 20 capturas por segundo. En la primera captura el balón se desplaza a 12 mm, en la segunda captura el balón se desplaza 4 veces la distancia anterior, posteriormente 9 veces y así sucesivamente. ¿Cuál es la variable relevante aquí? ¿Es aleatoria o no?

Ejercicio 1

Negocios

Cobre cerró con ganancias tras favorables datos económicos en China

El metal rojo terminó las operaciones a US\$2,59682 la libra comparado con US\$2,56098 la libra del cierre anterior.

Reuters
/ 23/06/2015 - 09:53

0 [url http://fw.to/eFvGHJa](http://fw.to/eFvGHJa) -A +A

El precio contado del cobre cerró con un avance de 1,4% este martes en la Bolsa de Metales de Londres después de que datos manufactureros alentadores del principal consumidor mundial, China, mejoraron el humor de los inversionistas, aunque el aumento fue contenido por una demanda estacionalmente más débil en el futuro cercano y por la fortaleza del dólar.

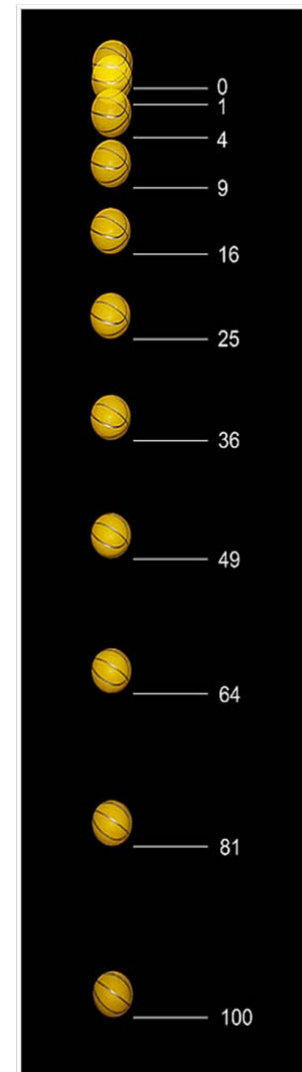
El metal rojo terminó las operaciones a US\$2,59682 la libra comparado con US\$2,56098 la libra del cierre anterior.

El índice preliminar PMI de HSBC/Markit para las manufacturas chinas mostró una contracción, pero también sugirió una estabilización. El componente de nuevos pedidos mostró expansión, mientras que los pedidos de exportaciones bajaron a un ritmo más lento.

Históricamente, la demanda de China por metales industriales se desacelera en el tercer trimestre.

El avance del dólar pesó sobre el metal, ya que la fortaleza de la divisa estadounidense hace que las materias primas como el cobre resulten más caras para empresas fuera de Estados Unidos.

Ejercicio 2



TIPOS DE VARIABLES ALEATORIAS (15 MINUTOS)

Variables Aleatorias Discretas

Una variable aleatoria discreta es aquella que solamente puede tomar valores enteros y se represente mediante un número finito de valores o sucesión infinita de valores (0, 1, 2, 3,...). Por ejemplo:

- Para elaborar una encuesta sobre PYMES, los encuestadores le preguntan a 1.000 personas por sus ingresos. Los encuestados tienen la posibilidad de declarar o no sus ingresos. Variable aleatoria (X): Número de encuestados que responden cuánto es su ingreso. Posibles valores de X: 0, 1, 2,..., 1000. Esto es una sucesión finita.
- Observar el número de personas que compra un boleto de bus en el Terminal San Borja. Variable aleatoria (X): Número de personas que compra boleto en el terminal. Posibles valores de X: 0, 1, 2,... Esto es una sucesión infinita.
- En la inauguración de un casino se invita a los estudiantes a responder una encuesta de satisfacción con la comida. Las respuestas pueden tomar valores entre 1 "muy insatisfecho" y 5 "muy satisfecho." Variable aleatoria (X): Nivel de satisfacción del estudiante. Posibles valores de X: 1, 2, 3, 4, 5. En este caso los resultados no se describen en valores numéricos.

VARIABLES ALEATORIAS CONTINUAS

Una variable aleatoria es continua cuando se puede representar mediante cualquier valor numérico dentro de un intervalo o colección de intervalos. Por ejemplo:

- Observar lo que se demora una orden en un local de comida rápida. Variable aleatoria (X): Tiempo que demora en entregar la orden. Posibles valores de X: $X \geq 0$.
- Construir un estadio para la U. Variable aleatoria (X): Porcentaje del proyecto terminado en 1.000 años. Posibles valores de X: $0 \leq X \leq 100$.
- Probar a qué temperatura ocurre un proceso químico nuevo. Variable aleatoria (X): Temperatura a la que tiene la reacción deseada (min. 80°C; máx. 120°C). Posibles valores de X: $80 \leq X \leq 120$.

EJERCICIO SOBRE LOS TIPOS DE VARIABLES ALEATORIAS

Determinar si las siguientes variables, son discretas o continuas:

1. La cantidad de personas que atiende la cafetería en una hora.
2. Si la cafetería logra o no cubrir sus costos en un día.
3. El tiempo que pasa entre que una y otra persona es atendida.
4. Cantidad de productos defectuosos en una línea de producción.
5. Producto está defectuoso.
6. Litros de bebidas desperdiciados en CCU.
7. Ingreso mensual en pesos de los clientes de una tienda.
8. Edad de las personas entrevistadas en la calle para un nuevo producto.

DISTRIBUCIONES DE PROBABILIDAD DISCRETA (10 MINUTOS)

La **distribución de probabilidad** de una variable aleatoria muestra cómo se reparten las probabilidades entre los distintos valores de la variable aleatoria. La distribución de probabilidad está definida por una **función de probabilidad** con la cual se puede obtener la probabilidad de cada valor de la variable aleatoria.

Para el caso de una **variable aleatoria discreta**, la distribución de probabilidad está definida por una función de probabilidad ($f(x)$). Esta función de probabilidad para una variable aleatoria discreta debe satisfacer las siguientes condiciones:

- $f(x_i) \geq 0 \quad \forall x_i$, es decir, la función de probabilidad siempre debe entregar valores mayores o iguales a cero, nunca negativos, para todo valor de x.
- $\sum_{i=1}^n x_i = 1$, es decir, al sumar la probabilidad de todos los valores posibles que puede tomar x, el resultado siempre será uno.

EJERCICIO SOBRE LAS DISTRIBUCIONES DE PROBABILIDAD DISCRETA

Un ejemplo que podemos replicar ahora, es ver la distribución de probabilidad de lanzar una moneda dos veces. En un lanzamiento la moneda puede tomar dos resultados con probabilidades equivalentes: cara (0.5) o cruz (0.5). Por lo tanto, en dos lanzamientos tenemos que:

- (Cruz, Cruz), $0.5 \times 0.5 = 0.25$
- (Cara, Cruz), $0.5 \times 0.5 = 0.25$
- (Cruz, Cara), $0.5 \times 0.5 = 0.25$
- (Cara, Cara), $0.5 \times 0.5 = 0.25$

En este ejemplo la suma de las probabilidades de los resultados es igual a 1 y ninguna probabilidad es menor a 0. ¿Cuál sería la distribución de probabilidad de...?

- No obtener ninguna cara $f(0) = ?$
- Obtener solamente una cara $f(1) = ?$
- Obtener dos caras $f(2) = ?$

TIPOS DE DISTRIBUCIÓN DE PROBABILIDAD DISCRETA (20 minutos)

Distribución Bernoulli

Es la distribución discreta más común. Se usa para cualquier caso en que tengamos solo dos decisiones. El ejemplo más clásico es el lanzamiento de una moneda, pero casi cualquier cosa puede "dicotomizarse". Por ejemplo, si alguien vota por el candidato A o B, si una empresa logra cumplir o no sus metas, si un departamento de una empresa tuvo que despedir a alguien en un año o no. La forma de la distribución bernoulli es la siguiente:

$$f(x_i) = p^{x_i} \cdot (1-p)^{1-x_i}$$

Analicemos la fórmula anterior. Matemáticamente, es la probabilidad de que x tome el valor 1 elevado a x , por el complemento, elevado a $(1-x)$, es decir, la probabilidad de que x tome el valor 0 elevado a $(1-x)$. Como x solo puede ser 1 o 0, los resultados solo pueden ser p , o bien, $1-p$. Por ejemplo, si x es 1 y decimos que $p=0.7$ si una persona vota por el candidato A, la probabilidad de que vote por B será $1-p=0.3$.

Distribución Binomial

La distribución bernoulli se aplica a una variable en particular, mientras que la distribución binomial se aplica a una suma de variables Bernoulli. Por ejemplo, la distribución binomial se utiliza cuando lanzo una moneda 10 veces y quiero saber la probabilidad de que salgan 4 caras. La forma de la distribución binomial es la siguiente:

$$f(x) = \binom{n}{x} \cdot p^{x_i} \cdot (1-p)^{n-x_i}$$

Esta fórmula es muy similar a la función Bernoulli, con la excepción del término conocido como coeficiente binomial y el $(n-x)$ en el exponente. El desarrollo de la combinatoria es igual a:

$$\binom{n}{x} = \frac{n!}{(n-x)! x!}$$

Es decir, n factorial dividido por $(n-x)$ factorial por x factorial.

¿Cuál es el uso del coeficiente binomial? Este coeficiente nos da cuenta de la cantidad de formas que podemos elegir x elementos de un conjunto n . Siguiendo el ejemplo de lanzar una moneda 10 veces y saber la probabilidad de que salgan 4 caras, el coeficiente binomial sería $\binom{10}{4}$. Este coeficiente indica todas las posibilidades de obtener 4 caras de un conjunto de 10 lanzamientos y equivale a 210.

Distribución Poisson

Esta distribución tiene un uso muy particular que la hace muy interesante. En concreto, la distribución poisson se aplica a variables que miden cuántos eventos ocurren en un ciclo, como por ejemplo, cuántos autos entran a Ciudad Empresarial en una hora. La forma de la distribución poisson es la siguiente:

$$f(x) = \frac{\lambda^x \exp(-\lambda)}{x!}$$

Esto es, lambda elevado a x, por exponencial elevado a menos lambda, dividido por x factorial. Lambda representa el promedio de la función. Por lo tanto, la probabilidad va a ir aumentando a medida que x se acerque a lambda. Volviendo al ejemplo de cuántos autos entran a Ciudad Empresarial en una hora y suponiendo que lambda es igual a 200, podemos decir que si en promedio entran 200 autos por hora a Ciudad Empresarial, es más probable que en una hora entren 190 a que entren 5.

Ejercicio Sobre los Tipos de Distribución de Probabilidad Discreta

De las siguientes variables, ¿cuáles podrían modelarse con las distribuciones descritas que hemos aprendido?

1. La cantidad de personas que atiende la cafetería en una hora.
2. Si la cafetería logra cubrir sus costos en un día o no.
3. El tiempo que pasa entre que una y otra persona es atendida.
4. Cantidad de productos defectuosos en una línea de producción.
5. Producto está defectuoso.
6. Litros de bebidas desperdiciados en CCU.
7. Ingreso mensual en pesos de los clientes de una tienda.
8. Edad de las personas entrevistadas en la calle para un nuevo producto.

STATA (15 MINUTOS)

Con las distribuciones que revisamos podemos replicar problemas de la vida real. Por ejemplo, supongamos que estamos administrando un *call center* y queremos saber cuántas personas atiende un trabajador en una hora. Si el único dato que tenemos es que en promedio atiende a 10 personas, podríamos simular, usando la función de distribución, a cuántas personas atiende en cada hora. Así tenemos valores que podrían perfectamente ser reales, como se muestra a continuación:

```
. set obs 30  
obs was 0, now 30  
  
. set seed 1234  
  
. gen atencion = rpoisson(10)  
  
. br  
  
.
```

	atencion
1	8
2	9
3	9
4	9
5	16
6	15
7	8
8	10
9	9
10	11
11	10
12	7
13	5
14	10
15	9
16	9
17	13
18	7
19	10
20	12
21	7
22	7
23	6
24	8
25	11
26	10
27	4
28	13
29	9
30	6

Por lo tanto, puede que en una hora atienda a 4 personas (fila 27) o 15 (fila 6). Pero siendo 10 personas el promedio de atención, ¿podemos decir que atender a 4 personas tiene la misma probabilidad que atender a 15?

La respuesta es no. Recordemos que estas probabilidades están modeladas por la distribución Poisson ya que estamos hablando de un evento (atender a una persona) que se repite en un ciclo igual (una hora). Veamos sus probabilidades:

```
. set obs 30
obs was 0, now 30

. set seed 1234

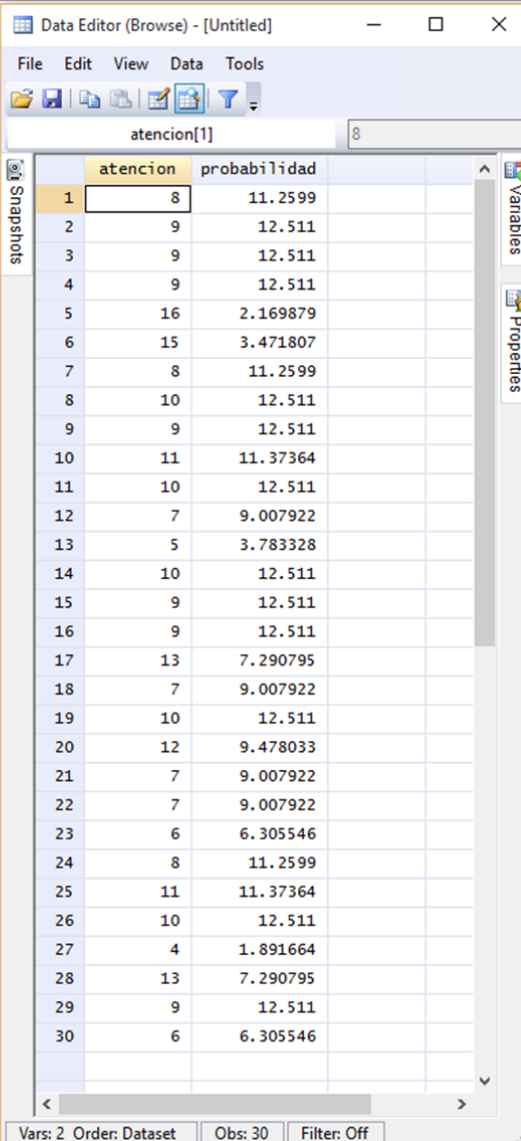
. gen atencion = rpoisson(10)

. br

. gen probabilidad = poissonp(10,atencion)*100

. br

.
```



	atencion	probabilidad
1	8	11.2599
2	9	12.511
3	9	12.511
4	9	12.511
5	16	2.169879
6	15	3.471807
7	8	11.2599
8	10	12.511
9	9	12.511
10	11	11.37364
11	10	12.511
12	7	9.007922
13	5	3.783328
14	10	12.511
15	9	12.511
16	9	12.511
17	13	7.290795
18	7	9.007922
19	10	12.511
20	12	9.478033
21	7	9.007922
22	7	9.007922
23	6	6.305546
24	8	11.2599
25	11	11.37364
26	10	12.511
27	4	1.891664
28	13	7.290795
29	9	12.511
30	6	6.305546

Aquí observamos que la probabilidad de que en una hora se atienda solamente a 4 personas es muy baja (1.89%), mientras que la probabilidad de que atienda a 15 personas es un poco más alta (3.47%). Si nos acercamos más al promedio, vemos que la probabilidad de que atienda a 10 personas es de 12%.

Esto es un buen ejemplo de cómo conocer las distribuciones de probabilidad nos puede ayudar a anticipar escenario más probable. En este caso, ¿qué es más probable que suceda? ¿Atenderán entre 9 a 12 personas por hora o más de 16? Es mucho menos riesgoso apostar por la primera posibilidad (9 a 12 personas por hora). Saber esto nos puede permitir tomar decisiones informadas, como por ejemplo, estimar cuántas personas tendremos que contratar para un funcionamiento óptimo del call center.

Descripción de los Comandos de Stata Utilizados en Esta Clase

- `<set obs #>`: Indica cuantas observaciones queremos crear. # es reemplazado por dicho número.
- `<set seed #>`: Cuando trabajamos con simulaciones y creamos números aleatorios, es importante siempre fijar un punto de partida o semilla (#) para que al replicar el ejercicio se obtengan los mismos resultados.

- `<gen newvar = exp>`: El comando `gen` es una abreviación para `generate`. Lo que hace es generar nuevas variables. El nombre de la variable será el que se reemplace en `newvar` y la forma que tendrá la variable es lo que se especifica en `exp`.
- `<rpoisson(#)>`: Genera números aleatorios que puedan describirse con una distribución poisson con promedio `#`.
- `
`: El comando `br` es una abreviación de `browse`. Lo que hace es mostrar la base de datos.
- `<poissonp(m,k)>`: Entrega la probabilidad de un número `k`, si la distribución es poisson con media `m`.