Probabilidad y estadística

Carina

16 de diciembre de 2018

Índice

1. PROBABILIDAD Y ESTADISTICA

Los primeros estudios de probabilidad fueron motivados por la posibilidad de acertar o perder en apuestas en los juegos de azar. (Este tema involucraba mucho dinero entre 2 jugadores eligiendo un número del 1 al 6), cada jugador elije un número talque cada jugador no repite número. Si un dado es lanzado, y si ha aparecido 2 veces el número de un jugador, entonces se puede hacer apuestas. Uno de los apostadores, Antonio de Gombaud, popularmente conocido como el caballero De Mere, deseando conocer la respuesta al problema plantea a Blaise Pascal (1623-1662) la situación. Pascal a su vez consulta con Pierre de Fermat (1601-1665) e inician un intercambio de cartas a propósito del problema. Esto sucede en el ano de 1654. Con ello se inician algunos esfuerzos por dar solución a éste y otros problemas similares que se plantean. Con el paso del tiempo se sientan las bases y las experiencias necesarias para la búsqueda de una teoría matemática que sintetice los conceptos y los métodos de solución de los muchos problemas particulares resueltos a lo largo de varios años.
El matemático ruso A. N. Kolmogorov (1903-1987) propone ciertos axiomas que a la postre resultaron adecuados para la construcción de una teoría de la probabilidad. Esta teoría prevalece hoy en día y ha adquirido el calificativo de teoría clásica. Actualmente la teoría clásica de la probabilidad se ha desarrollado y extendido enormemente gracias a muchos pensadores que han contribuído a su crecimiento, y es sin duda

Blaise Pascal
(Francia, 1623–1662)

Pierre de Fermat
(Francia, 1601–1665)

una parte importante y bien establecida de las matemáticas. Ha resultado útil para resolver problemas puramente matemáticos, pero sobre todo y principalmente, para modelar situaciones reales o imaginarias, en donde el azar es relevante.

1.1. Definición

ESTADISTICA. Es el conjunto de métodos y procedimientos que implican: Recopilación, presentación, organización y análisis de datos, para que a partir de estos datos se pueda inferir y llegar a concluciones válidas y tomar decisiones razonables.

PROBABILIDAD. Parte de las matemáticas que estudia fenómenos o experimentos aleatorios.

1.2. Su uso

Se puede predecir por ejemplo el resultado de un experimento tan simple como el lanzamiento de una moneda.

1.3. Ventajas y desventajas

La probabilidad se usa en la toma de decisiones, por ejemplo: Una compañía produce un detergente líquido que se envasa en botellas de 500ml, las cuales son llenadas por una máquina. Las botellas con volumen mayor que 500ml representa pérdidas para la compañía, y las botellas con volumen menor que 500ml representan pérdidas para el consumidor. La compañía necesita mantener 500ml de volumen neto. Para mantener el control, la compañía hace el esquema de muestreo: Selecciona 10 botellas 4 veces al dá, y se determina su contenido neto promedio, si está cercano a 498ml y 502ml el proceso está bajo control, de otra manera, se encontrará fuera de control (en este caso se detendrá el llenado analizando cual es la causa del problema). La evaluación de estos riesgos puede hacerse utilizando probabilidad.

2. PROBABILIDAD

2.1. Teoría de conjuntos

Experimentos aleatorios: Es aquel experimento que cuando se repite bajo las mismas condiciones iniciales, el resultado que se obtiene no siempre es el mismo.
Ejemplos: extracción de una carta de una baraja de 52, lanzar una moneda, tirar un dado.

En principio no sabemos cuál será el resultado del experimento aleatorio, así que conviene agrupar todos los resultados posibles.

Espacio muestral o espacio muestra Ω: Es el conjunto de todos los posibles resultados del experimento, se le demota como Ω (Omega). En otros textos se usa también **S (sampling space o espacio muestral).**

Si el espacio miestral tiene un número finito de elementos o finito numerable, entonces se dice que este es discreto. Si el espacio muestal tiene como elementos todos los puntos de un intervalo real, entonces se dice que es continuo.
Ejemplos:

- a) $\Omega = \{$ lanzamiento de un dado $\}$
  $= \{1, 2, 3, 4, 5\}$
- b) $\Omega = \{$ tiempo de duración de un foco $\}$
  $= \{t, t \geq 0\}$

Evento o suceso: Es cualquier subconjunto del espacio muestral y lo denotaremos por las primeras letras del alfabeto en mayúsculas.

Ejemplos:

- a) $\Omega = \{$ Obtener un número impar al lanzar un dado. $\}$
  $= \{1, 3, 5\}$
- b) $\Omega = \{$ Obtener al menos una cara al lanzar una moneda 2 veces $\}$
  $= \{cx, xc, cc\}$

Si por ejemplo sale el número 3, se observa el ocurrimiento del evento $A = \{1, 3, 5\}$.
Como los **EVENTOS** son subconjuntos del espacio muestral Ω, entonces es posible aplicar la teoría de conjuntos. Algunas de estas propiedades serán de gran utilidad en el estudio de la PROBABILIDAD Y LA ESTADÍSTICA.

2.2. Conjuntos

Suponemos que el espacio muestral Ω de un experimento aleatorio es nuestr conjunto universal, y cualquier elemento de Ω lo denotamos por ω.

El conjunto vacío es $\emptyset = \{,\}$. Otros símbolos comunes son: $\notin, \in, \subseteq, \subseteq, \nsubseteq$
Si **A** y **B** son dos subconjunto cualquiera de Ω, y usamos las operaciones básicas de conjuntos: unión (A $\cup$ B), intersección (A $\cap$ B), diferencia (A$-$B) y complemento (A^c).
Ejemplo

$A \cup B = \{\omega \in \Omega : \omega \in A \quad o \quad \omega \in B.\}$
$A \cap B = \{\omega \in \Omega : \omega \in A \quad y \quad \omega \in B.\}$
$A - B = \{\omega \in \Omega : \omega \in A \quad y \quad \omega \notin B.\}$
$A^c = \{\omega \in \Omega : \omega \notin A.\}$

Ejemplos de experimentos aleatorios

- El lanzamiento de 3 monedas hasta obtener 2 águilas.
- Lanzamiento de una moneda 3 veces hasta obtener 2 águilas.
- Lanzamiento de una moneda 3 veces y contar la cantidad de soles que aparecen en los lanzamientos.
- Lanzamiento de un dado observando el número de la cara superior.
- Lanzamiento de 2 dados y la realización del conteo de la suma de las caras.
- Sea un lote de 60 artículos que tiene 10 defectuosos. Entonces, se define el proceso de seleccionar 1 artículo sin remplazo y anotar los resultados hasta obtener el último defectuoso.
- Observar las cantidades mínimas y máximas de personas que llegan a la terminal de autobuses cada día y cada 5 minutos.
- El conocimiento del curso de una acción referente a una empresa en la bolsa de valores es uno de los principales problemas que todo accionista quisiera saber cómo predecir. Este es un problema financiero muy complejo que depende de muchos factores, incluyendo los políticos, por lo que no se puede controlar el curso de la acción ya que esta se encuentra envuelta en mucha incertidumbre; por tanto, solo es posible indicar un rango de valores posibles en el que se tengan evidencias que podrán encon- trarse en el curso de la bolsa para dicha acción. En el caso del dólar podríamos tener evidencias de que al día siguiente su costo estará entre 12,40 y 12,80 pesos, pero en realidad no conocemos cuál será su cotización exacta, puesto que este estará influido por factores que pueden tener mucha incertidumbre, como situaciones políticas.

2.3. EJEMPLOS: ESCRIBE EL ESPACIO MUESTRAL EN CADA CASO

- El experimento sobre el lanzamiento de una moneda se realiza 3 veces. Anota el resultado. (NOTA: s es sol, y c es cara)
  Solución:
  $S = \{sss, ssc, scs, css, scc, csc, ccs, ccc\}$
- Lanzamiento de una moneda 3 veces y se anota la cantidad de águilas que aparecen. De ese modo, 0 representa la ausencia de águilas.
  Solución:
  $S = \{0, 1, 2, 3\}$
- Lanzamiento de 2 dados y se anota la suma de las caras superiores.
  Solución:
  $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
- Lanzamiento de 2 dados y se anota la diferencia del valor mayor menos el menor de sus caras superiores.
  Solución:
  $S = \{0, 1, 2, 3, 4, 5\}$
- Lanzamiento de 1 dado 2 veces, de las cuales se toma la diferencia del valor del primer resultado menos el valor del segundo número de la cara superior de otro dado.
  Solución:
  $S = \{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$

- El administrador de una red logística de autobuses debe tomar la decisión de cómo ordenar la distribución de dos de 3 autobuses con el fin de que viajen a otra ciudad en 2 días sucesivos (sin repetir un autobús dos días seguidos). Represente con a_1, a_2 y a_3 los tres autobuses. Ordene los viajes de tal forma que a_1a_3 significa que el autobús a_1 viaja a la otra ciudad el primer día y el autobús a_3 el segundo día. Establezca los puntos muestrales de este experimento.
  Solución:
  $S = \{a_1a_2, a_1a_3, a_2a_1, a_3a_2, a_2a_3, a_3a_1\}$

- El administrador de una red logística de autobuses tiene que tomar la decisión de cómo distribuir dos de tres autobuses para viajar a otra ciudad. Represente con a_1, a_2 y a_3 a los tres autobuses y describa el espacio muestral del experimento: Seleccionar dos autobuses para viajar a la otra ciudad".
  Solución:
  $S = \{a_1a_2, a_1a_3, a_2a_3\}$

- Una agencia comercial compra papelería a uno de tres vendedores V_1, V_2 y V_3 . El pedido se ordena en dos días sucesivos (sin repetir vendedor dos días), un pedido por día, tal que V_1V_3, lo que significa que el vendedor V_1 recibe el pedido el primer día y el vendedor V_3 lo recibe el segundo día. Establezca los puntos muestrales de este experimento.
  Solución:
  $S = \{V_1V_2, V_1V_3, V_2V_3\}$

2.4. ESCRIBE EL EVENTO EN CADA CASO

- Se lanza una moneda 3 veces y se anotan los resultados posibles. Sea el evento E={ aparece solo 1 águila (c). }
  Solución:
  $E = \{ssc, scs, css\}$

- El lanzamiento de una moneda 3 veces, escribe el evento como el conteo de águilas que aparecen.
  Solución:
  $E = \{1\}$

- El lanzamiento de 1 dado. Escribe el evento que denota el número de la cara del dado que "no" es mayor que 4.
  Solución:
  $E = \{1, 2, 3, 4\}$

- El lanzamiento de 1 dado. Escribe el evento que denota el número de la cara del dado que es mayor que 4.
  Solución:
  $E = \{5, 6\}$

2.5. Modelos determinísticos y probabilísticos

Uno de los objetivos del estudio de la probabilidad es desarrollar estructuras conceptuales que permitan comprender los fenómenos que ocurren en procesos aleatorios para poder predecir los efectos que de ellos se deriven. De la experiencia, se deduce fácilmente que para poder estudiar un fenómeno es necesaria su imitación o reproducción en una cantidad suficiente, a fin de que su investigación sea lo más precisa posible.

Esta necesidad es lo que da origen a los **modelos**. Ahora bien, ¿qué entendemos por **modelo** y qué lo origina? Por modelo, entenderemos **la representación o reproducción de los fenómenos** [1]**.**
Los modelos pueden ser de diferentes tipos, pero solo veremos en clase áquellos que son *"modelos matemáticos"*. Veamos a continuación la definición de *modelo matemático* que se utiliza durante todo el curso. Un *modelo matemático* es una representación simbólica de un fenómeno cualquiera, realizada con el fin de estudiarlo mejor, dichas representaciones puede ser fenómenos físicos, económicos, sociales, etcétera.
Los modelos matemáticos pueden clasificarse en determinísticos y probabilísticos, y para poderlos diferenciar es necesario conocer su definición y algunos ejemplos. Primero, presentamos la definición de **modelos determinísticos.** Cuando se realiza el modelo matemático de un fenómeno y en este se pueden manejar los factores que intervienen en su estudio con el propósito de predecir sus resultados, se llamará modelo determinístico. A continuación se presentan algunos ejemplos de modelos determinísticos:

2.5.1. Ejemplos de Modelos determinísticos

- El lanzamiento de una moneda con ambos lados iguales (p. ej., águilas). Al plantear este modelo es posible determinar que siempre es posible predecir el resultado (suponiendo que la moneda no puede quedar en posición vertical), puesto que solo hay una opción: águila.

- Cuando tenemos una inversión **c** a una tasa **r**, podemos calcular su Valor Presente Neto, **VPN** (**c**). El modelo es determinístico, puesto que tiene una inversión fija **c** a una tasa fija **r**; por tanto, es posible predecir el resultado que ocurrirá al cabo de n años mediante el uso de la siguiente fórmula:

$$VPN(c) = \frac{c}{(1+r)^n} \tag{1}$$

  Por ejemplo, si vamos a recibir c = 150 000 pesos dentro de cuatro años, pero queremos saber cuánto vale hoy, VPN(c), debemos descontar los intereses que se generan desde hoy hasta dentro de cuatro años. Si el interés anual es de 8 % en operaciones a cuatro años, entonces el día de hoy debemos invertir:

$$VPN(c) = \frac{150000}{(1+0{,}08)^4} = 110254{,}50 \tag{2}$$

  Entonces, si iniciamos la inversión con 110 254.50 al cabo de cuatro años tendremos 150 000 pesos.

- Sea una economía en equilibrio determinada por el modelo económico de entradas y salidas de **Wassiley Leontief**, aplicado a tres empresas distintas.

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + e_1 &= x_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + e_2 &= x_2 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + e_3 &= x_3
\end{aligned}$$

[1] En general, sabemos que cualquier modelo físico es una aproximación de la realidad; no obstante, este no la puede representar en forma exacta, esto se debe a que en cada fenómeno intervienen infinidad de factores y no es posible involucrarlos a todos en el modelo. Por esta razón, salvo que se diga otra cosa, en el texto se consideran los modelos conocidos sobre diferentes fenómenos físicos como determinísticos. Por ejemplo, en el circuito anterior, la caída real de voltaje está influenciada por los factores: humedad, calentamiento del alambre conductor, temperatura, etcétera, que para fines prácticos se pueden considerar despreciables. De manera similar, en el modelo del movimiento de un cuerpo sobre una superficie con fricción, se considera que las otras fuerzas que intervienen son despreciables.

Donde: x_i representa la producción de la empresa i; y e_i representa la demanda externa sobre la empresa i; y a_ij representa el número de unidades de producto de la empresa i necesarias para producir una unidad de producto de la industria j. Conociendo la demanda externa por empresa y la demanda interna entre empresas, con este modelo es posible predecir la producción de cada empresa.

- El modelo de una compañía donde se elaboran dos productos al pasar en forma consecutiva, a través de una línea de producción, por tres máquinas distintas. En este caso, el tiempo por máquina asignado a los dos productos está limitado por una cantidad determinada de horas por día; el tiempo de producción y la ganancia por artículo de cada producto se pueden establecer de manera que al combinar los productos podemos obtener una ganancia óptima.
  En el modelo anterior se puede notar que estamos controlando los diferentes parámetros que intervienen. Por tanto, al establecer el modelo matemático correspondiente y los valores para los factores es posible predecir su resultado.

- Se puede diseñar un modelo que muestre la influencia de la fuerza de fricción sobre un cuerpo que se mueve en una superficie. Con este modelo se puede concluir que en superficies más ásperas se tiene mayor fuerza de rozamiento. El modelo anterior es determinístico, ya que en este se manejan superficies y se puede predecir el resultado. Por ejemplo, la distancia a la que se puede detener el móvil; esto es, si se mueve el cuerpo con una fuerza inicial y cambiamos las asperezas podremos establecer una fórmulamatemática que indique (como resultado de un cálculo numérico) la distancia en la que se detendrá el móvil.

2.6. Modelos probabilísticos

El otro tipo de modelos ocurre cuando **no podemos controlar los factores que intervienen en dichos modelos**. A partir de lo cual surge la definición de **modelo probabilístico o estocástico.**

Los modelos probabilísticos o **modelos estocásticos son aquellos modelos matemáticos de los fenómenos en los cuales no se pueden controlar los factores que intervienen en su estudio**, además de que dichos factores ocurren de tal manera que no es posible predecir sus resultados.

Los modelos probabilísticos son de gran interés; por tanto, para una mejor comprensión de estos se presentan los siguientes ejemplos:

- Los modelos clásicos probabilísticos se refieren a los juegos de azar, como el lanzamiento de una moneda equilibrada (es decir, que no está cargada a ningún lado), para determinar el resultado que va a ocurrir. En el lanzamiento de un dado no equilibrado (esto es, que un lado del dado no pesa más que los otros) no es posible predecir qué número quedará en la parte de arriba del dado

- En el lanzamiento de una moneda equilibrada 10 veces para obtener cinco águilas, el modelo es de tipo probabilístico, puesto que no podemos predecir el resultado que va a ocurrir en el siguiente lanzamiento.

- Las cartas o fichas que le tocarán a una persona al inicio de una partida de un juego de cartas o domino, respectivamente.

- En el ejemplo 2 de los ejemplos de modelos determinísticos, la tasa anual de inversión para un año determinado en realidad está condicionada a situaciones de incertidumbre del país; por consiguiente, bajo estas condiciones no podemos predecir el VPN para un año determinado si no conocemos con anterioridad la tasa r.

- En una línea de producción, al realizar el control de calidad de los artículos se detecta cierta cantidad de productos defectuosos; no es posible determinar la cantidad o porcentaje de estos en la línea.

- Si deseamos conocer los ingresos por acción para una compañía de teléfonos, estos se pueden estimar mediante el PIB (Producto Interno Bruto) que se mide en millones de pesos. Entonces, establecemos, mediante una ecuación, un modelo para su estimación, pero no podemos saber con exactitud sus resultados.

- El conocimiento del curso de una acción referente a una empresa en la bolsa de valores es uno de los principales problemas que todo accionista quisiera saber cómo predecir. Este es un problema financiero muy complejo que depende de muchos factores, incluyendo los políticos, por lo que no se puede controlar el curso de la acción ya que esta se encuentra envuelta en mucha incertidumbre; por tanto, solo es posible indicar un rango de valores posibles en el que se tengan evidencias que podrán encontrarse en el curso de la bolsa para dicha acción. En el caso del dólar podríamos tener evidencias de que al día siguiente su costo estará entre 20 y 21 pesos, pero en realidad no conocemos cuál será su cotización exacta, puesto que este estará influido por factores que pueden tener mucha incertidumbre, como situaciones políticas.

- Si deseamos conocer el lugar de caída de un satélite que se salió de su órbita y se dirige a la Tierra no podemos predecir el lugar donde caerá, puesto que no es posible controlar su movimiento; por tanto, solo es posible indicar una región en donde se cree que caerá, con un valor numérico que represente la aseveración.

Al reproducir cualquier fenómeno, ya sea de manera determinística o probabilística, estamos experimentando, por lo que es necesario aclarar lo siguiente: qué entenderemos por experimento al utilizar un modelo matemático de tipo probabilístico.

Llamaremos experimento aleatorio al proceso de obtención de una observación en que se cumple alguna de las siguientes condiciones:

- a) Todos los resultados posibles son conocidos.

- b) Antes de realizar el experimento el resultado es desconocido.

- c) Es posible repetir el experimento en condiciones ideales.

2.7. Experimentos determinísticos

Además de los experimentos aleatorios, también tenemos los determinísticos, de los cuales enseguida se presenta su definición y se muestran algunos ejemplos para su mejor comprensión [2].
Ejemplos:

- En un modelo de valor presente neto de una serie de flujos de efectivo V_i de una inversión **c**, a una tasa fija **r**, podemos calcular su **VPN(c)** al cabo de **n** años, el cual se calcula de la siguiente forma:

[2] Al proceso por el cual se describen los fenómenos cuyos resultados pueden predecirse, lo llamaremos experimento determinístico.

$$VPN(c) = - \quad \text{inversión} + \sum_{i}^{n} \frac{V_i}{1+r^i} \tag{3}$$

- La mezcla de sustancias químicas para la obtención de algún compuesto.

3. Interpretaciones de la probabilidad

En la actualidad, la palabra probabilidad es empleada con demasiada frecuencia por las personas; por ejemplo, en expresiones como: "Es probable que hoy estudie para probabilidad y estadística"; El equipo mexicano de fútbol está jugando mal, y es muy probable que en su siguiente partido pierda; El cielo está bastante despejado; por tanto, no hay muchas posibilidades de que llueva; entre otras. Como se puede notar en las expresiones anteriores, las palabras relacionadas con la probabilidad tienen la característica de basarse en **sucesos que pueden ser verdaderos**, además de que a causa de los hechos observados (resultados preliminares, tiempo, etcétera), se puede hablar de la posibilidad de su ocurrencia.

A pesar de los esfuerzos realizados por muchos matemáticos para asignar de forma única la probabilidad a un suceso, todo ha sido en vano, pues desde los inicios de su estudio hasta nuestros días no existe una forma única de asignación de probabilidades. Solo contamos con diferentes corrientes de probabilidad, las cuales se aplican para asignar un valor numérico a la posibilidad de la ocurrencia de algún suceso probabilístico. De hecho, el verdadero significado de la probabilidad aún se considera conflictivo; por tanto, primero trataremos las dos corrientes más comunes de la probabilidad (Corriente clásica (a priori, Corriente bayesiana (a posteriori).

3.1. Corriente clásica (a priori)

En la corriente clásica se consideran espacios muestrales uniformes, es decir, se asignan probabilidades a eventos con base en resultados equiprobables (igualmente verosímiles). Esto es, los clasistas asignan la misma probabilidad a cada punto del espacio muestral (es decir: $\frac{1}{n}$, donde n es la cantidad de elementos del espacio muestral); posteriormente, para obtener la probabilidad de la ocurrencia de un evento E, se suma la cantidad de elementos de E y se multiplica por la probabilidad de un elemento del espacio muestral $\frac{1}{n}$(o se divide entre n).

Cabe apuntar que de lo anterior se deduce que la probabilidad de los puntos muestrales se establece a priori; es decir, antes de cualquier experimento. Resolviendo el ejemplo anterior en la forma clásica tendremos, lo siguiente: Se lanza una moneda equilibrada tres veces y se anotan los resultados posibles que aparecen; sea el evento E: "obtención de dos soles en los tres lanzamiento", la pregunta es: ¿cuál es la probabilidad de que ocurra el evento E? Para responder a la pregunta, primero obtenemos el espacio muestral desde el punto de vista clásico; de este modo, representando águila por a y sol por s, tendremos:
$S = \{sss, ssa, sas, ass, saa, asa, aas, aaa\}$.
En estos casos, ssa representa que los primeros dos lanzamientos resultaron soles y el tercer lanzamiento águila. Considerando 1 que cada punto del espacio muestral es equiprobable con probabilidad de ocurrencia $\frac{1}{8}$, tendremos que la probabilidad del evento E (resulten dos soles en los tres lanzamientos) se resuelve al conocer la cantidad de elementos del evento:
$E = \{ssa, sas, ass\}$
En este caso, como E contiene tres elementos tenemos que la probabilidad de que ocurra el evento E es:
$P(E) = \frac{3}{8}$.

3.2. Corriente bayesiana (a posteriori)

En la corriente bayesiana se asignan probabilidades a los eventos después del experimento. Es decir, la asignación de probabilidades está basada en el conocimiento de la ocurrencia de eventos que estén en dependencia con el evento de estudio. Por ejemplo, si queremos asignar una probabilidad al evento de que el día 11 de septiembre llueva y tenemos la siguiente información:

- a) Los días 9 y 10 de septiembre no llovió.
- b) Los días 9 y 10 de septiembre llegó un huracán a 400 kilómetros de distancia y llovió ambos días.

Es obvio suponer que la asignación de probabilidades en ambos casos es muy diferente, ya que tenemos información que hace cambiar nuestra asignación de probabilidades. En tal situación decimos que la información obtenida influyó en la asignación de probabilidades. Otro ejemplo, es el caso anterior cuando se lanza una moneda equilibrada tres veces y se cuenta la cantidad de soles que aparecen, el evento E: "obtención de dos soles en los tres lanzamientos"; la pregunta es: ¿cuál es la probabilidad de que ocurra el evento E?, si se sabe que el primer lanzamiento resultó sol.
Esta corriente de probabilidad es la base motora de la teoría de decisiones, puesto que cualquier toma de decisiones está influida por todo tipo de información que se pueda tener sobre un fenómeno en estudio. El uso de esta corriente es posible en la parte de decisiones llamada árboles de decisión.

4. EJERCICIOS RESUELTOS

4.1. Problema 1

Un juego consiste en extraer de manera aleatoria dos pelotas al mismo tiempo de una urna que contiene cinco pelotas numeradas de 1 a 5, de igual forma y tamaño. La persona gana si las dos pelotas extraídas tienen número par, en otro caso la persona pierde. Calcule la probabilidad de que la persona gane.

Solución

Se extraen dos pelotas aleatoriamente de un total de 5, numeradas del 1 al 5. El espacio muestral resulta:
$E = \{1-2, 1-3, 1-4, 1-5, 2-3, 2-4, 2-5, 3-4, 3-5, 4-5\}$.
El evento E lo definimos, como: $E = \{$ las dos pelotas extraídas tienen número par $\} = \{2-4\}$
Por tanto, del espacio muestral encontrado, y considerando a los puntos muestrales equiprobables tenemos que la probabilidad del evento E estará dada por: $P(E) = \frac{1}{10}$

4.2. Problema 2

En una cadena de televisión se hizo una encuesta a 2 500 personas para saber la audiencia de un debate y de una película que se emitieron en horas distintas: 2 100 vieron la película, 1 500 vieron el debate y 350 no vieron ninguno de los dos programas. Si elegimos al azar a uno de los encuestados:

- a) ¿Cuál es la probabilidad de que viera la película y el debate?
- b) ¿Cuál es la probabilidad de que viera la película, sabiendo que vio el debate?
- c) Sabiendo que vio la película, ¿cuál es la probabilidad de que viera el debate?

Organizamos la información en una tabla de doble entrada, completando los datos que faltan:

Solución:

Sea $D = \{$ vió el debate$\}$ y $P = \{$ vió la película$\}$

- a) $P(D \cap P) = \frac{1450}{2500} = \frac{29}{50} = 0{,}58$
- b) $P(P \mid D) = \frac{1450}{1500} = \frac{29}{30} = 0{,}97$
- c) $P(D \mid P) = \frac{1450}{2100} = \frac{29}{42} = 0{,}69$

	DEBATE	NO DEBATE	
PELÍCULA	1450	650	2100
NO PELÍCULA	50	350	400
	1500	1000	2500

5. Distribuciones discretas de probabilidad

Una v.a discreta asume cada uno de sus valores con una cierta probabilidad. Con mucha frecuencia es conveniente representar con una fórmula todas las probabilidades de una v.a X. Dicha fórmula, necesariamente, debe ser función de los valores numéricos x, y que se representa por $f(x), g(x), h(x)$, etc. Por lo tanto, $f(x) = P(X = x)$. Al conjunto de pares ordenados $(x, f(x))$ se le llama **función de probabilidad o distribución de probabilidad** de la v.a discreta X.

Concepto: El conjunto de pares ordenados $(x, f(x))$ es una **función de probabilidad o distribución de probabilidad** e la v.a discreta X si satisface las siguientes condiciones:

$$f(x) \geq 0, \quad \text{para x un número real} , \tag{4}$$

$$\sum_x f(x) = 1, \tag{5}$$

$$P(X = x) = f(x). \tag{6}$$

5.1. Ejemplos:

1)

Una moneda se lanza dos veces, entonces $\Omega = \{cc, cs, sc, ss\}$. Sea X la v.a. que consiste en observar el número de caras.
Solución:

Espacio muestral	x
cc	2
cs	1
sc	1
ss	0

La función de probabilidad es:

$f(x = 0) = P(x = 0) = \frac{1}{4}$, $f(x = 1) = P(x = 1) = \frac{2}{4}$, $f(x = 2) = P(x = 2) = \frac{1}{4}$

x	$P(x)$
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

2)

Se tiene un lote de 25 artículos de los cuales 5 son defectuosos se eligen 4 al azar. Sea Y la v.a que representa el número de artículos defectuosos encontrados. Obtener la distribución de probabilidades de la v.a Y si los artículos se eligen sin sustitución.

Solución:

Sea D =artículo defectuoso, por lo tanto, D^c =artículo no defectuoso.

$P(D) = \frac{5}{25}$ $\qquad$ $P(D^c) = \frac{20}{25}$

$$f(0) = P(y=0) = P(D_1^c \cap D_2^c \cap D_3^c \cap D_4^c) = \frac{20}{25}\frac{19}{24}\frac{18}{23}\frac{17}{22} = \frac{4845}{12650} \tag{7}$$

$$f(1) = P(y=1) = 4 \cdot P(D_1 \cap D_2^c \cap D_3^c \cap D_4^c) = 4 \cdot \frac{5}{25}\frac{20}{24}\frac{19}{23}\frac{18}{22} = \frac{5700}{12650} \tag{8}$$

$$f(2) = P(y=2) = 6 \cdot P(D_1 \cap D_2 \cap D_3^c \cap D_4^c) = 6 \cdot \frac{5}{25}\frac{4}{24}\frac{20}{23}\frac{19}{22} = \frac{1900}{12650} \tag{9}$$

$$f(3) = P(y=3) = 4 \cdot P(D_1 \cap D_2 \cap D_3 \cap D_4^c) = 4 \cdot \frac{5}{25}\frac{4}{24}\frac{3}{23}\frac{20}{22} = \frac{200}{12650} \tag{10}$$

$$f(4) = P(y=4) = P(D_1 \cap D_2 \cap D_3 \cap D_4) = \frac{5}{25}\frac{4}{24}\frac{3}{23}\frac{2}{22} = \frac{5}{12650} \tag{11}$$

(12)

y	$P(y)$
0	$\frac{4845}{12650}$
1	$\frac{5700}{12650}$
2	$\frac{1900}{12650}$
3	$\frac{200}{12650}$
4	$\frac{5}{12650}$

3)

Una caja que contiene 4 monedas de 100 y 2 de 50 , se seleccionan tres de ellas al azar sin reemplazo. Determine la distribución de probabilidad para el total de las tres monedas.

Solución:

Sea X = Dinero total de las 3 monedas, $\qquad$ A =moneda de 100 pesos, $\qquad$ B =moneda de 50 pesos.

Así, la distribución de probabilidad se escribe como la segunda tabla:

monedas	x
AAA	300
AAB	250
ABA	250
BAA	250
ABB	200
BAB	200
BBA	200
BBB	150

x	$P(x)$
150	$\frac{1}{8}$
200	$\frac{3}{8}$
250	$\frac{3}{8}$
300	$\frac{1}{8}$

4)

De una caja que contiene 4 pelotas negras y 2 verdes, se seleccionan 3 de ellas en sucesión con reemplazo. Encuentre la distribución de probabilidad para el número de pelotas verdes.

Solución:

Sea $X =$ múmero de pelotas verdes, $V =$pelota verde, $N =$pelota negra.
Así, la distribución de probabilidad se escribe como la segunda tabla:

pelotas	x
VVV	3
VVN	2
VNV	2
NVV	2
VNN	1
NVN	1
NNV	1
NNN	0

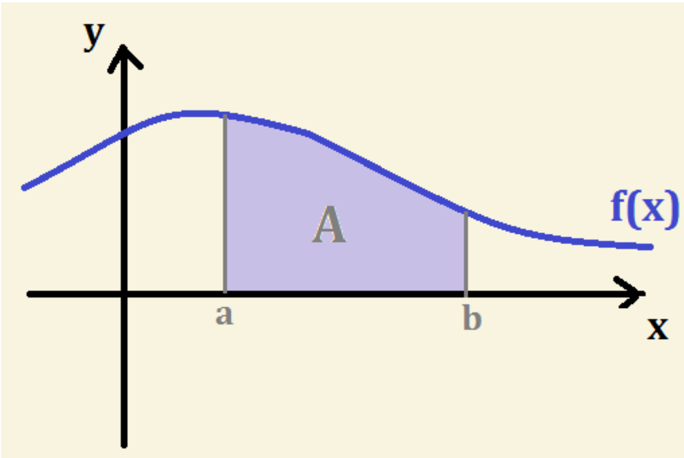
x	$P(x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

6. Distribuciones continuas de probabilidad

Una v.a continua tiene probabilidad cero de asumir cualquiera de sus valores. Luego, su distribución de probabilidad no puede darse en forma tabular. Como una distribución de probabilidad de una v.a continua no puede presentarse en forma tabular, si puede tener una fórmula. Esta fórmula es una función, es decir, $f(x)$ y para este tipo de variables se llama función de densidad de probabilidad o función de densidad.

Las áreas bajo la curva representarán las probabilidades, por lo tanto, el gráfico de la función de densidad se ubica siempre sobre el eje x. Una función de densidad se construye de tal forma que el área comprendida bajo la curva es siempre igual a uno, cuando se calcula sobre todo el recorrido de la v.a X.

Así, $P(a < x < b) = \int_a^b f(x)dx$



Concepto : La función $f(x)$ es una **función de densidad de probabilidad** para la v.a continua X, definida en el conjunto de los números reales, si:

$$f(x) \geq 0, \quad \text{para x un número real ,} \tag{13}$$

$$\int_{-\infty}^{\infty} f(x)dx) = 1, \tag{14}$$

$$P(a < x < b) = \int_a^b f(x)dx \tag{15}$$

7. Esperanza o valor esperado

El valor esperado se usa como una medida de centro de una distribución de probabilidad de una v.a.
Sea X una v.a con función de probabilidad o función de densidad $f(x)$. El valor esperado de X, simbolizado por $E(X)$ es:

$$E(X) = \begin{cases} \sum_x x \cdot f(x), \text{ si } X \text{ es una v.a. discreta} \\ \int_{\infty}^{\infty} x \cdot f(x), \text{ si } X \text{ es una v.a. continua} \end{cases} \tag{16}$$

Si $g(x) = (X - \mu)^2$, donde $\mu = E(X)$, entonces $E(g(X))$ se le llama la **varianza** de la v.a. X y se simboliza como σ^2:

$$\sigma^2 = E(g(X)) = \begin{cases} \sum_x (X - \mu)^2 \cdot f(x), \text{ si } X \text{ es una v.a. discreta} \\ \int_{\infty}^{\infty} (X - \mu)^2 \cdot f(x), \text{ si } X \text{ es una v.a. continua} \end{cases} \tag{17}$$

donde, $\sigma = \sqrt{Var(X)}$ se conoce como **desviación estándar**. σ mide la dispersión de los valores de la v.a. X con respecto a la media μ.

7.1. Propiedades de la esperanza

Sea X una v.a., entonces,

- 1) $E(c) = c$ (c es una constante, que puede ser un número.)

- 2) $E(c \cdot X) = c \cdot E(X)$
- 3) $\sigma^2 = E(X^2) - E(X)^2$
- 4) $Var(X) = \sigma^2 = E(X^2) - E(X)^2$

7.2. Propiedades de la Varianza

Sea X una v.a., entonces,

- 1) $Var(c) = 0$
- 2) $Var(c \cdot X) = c^2 \cdot Var(X)$
- 3) $Var(X + a) = Var(X)$

7.2.1. Ejemplos

1)

Se lanza una moneda tres veces, si las tres veces aparece cara o parece cruz un jugador gana 5 pesos, pero si no es así pierde 3 pesos. ¿Cuál es la esperanza de este juego?

Solución

Sea X la v.a que denota ganancia o pérdida. Y el espacio muestra es:

$\Omega = \{(ccc), (cxc), (ccx), (xcc), (xxc), (xcx), (cxx), (xxx)\}$

x	$P(x)$
5	$\frac{2}{8}$
-3	$\frac{6}{8}$

$E(X) = 5\left(\frac{1}{4}\right) - 3\left(\frac{3}{4}\right) = -1$

Es decir, el jugador pierde en promedio 1 peso por lanzamiento de las tres monedas.

2)

Las ventas por hora de una máquina automática puede ser 20 , 21 o 22 cajetillas de cigarros con probabilidad 0,3, 0,5 y 0,2 respectivamente. ¿Cuál es la venta esperada por hora para esta máquina? ¿Cuál es la varianza de ventas por hora?

Solución

Si X = ventas de cigarros por hora.

$E(X) = (20)(0{,}3) + (21)(0{,}5) + (22)(0{,}2) = 20{,}9$

Así la venta esperada por hora resulta de 20,9 cajetillas.

$Var(X) = E(X^2) - E(X)^2$

x	$P(x)$
20	0,3
21	0,5
22	0,2

$E(X)^2 = (400)(0{,}3) + (441)(0{,}5) + (484)(0{,}2) = 437{,}3$

$Var(X) = 437 - 436 = 0{,}49$
Resulta que la varianza de ventas por hora respecto al promedio de 20,9 decigarros vendidos al día es 0,49.

3)

El número de autos X que pasan a través de una lavadora entre las $4:00$ pm y las $5:00$ pm de los viernes tiene la siguiente distribución de probabilidad:

x	$P(x)$
4	$\frac{1}{12}$
5	$\frac{1}{12}$
6	$\frac{1}{4}$
7	$\frac{1}{4}$
8	$\frac{1}{6}$
9	$\frac{1}{6}$

Solución

Sea $g(X) = 2X - 1$ que que representa la cantidad de dinero, en dólares, que el gerente del negocio le paga al encargado. Encuentre las ganancias esperadas del encargado en este período en particular.

$$E(X) = 4\left(\frac{1}{12}\right) + 5\left(\frac{1}{12}\right) + 6\left(\frac{1}{4}\right) + 7\left(\frac{1}{4}\right) + 8\left(\frac{1}{6}\right) + 9\left(\frac{1}{6}\right) \tag{18}$$

$$E(X) = \frac{41}{6} \tag{19}$$

$$E(2X-1) = 2 \cdot E(X) - 1 \tag{20}$$

$$E(2X-1) = 2 \cdot \frac{41}{6} - 1 \tag{21}$$

$$E(2X-1) = \frac{38}{3} \tag{22}$$

La ganancia esperada del encargado es 12,67 dólares estre las $4:00$ pm. y las $5:00$ pm.

4)

Por invertir en unas acciones en particular, una persona puede obtener ganancias de \$4000 con una probabilidad de 0,3, o una pérdida de \$1,000 con una probabilidad de 0,7. ¿Cuál es la ganancia que espera esta persona?.
Solución

$\$$	$P(\$)$
4000	0,3
-1000	0,7

$E(\$) = 4000(0{,}3) - 1000(0{,}7) = 1200 - 700 = 500,$
entonces, la ganancía esperada es de \$500.

5)

Un distribuidor de joyas antiguas está interesado en comprar un collar de oro para el cual las probabilidades son 0,22, 0,36, 0,28 y 0,14 respectivamente, de que la poseedora estaría dispuesta a venderla en \$250,000, en \$150,000, al costo \$100,000 o con una pérdida de \$150,000. ¿Cuál es la utilidad que ella espera?.

Solución

$\boxed{X = \text{es la v.a. que indica el precio}}$

X	$P(X)$
250,000	0,22
150,000	0,36
100,000	0,28
-150,000	0,14

$E(X) = 250,000\Big(0{,}22\Big) + 150,000\Big(0{,}36\Big) + 100,000\Big(0{,}28\Big) - 150,000\Big(0{,}14\Big) = 116,000,$
entonces, la ganancía esperada es de \$500.

6)

Calcule la varianza de $h(X) = 2X + 3$ donde X es una v.a. con una distribución de probabilidad:

X	$P(X = x)$
0	$\frac{1}{4}$
1	$\frac{1}{8}$
2	$\frac{1}{2}$
3	$\frac{1}{8}$

$E(X) = 0\left(\frac{1}{4}\right) + 1\left(\frac{1}{8}\right) + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{8}\right) = \frac{3}{2}$

$Var(X) = E(X^2) - E(X)^2$
$E(X^2) = 0^2\left(\frac{1}{4}\right) + 1^2\left(\frac{1}{8}\right) + 2^2\left(\frac{1}{2}\right) + 3^2\left(\frac{1}{8}\right) = \frac{13}{4}.$

7)

Supóngase usted juega con un dado. Un jugador gana \$20 si obtiene 2, \$40 si obtiene 4, y pierde \$30 si obtiene 6, en tanto que ni pierde ni gana si obtiene otro resultado. Hallar la suma esperada de dinero ganado.

Solución

Sea X = la v.a. que representa el dinero ganado.

Ω	x	$P(X=x)$
1	0	$\frac{1}{6}$
2	20	$\frac{1}{6}$
3	0	$\frac{1}{6}$
4	40	$\frac{1}{6}$
5	0	$\frac{1}{6}$
6	-30	$\frac{1}{6}$

$$E(X) = 0\left(\frac{1}{6}\right) + 20\left(\frac{1}{6}\right) + 0\left(\frac{1}{6}\right) + 40\left(\frac{1}{6}\right) + 0\left(\frac{1}{6}\right) - 30\left(\frac{1}{6}\right) = 5$$

8)

En una lotería hay 200 premios de \$150, 20 premios de \$750 y 5 premios de \$3000 Suponer que se colocan a la venta $10,000$ boletos. ¿Cuál es el precio justo que se debe pagar por un boleto?
Solución

Sea X = la v.a. que representa el dinero ganado.

Ω	x	$P(X=x)$
10,000 - 225	0	0,9775
200	150	0,02
20	750	0,002
5	3000	0,0005

$$E(X) = 0\Big(0{,}9775\Big) + 150\Big(0{,}02\Big) + 750\Big(0{,}002\Big) + 3000\Big(0{,}0005\Big) = 6$$

El precio justo a pagar por un boleto es \$6. Sin embargo una lotería gana dinero, el precio del boleto sería más alto.

8. Distribuciones discretas de probabilidad

8.1. Distribuciones de Bernoulli

El experimento más sencillo es aquel que puede resultar en uno de dos resultados posibles.
Ejemplo

- a) sexo de un niño al nacer.
- b) obtener cara o cruz al lanzar una moneda

- c) sexo de una niña al nacer

El experimento con dos resultados posibles se denomina ensayo Bernoulli.

Cualquier experimento puede usarse para definir un ensayo Bernoulli, simplemente denotando algún evento A como éxito y su complemento A^c como fracaso.

La distribución de probabilidad para un ensayo Bernoulli depende sólo de un parámetro p probabilidad de éxito, y entonces $1-p$ es la probabilidad de fracaso $(1-p=q)$, donde $0<p<1$.
Si X es la v.a. definida por

$$\sigma^2 = X(x) = \begin{cases} 1, \text{ si } X \in A \\ 0, \text{ si } X \in A^c \end{cases} \tag{23}$$

Entonces X se llama v.a Bernoulli con parámetro p. La distribución de probabilidad de una v.a Bernoulli es de la siguiente forma:

$$P(x=1) = P(A) = p \tag{24}$$

$$P(x=0) = P(A^c) = 1-p = q \tag{25}$$

X	$f(X=x)$
0	q
1	p

la cual se puede resumir de la siguiente forma:

$$f(x) = p^x q^{1-x}, \quad \text{x=0,1.} \tag{26}$$

RESUMEN

La variable de Bernoulli se nombró saí en honor al matemático y físico suizo Daniel Bernoulli (1700-1782) nacido e Holanda. Entre sus principales aportaciones destaca el descubrimiento de uno de los principios básicos del comportamiento de los fluidos.

Un experimento se llama de Bernoulli, cuando solo tiene dos resultados posibles, alos que denominamos **éxito** y **fracaso**, y son denotados por p y q, respectivamente. Por otro lado cada prueba del experimento se le puede llamar **ensayo de Bernoulli**. Así a la variable aleatoria X definida en un experimento de Bernoulli que representa al éxito en un ensayo de Bernoulli le llamamos variable aleatoria de Bernoulli.
Una pregunta importante sería ¿Si el modelo se refiere solo a dos posibilidades de los valores de la variable aleatoria, cuál es su importancia en las aplicaciones?
Cuando se hacen estudios de mercado, casi todas las encuestas están diseñadas de tal forma que la respuesta a las preguntas que las conforman es **dicotómica** binaria. Es decir, preguntas del tipo:

- Le gusta o no un producto.
- Ve las noticias X a las 9 am.
- Está de acuerdo con la política del presidente sobre los indocumentados.

En términos estrictos el proceso de Bernoulli se caracteriza por lo siguiente:

- i. El experimento consta de ensayos repetidos.
- ii. Cada ensayo produce un resultado que se puede clasificar como éxito o fracaso.
- iii. La probabilidad de un éxito, que se denota con p, permanece constante de un ensayo a otro.
- iv. Los ensayos repetidos son independientes.

Considere el conjunto de experimentos de Bernoulli en el que se seleccionan tres artículos al azar de un proceso de producción, luego se inspeccionan y se clasifi can como defectuosos o no defectuosos. Un artículo defectuoso se designa como un éxito. El número de éxitos es una variable aleatoria X que toma valores integrales de ce ro a 3. Los ocho resultados posibles y los valores correspondientes de X son

Resultado	NNN	NDN	NND	DNN	NDD	DND	DDN	DDD
x	0	1	1	1	2	2	2	3

Como los artículos se seleccionan de forma independiente y se asume que el proceso produce 25 % de artículos defectuosos,

$$
\begin{aligned}
P(NDN) &= P(N)P(D)P(N) && (27)\\
&= \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) && (28)\\
&= \frac{9}{64}. && (29)
\end{aligned}
$$

La distribución de probabilidad de X es:

x	0	1	2	3
$f(x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

Como podemos apreciar, en realidad estas variables aleatorias resultan en una gran cantidad de situaciones de un grupo o población de estudio, que de hecho nos interesaría conocer la cantidad de personas que están a favor o en contra de la pregunta **dicotómica** formulada.

Por otro lado, en el estudio de la probabilidad y la estadística, el concepto de independencia juega un papel muy importante. Por consiguiente, existen muchos fenómenos probabilísticos discretos donde las pruebas de los experimentos se consideran independientes. Uno de estos modelos discretos que tiene mucho auge en los experimentos con pruebas independientes es el **modelo binomial.**

8.2. Distribución Binomial

Si un experimento que consiste de n ensayos Bernoulli independientes, cada uno con probabilidad de éxito p, se llama experimento binomial con n ensayos y parámetro p.
Ensayos independientes indica que los ensayos son eventos independientes, esto es, lo que ocurre en un ensayo no influye en el resultado de cualquier otro ensayo. El espacio muestral para un experimento binomial es el producto cartesiano de los espacios muestrales de los ensayos Bernoulli consigo mismo n veces.

Sea X el número total de éxitos en un experimento binomial con n ensayos y parámetro p.Entonces X se llama v.a binomial con parámetro n y p. Donde su distribución de probabilidades es:

$$f(x) = \binom{n}{x} \cdot p^x q^{n-x}, \quad x = 0, 1, 2, ..., n \tag{30}$$

$$f(x) = \frac{n!}{(n-x)!x!} \cdot p^k q^{n-x}. \tag{31}$$

El valor esperado o esperanza es $E(X) = np$, y la varianza es $Var(X) = npq$.

8.2.1. Ejemplos

1)

Cinco dados son lanzados una vez

- a)¿Cuál es la probabilidad de obtener un dado con el número 3.
- b)¿Cuál es la probabilidad de obtener dos dados con el número 3.
- c)¿Cuál es la probabilidad de obtener **al menos** un dado con el número 3.

Solución

Sea $n = 5$, X es la variable aleatoria que cuenta el número de dados con el número 3.

Y el evento de obtener el número 3 es: $D = \{0, 1, 2, 3, 4, 5\}$
La probabilidad para obtener un **solo** dado con el número 3 es $P(D) = \frac{1}{6}$ y, la probabilidad de NO obtener un **solo** dado con el número 3 es $P(D) = \frac{5}{6}$.
La distribución de probabilidad es:

$$f(x) = \binom{5}{x} \cdot \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{5-x}, \quad x = 0, 1, 2, ..., 5 \tag{32}$$

$$f(x) = \frac{5!}{(5-x)!x!} \cdot \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{5-x}. \tag{33}$$

- a) $f(x=1) = \binom{5}{1} \cdot \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{5-1} = (5) \cdot \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^4 = 0{,}4018$
  R—> $dbinom(x, n, p) = dbinom(1, 5, 1/6)$

- b) $f(x=2)=\binom{5}{2}\cdot\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{5-2}=(10)\cdot\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3=0{,}16075$

  R—> $dbinom(x,n,p)=dbinom(2,5,1/6)$

- c) $f(x\geq 1)=1-f(x=0)$.

  $f(x=0)=\binom{5}{0}\cdot\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{5-0}=(1)\cdot\left(1\right)\left(\frac{5}{6}\right)^5=0{,}4018.$

  Por lo tanto, $f(x\geq 1)=1-\left(\frac{5}{6}\right)^5=0{,}598122$

  R—> $1-dbinom(x,n,p)=1-dbinom(1,5,1/6)$

R—> $plot(dbinom(0:6,5,1/6))$

plot(dbinom(0 : 5, 5, 1/6),type="h",lwd=8,col="green",ylim = c(0,0,5),xlab="# de dados con el #3",ylab="Probabilidad",main="distrin de probab") [a]

[a] La gráfica empieza desde 0, hasta 5, el uno significa primer dato.

2)

La probabilidad de que una cierta clase de componente pase con éxito una determinada prueba de impacto es $\frac{4}{5}$. Encuentre la probabilidad de que exactamente 2 de los siguientes cuatro componentes que se prueben pasen la prueba.

Solución

$n=4$, X es la v.a. de pasar con éxito en la prueba de impacto. $X=\{0,1,2,3,4\}$

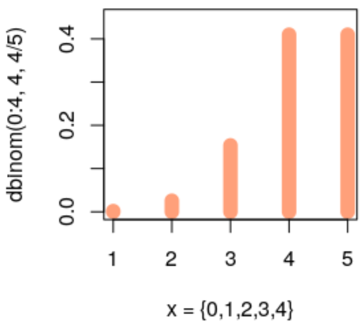
$p=\frac{4}{5}$ y $q=\frac{1}{5}$, entonces la distribución de probabilidad es:

$$f(x)=\binom{4}{x}\cdot\left(\frac{4}{5}\right)^x\left(\frac{1}{5}\right)^{4-x},\quad x=0,1,2,...,4 \tag{34}$$

$$f(x=2)=\binom{4}{2}\cdot\left(\frac{4}{5}\right)^2\left(\frac{1}{5}\right)^{4-2}=(6)\cdot\left(\frac{4}{5}\right)^2\left(\frac{1}{5}\right)^2=0{,}1536. \tag{35}$$

R—> $dbinom(x,n,p)=dbinom(2,4,4/5)$

plot(dbinom(0 : 4, 5, 4/5),type="h",col="lightsalmon",lwd=11,xlab = "$x=\{0,1,2,3,4\}$",ylim = c(0,0,45))



La probabilidad de que exactamente dos de las siguientes piezas cuatro de ellas que se prueben pasen la prueba es de 0,1536.

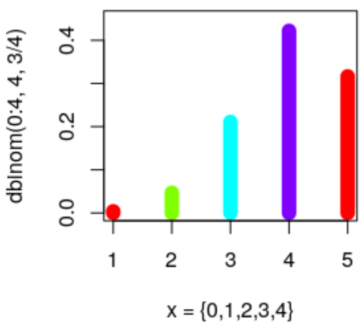
Ahora si el éxito de la prueba de impacto es $\frac{3}{4}$, el éxito que que dos de los siguientes cuatro componentes que se prueben pasen la prueba es

$$f(x=2) = \binom{4}{2} \cdot \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^3 = (6) \cdot \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^2 = 0{,}21093. \tag{36}$$

R—> $dbinom(x,n,p) = dbinom(2,4,3/4)$

La probabilidad de que exactamente dos de las siguientes piezas cuatro de ellas que se prueben pasen la prueba es de 0,21093.

plot(dbinom($0:4,4,3/4$),type = ”h”,col=rainbow(4),lwd= 11,xlab = ”$x = \{0,1,2,3,4\}$”,ylim = $c(0,0{,}45)$)



3)

Una empresa produce una línea de productos, talque el 35 % de los producidos en un mes es defectuoso. ¿cuál es la probabilidad de que los siguientes 1000 productos manufacturados en esa línea

- a) menos de 354 productos sean defectuosos?
- b) entre 342 y 364 productos sean defectuosos?, y

- c) exactamente 354 productos sean defectuosos?. En un mes.

Solución

$n = 1000$, Sea la v.a. $X = \{0, 1, 2, ..., 1000\}$que cuenta los productos defectuosos mensuales.

Probabilidad de los **defectuosos** en un mes es: $p(D) = 0{,}35$
Probabilidad de los **NO-defectuosos** en un mes es: $q(D) = 0{,}65$.
El valor medio total de los productos totales es: $\mu = n \cdot p = 1000(0{,}35) = 350$
La desviación estandar es: $\sigma = \sqrt{npq} = \sqrt{1000(0{,}35)(0{,}65)} = 15{,}08$.
$p(x \leq 354)$.
Para esto, tenemos que buscar en la tabla de la distribución normal estándar el valor de p para el valor mas cercano a la izquierda de $x = 354$.

$z = \frac{353{,}5-\mu}{\sigma} = 0{,}23 \quad p(0{,}23) = 0{,}091$. **(según la tabla de abajo).**

TABLA DE DISTRIBUCIÓN NORMAL ESTÁNDAR

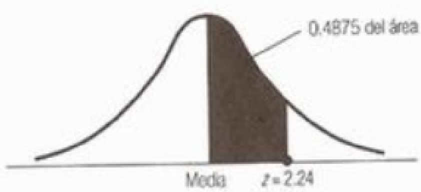


Tabla distribución de probabilidad normal estándar

*Áreas bajo la curva de distribución de probabilidad normal estándar, entre la media y valores positivos de z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	$\boxed{0.0910}$	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

- a) Por lo tanto, $p(x \leq 354) = 0{,}5 + 0{,}091 = 0{,}5091$. **Se suma primero $0{,}5$, pues es la mitad de la gráfica, esto significa que a la mitad de la gráfica de campana está el valor de μ.**

$R--->$ pnorm(353.5,350,15.08)

- b) Entre 342 y 364 defectuosos. A partir de $\mu = 350 \pm 15{,}0831$
  $z = \frac{353{,}5-\mu}{\sigma} = 0{,}23.$
  $z = \frac{341{,}5-\mu}{\sigma} = -0{,}56 \quad p(-0{,}56) = 0{,}2123.$**(según la tabla de arriba).**
  Por lo tanto, la probabilidad entre $p(342 < x < 354) = 0{,}23 + 0{,}2123 = 0{,}5438.$
  $R--->$ pnorm(364,350,15.0831)-pnorm(342,350,15.0831)
- c) La probabilidad de 354 productos defectuosos. Alrededor de $\mu = 350$.
  $z = \frac{353{,}5-\mu}{\sigma} = 0{,}23 \quad p(0{,}23) = 0{,}091.$
  $z = \frac{354{,}5-\mu}{\sigma} = 0{,}29 \quad p(0{,}29) = 0{,}1179.$**(según la tabla de arriba).**

4)

Una fábrica ha registrado el consumo de KWH(KiloWatts por hora) durante las últimas 25 semanas, mostrando los resultados en la siguiente tabla:

Semana	KWH	Semana	KWH	Semana	KWH
1	9	9	6	17	7
2	5	10	7	18	4
3	8	11	6	19	8
4	10	12	8	20	11
5	11	13	9	21	9
6	12	14	8	22	7
7	5	15	9	23	6
8	7	16	10	24	8
				25	10

- Calcular la función de probabilidad.
- Estimar la demanda de energía eléctrica para la próxima semana.

Solución

- a) Así $X = \{$va.a que representa el consumo de KWH durante la semana$\} = \{4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Así, se tabula la función de distribución para la v.a. X, como sigue:
- b) El valor esperado es entonces $E(X) = 4(1/5) + 5(2/25) + 6(3/25) + 7(4/25) + 8(5/25) + 9(4/25) + 10(3/25) + 11(2/25) + 12(1/25) = 8$

5)

En una rifa se venden 5000 billetes a \$1 cada uno. El único premio del sorteo es de \$1800. Calcular el resultado que debe esperar una persona que compra 3 billetes.

Solución

La variable aleatoria es $x = \{$cantidad de dinero ganado$\}$

$X = x$	$P(X = x) = p$
4	$\frac{1}{25}$
5	$\frac{2}{25}$
6	$\frac{3}{25}$
7	$\frac{4}{25}$
8	$\frac{5}{25}$
9	$\frac{4}{25}$
10	$\frac{3}{25}$
11	$\frac{2}{25}$
11	$\frac{2}{25}$
12	$\frac{1}{25}$

- Si gana la rifa : $1800 - 3 = 1777$ *pues gastó* 3 *en el billete.* Por lo tanto, $P(ganar) = 3/5000$.
- Si **no** gana la rifa: $P(perder) = 4997/5000$

La distribución de probabilidad es:

X		$P(X)$
ganar	1777	$3/5000 = 0{,}0006$
perder	-3	$4997/5000 = 0{,}9994$

El valor esperado es:

$$
\begin{aligned}
E(X) &= (1777)\left(\frac{3}{5000}\right) + (-3)\left(\frac{4997}{5000}\right) && (37) \\
&= -\left(\frac{9660}{5000}\right) && (38) \\
&= -1{,}932 && (39)
\end{aligned}
$$

Así el valor esperado para el comprador de 3 billetes es perder en promedio 1,932.

6)

La probabilidad de obtener **al menos** un 6 cuando se lanza 4 veces un dado.

Solución

Sea el evento $A = \{\text{Obtene al menos un 6}\}$ y $A^c = \{\text{No sacar ningún 6}\}$

Así, la probabilidad del evento A se calcula como:

$$\begin{aligned} P(A^c) &= P(A_1^c) \cap P(A_2^c) \cap P(A_2^c) \cap P(A_3^c) \cap P(A_4^c) && (40)\\ &= (\frac{5}{6})(\frac{5}{6})(\frac{5}{6})(\frac{5}{6}) && (41)\\ &= (\frac{5}{6})^4 && (42)\\ &= 0{,}482253 && (43) \end{aligned}$$

Entnces, la probabilidad de obtener al menos un seis (6) en los 4 lanzamientos es:

$$\begin{aligned} P(A) &= 1 - P(A^c) && (44)\\ &= 1 - (\frac{5}{6})^4 && (45)\\ &= 0{,}5177. && (46) \end{aligned}$$

7)

La probabilidad de que cierta clase de componente sobreviva a una prueba de choque es de $\frac{3}{4}$. Calcule la probabilidad de que sobrevivan exactamente 2 de los siguientes 4 componentes que se prueben.

Solución

Suponiendo que las pruebas son independiente y la probabilidad de cada prueba es $p = 3/4$, se obtiene que la distribcíon de la probabilidad es

$$\begin{aligned} f(x) &= \binom{4}{2} \cdot (\frac{3}{4})^2 \cdot (\frac{1}{4})^2 && (47)\\ &= \left(\frac{4!}{2!2!}\right)\left(\frac{3^2}{4^4}\right) && (48)\\ & \frac{27}{128}. && (49) \end{aligned}$$

¿De dónde proviene la palabra binomial?

La distribución binomial deriva su nombre del hecho de que los $n+1$ términos en la expansión binomial de $(p+q)^n$ corresponden a los diversos valores de f(x) para $x = 0, 1, 2, ..., n$. Es decir,

$$\begin{aligned} (p+q)^n &= \binom{n}{0} q^n + \binom{n}{1} pq^{n-1} + \binom{n}{2} p^2 q^{n-2} + ... + \binom{n}{n} p^n && (50)\\ &= f(x=0) + f(x=1) + f(x=2) + ... + f(x=n) && (51) \end{aligned}$$

Como la probabilidad total es 1, entonces $p + q = 1$, entonces

$$\sum_{0}^{n} f(x) = 1, \tag{52}$$

esta es una condición que se debe cumplir para cualquier distribución de probabilidad.

8.3. Distribución Hipergeométrica

Tanto la distribución binomial como la distribución hipergeométrica persiguen un mismo objetivo: el número de éxitos en una muestra que contiene observaciones. Lo que establece una diferencia entre estas dos distribuciones de probabilidad discreta es la forma en que se obtiene la información. Para el caso de la distribución binomial la información de la muestra se toma con reposición de una muestra finita, o sin reposición de una población infinita. Para el modelo hipergeométrico la información de la muestra se toma sin reposición de una población finita. Por lo tanto, la probabilidad de éxito, p,es constante a lo largo de todas las observaciones de un experimento binomial, en cambio, en una distribución hipergeométrica el resultado de una observación afecta el resultado de las obseravciones previas.

En general, el interés que se tiene es en la probabilidad de seleccionar x éxitos de los k posibles resultados o artículos también considerados éxitos y $n-x$ fracasos de los $N-k$ posibles resultados o artículos también considerados fracasos, cuando una muestra aleatoria de tamaño n se selecciona de N resultados o artículos totales. Esto se conoce como un experimento hipergeométrico.
La función de probabilidad de una v.a X con distribución hipergeométrica es

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, ..., n \quad x = 1, 2, ..., n. \tag{53}$$

Donde, la esperanza y la varianza es:

$$E(X) = n \cdot \frac{k}{N}, \tag{54}$$

$$Var(X) = n \cdot \frac{k}{N}\left(\frac{N-k}{N}\right)\left(\frac{N-n}{N-1}\right) \tag{55}$$

8.3.1. Ejemplos

1)

Un comité compuesto por cinco personas se selecciona aleatoriamente de un grupo formado por tres administradores y cinco contadores. Encuentre la distribución de probabilidad para el número de administradores en el comité.

Solución

Sea $X =$ la variable aleatoria que cuenta en número de administradores.
$x = \{0, 1, 2, 3\}$

$N = 8$
$n = 5$
$k = 3$
$N - k = 8 - 3 = 5$, 5 posibles resultados

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \tag{56}$$

$$f(x) = \frac{\binom{3}{x}\binom{5}{5-x}}{\binom{8}{5}} \tag{57}$$

- Para $x = 0$, es decir 0 administradores en el grupo de 5.

$$f(0) = \frac{\binom{3}{0}\binom{5}{5-0}}{\binom{8}{5}} = \frac{1 \cdot 1}{56} = \frac{1}{56} \tag{58}$$

R—> $dhyper(x, k, N - k, n) = dhyper(0, 3, 8 - 3, 5)$

- Para $x = 1$, es decir 1 administrador en el grupo de 5.

$$f(1) = \frac{\binom{3}{1}\binom{5}{5-1}}{\binom{8}{5}} = \frac{3 \cdot 5}{56} = \frac{15}{56} \tag{59}$$

R—> $dhyper(x, k, N - k, n) = dhyper(1, 3, 8 - 3, 5)$

- Para $x = 2$, es decir 2 administradores en el grupo de 5.

$$f(2) = \frac{\binom{3}{2}\binom{5}{5-2}}{\binom{8}{5}} = \frac{3 \cdot 10}{56} = \frac{30}{56} \tag{60}$$

R—> $dhyper(x, k, N - k, n) = dbinom(2, 3, 8 - 3, 5)$

- Para $x = 3$, es decir 3 administradores en el grupo de 5.

$$f(3) = \frac{\binom{3}{3}\binom{5}{5-3}}{\binom{8}{5}} = \frac{1 \cdot 10}{56} = \frac{10}{56} \tag{61}$$

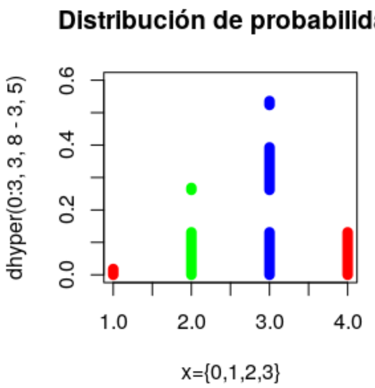
```
data.frame(X=0:3,Prob=dhyper(0:3,3,8-3,5))
```

R—> $dhyper(x, k, N-k, n) = dbinom(c(0, 1, 2, 3), 3, 8-3, 5)$

data.frame(X=0:3,Prob=dhyper(0:3,3,8-3,5))

La distribución de probabilidad es

X	$P(X)$
0	1/56
1	15/56
2	30/56
3	10/56



plot(dhyper(0 : 3, 3, 8 − 3, 5),type="h",$lwd = 8$,col=rainbow(3),xlab="$x = \{0, 1, 2, 3\}$",main="Distribución de probabilidad",lty=2,ylim=c(0,0.6))

2)

Entre 16 postulantes para un trabajo, 10 tenían un grado universitario. Si 3 de los postulantes son elegidos al azar para una entrevista. ¿Cuál es la probabilidad de que:

- a) ninguno tenga grado universitario?.
- b) tenga exactamente uno grado universitario?.
- c) dos tengan grado universitario?.
- d) los tres tengan grado universitario?.

Solución

Sea X = la variable aleatoria que cuenta el número de postulantes con título de la muestra.

$X = \{0, 1, 2, 3\}$.
$N = 16$, $k = 10$, $n = 3$, $N - k = 6$

- a) $x = 0$

$$f(0) = \frac{\binom{10}{0}\binom{16-10}{3-0}}{\binom{16}{3}} = \frac{1 \cdot 20}{560} = \frac{20}{560} = 0{,}035 \tag{62}$$

$$\boxed{\text{R—> } dhyper(x, k, N - k, n) = dbinom(0, 10, 16 - 10, 3)} \tag{63}$$

- b) $x = 1$

$$f(1) = \frac{\binom{10}{1}\binom{16-10}{3-1}}{\binom{16}{3}} = \frac{10 \cdot 15}{560} = \frac{150}{560} = 0{,}2678 \quad (64)$$

$$\boxed{\text{R—> } dhyper(x, k, N - k, n) = dbinom(1, 10, 16 - 10, 3)} \quad (65)$$

- c) $x = 2$

$$f(2) = \frac{\binom{10}{2}\binom{16-10}{3-2}}{\binom{16}{3}} = \frac{45 \cdot 6}{560} = \frac{270}{560} = 0{,}4821 \quad (66)$$

$\boxed{\text{R—> } dhyper(x, k, N - k, n) = dbinom(2, 10, 16 - 10, 3)}$

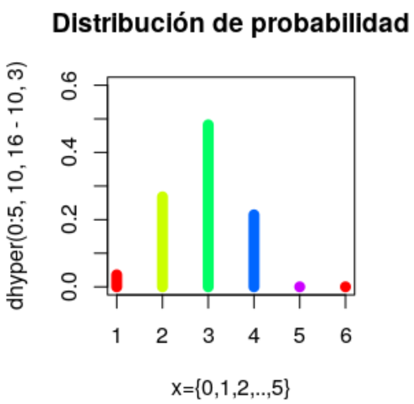
- d) $x = 3$

$$f(3) = \frac{\binom{10}{3}\binom{16-10}{3-3}}{\binom{16}{3}} = \frac{120 \cdot 1}{560} = \frac{120}{560} = 0{,}2143 \quad (67)$$

$\boxed{\text{R—> } dhyper(x, k, N - k, n) = dhyper(3, 10, 16 - 10, 3)}$

$\boxed{\text{R—> } dhyper(x, k, N - k, n) = dhyper(c(0, 1, 2, 3), 10, 16 - 10, 3)}$

plot(dhyper(0 : 5, 10, 16 – 10, 3),type=”h”,lwd= 8,col=rainbow(5),xlab=”x={0, 1, 2, ..., 5}”, main=”Distribución de probabilidad”,ylim=c(0,0.6))6))



3)

10 refrigeradores han sido devueltos al distribuidor por ser defectuosos. Supongamos, que 4 de esos son defectuosos y 6 tienen problemas leves. Si se examinan al azar 5. a) ¿Cuál es la probabilidad donde ninguno esté defectuoso? b) A lo sumo tenga fallas.

Solución

Sea X la variable aleatoria que cuenta el número de refrigeradores defectuosos.
$N = 10$

$k = 4$
$x = \{0, 1, 2, 3, 4\}$ $n = 5$

- a)

$$f(0) = \frac{\binom{4}{0}\binom{10-4}{5-0}}{\binom{10}{5}} = \frac{1 \cdot 6}{252} = \frac{6}{252} = 0{,}023 \qquad (68)$$

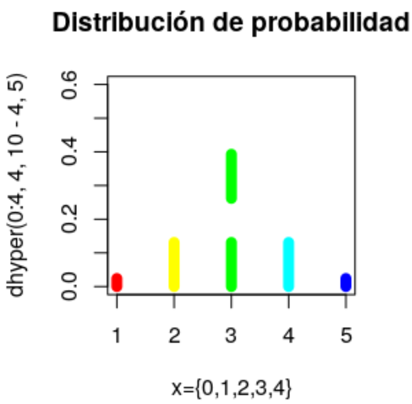
R—> $dhyper(x, k, N - k, n) = dhyper(0, 4, 10 - 4, 5)$

- b)

$$f(0) = 1 - f(0) = 1 - \frac{6}{252} = 1 - 0{,}023 = 0{,}9761 \qquad (69)$$

R—> $1 - dhyper(x, k, N - k, n) = 1 - dhyper(0, 4, 10 - 4, 5)$

plot(dhyper($0 : 4, 4, 10 - 4, 5$),type="h",lwd=8,col=rainbow(6),xlab="$x = \{0, 1, 2, 3, 4\}$",main="Distribución de probabilidad",lty=2,ylim=c(0,0.6))



8.4. Distribución de Poisson

Los experimentos que resultan en valores numéricos de una v.a X y que representan el número de resultados durante un intervalo de tiempo dado o en una región específica frecuentemente se llaman experimentos Poisson. El intervalo de tiempo dado puede ser de cualquier duración, por ejemplo, un minuto, un día, una semana, un mes o inclusive un año. Por tal motivo un experimento Poisson puede generar observaciones para una cierta v.a X que representen el número de llamadas telefónicas por hora que se recibe en una oficina, el número de días en que una determinada escuela se cierra en invierno debido a la nieve, o al número de juegos pospuestos debido a la lluvia durante una temporada de fútbol.

El número x de resultados que ocurren en un experimento Poisson se llama **v.a. de Poisson**. El número promedio de resultados se calcula de la forma $E(X) = \lambda t$, donde t es el tiempo o región específicos de interés.

La distribución es de la forma.

$$f(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \tag{70}$$
$$f(x) = \frac{e^{-\mu}\mu^x}{x!} \tag{71}$$

8.4.1. Ejemplos

1)

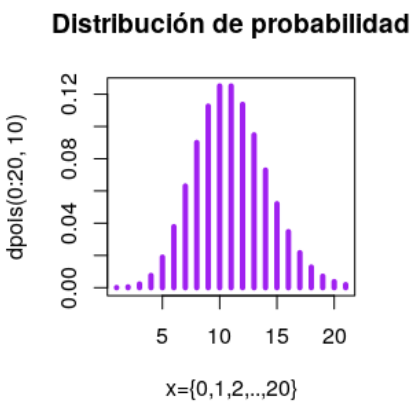
Si 10 es el número promedio de camiones que llega al estacionamiento al día. El estacionamiento puede atender máximo 15 camiones en un día. a) ¿Cuál ess la probabilidad de que en un día tengan que regresar camiones?

Solución

$\mu = 10$. X es la v.a. igual al número de camiones al día.

$$f(x) = \frac{e^{-\mu}\mu^x}{x!} \tag{72}$$
$$f(x > 15) = 1 - f(x \leq 15) \tag{73}$$
$$= 1 - (f(0) + f(1) + ... + f(15)) \tag{74}$$
$$= 1 - e^{-10}(1 + 10 + 100/2 + 1000/6 + ...) \tag{75}$$
$$= 0{,}00487 \tag{76}$$

R—> $1 - ppois(x, \mu) = 1 - ppois(15, 10)$ ó, de igual manera:
R—> $1 - sum(dpois(c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15), 10))$



plot(dpois(0 : 20, 10),type=”h”,lwd=4,col=”purple”,xlab=”x= {0, 1, 2, .., 20}”,main=”Distribución de probabilidad”)

9. EJERCICIOS

9.1. Problema 1

Si se lanza tres veces una moneda. Sea la variable aleatoria: X =número de caras que se obtienen. Se pide:

- a) Distribución de probabilidad de X.
- b) Función de distribución de X. Representación gráfica
- c) Media, varianza y desviación típica de X.
- d) Probabilidad de que salgan a lo sumo dos caras.
- e) Probabilidad de que salgan al menos dos caras.

Solución:

a) Espacio muestral: $\Omega = \{$ (ccc), (ccx), (cxc), (xcc),(cxx),(xcx),(xxc),(xxx) $\}$

$X(ccc) = 3$ $\qquad P(X = 3) = \frac{1}{8}$
$X(ccx) = X(cxc) = X(xcc) = 2$ $\qquad P(X = 2) = \frac{3}{8}$
$X(cxx) = X(xcx) = X(xxc) = 1$ $\qquad P(X = 1) = \frac{3}{8}$
$X(xxx) = 1$ $\qquad P(X = 0) = \frac{1}{8}$

La distribución de probabilidad será:

$X = x_i$	$P(X = x_i) = p_i$	$x_i \cdot p_i$	x_i^2	$x_i^2 \cdot p_i$
$x_1 = 0$	$\frac{1}{8}$	0	0	0
$x_2 = 1$	$\frac{3}{8}$	$\frac{3}{8}$	1	$\frac{3}{8}$
$x_3 = 2$	$\frac{3}{8}$	$\frac{6}{8}$	4	$\frac{12}{8}$
$x_4 = 3$	$\frac{1}{8}$	$\frac{3}{8}$	9	$\frac{9}{8}$
	1	$\frac{12}{8} = 1{,}5$		$\frac{24}{8} = 3$

b) La función de distribución: $F(x) = P(X \leq x) = \sum P(X = x_i) = \sum p_i$.

$x < 0 \quad F(x) = P(X \leq x) = 0$
$0 \leq x < 1 \quad F(x) = P(X = 0) = \frac{1}{8}$
$1 \leq x < 2 \quad F(x) = P(X < 2) = P(X = 0) + P(X = 1) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$
$2 \leq x < 3 \quad F(x) = P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$
$x = 3 \quad F(x) = P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$
$x > 3 \quad F(x) = P(\Omega) = 1$

$X = x_i$	0	1	2	3
$P(X = x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$P(X \leq x)$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{7}{8}$	1

c) Media, varianza y desviación de X:

Media: $E(x) = x \cdot P = \frac{12}{8} = 1{,}5$
$E(x^2) = x^2 \cdot P = \frac{24}{8} = 3$
Varianza:$E(x^2) - E(x)^2 = 3 - 1{,}5^2 = 0{,}75$
Desviación estandar: $\sigma = \sqrt{0{,}75}$

d) Probabilidad de que salgan a lo sumo dos caras

$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$

e) Probabilidad de que salgan al menos dos caras
$P(x \geq 2) = P(x = 2) + P(x = 3) = \frac{3}{8} + \frac{1}{8} = \frac{1}{2}$

9.2. Problema 2

Sea la variable aleatoria X = número de hijos por familia de una ciudad tiene la siguiente distribución de probabilidad:

X	0	1	2	3	4	5	6
$P(X = x)$	0,47	0,3	0,1	,06	0,04	0,02	0.01

Calcular:

- a) Valor esperado o esperanza matemática.
- b) Varianza y desviación estandar.
- c) Si el Ayuntamiento de la ciudad paga 2000 pesos por hijo, donde $Y = 2000x$, ¿cuál es la distribución de probabilidad?

Solución:
a)

$X = x$	$P(X = x) = p$	$x \cdot p$	x^2	$x^2 \cdot p$
$x = 0$	0,47	0	0	0
$x = 1$	0,3	0,3	1	0,3
$x = 2$	0,1	0,2	4	0,4
$x = 3$	0,06	0,18	9	0,54
$x = 4$	0,04	0,16	16	0,64
$x = 5$	0,02	0,10	25	0,5
$x = 6$	0,01	0,06	36	0,36
	1	1		2,74

Media: $E(x) = \sum_{x=0}^{6} x \cdot p = 1$. Si se toma al azar una familia de la ciudad, el número de hijos que se espera que tenga por término medio es uno.

b)
Varianza y desviación:
$E(x)^2) = \sum x \cdot p = 2{,}74$
Varianza: $E(x^2) - E(x)^2 = 2{,}74 - 1^2 = 1{,}74$
Desviación estandar: $\sqrt{1{,}74} = 1{,}32$

c)

$Y = y$	$P(Y = y) = p$
$y = 0$	0,47
$y = 2000$	,3
$y = 4000$	0,1
$y = 6000$	0,06
$y = 8000$	0,04
$y = 10000$	0,02
$y = 12000$	0,01
	1

(I) *Problem.* Toss a coin until the first head is obtained. Let X be the number of tosses until the first head is obtained. Find the probability distribution of X.

Solution.
$P(X = x) = (1-p)^{x-1}p$. This is called the geometric distribution.

(II) The *cumulative distribution function* (cdf) of a discrete random variable X is is defined as

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y).$$

(III) Find cdf of the geometric distribution.

Solution.
For positive integer x, $F(x) = 1 - (1-p)^x$.

(IV) The *expected value* of a discrete random variable X is defined as $\mathbb{E}X = \sum xp(x)$. The expectation of the geometric distribution is $1/p$.

(V) *Problem.* If two players A and B flip a biased coin alternately and the first player to obtain a head wins. The probability of obtaining a head is $p > 0$ at each toss. If A flips first, find the probability that A wins the game.

Solution.
Let X be the number of tosses A makes until he gets head.

$$\begin{aligned} P(A \text{ wins }) &= P(X=1) + P(X=2) + P(X=3) + \cdots \\ &= p + (1-p)^2 p + (1-p)^4 p + \ldots \\ &= \frac{p}{1-(1-p)^2} \end{aligned}$$

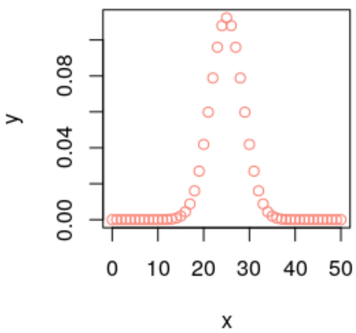
10. RStudio

El objetivo principal de este curso es proporcionar los elementos básicos para empezar a trabajar con el lenguaje de programación R en el ámbito de la Estadística. R, también conocido como "GNU S", es un entorno y un lenguaje para el cálculo estadstico y la generacion de gráficos. R implementt()a un dialecto del premiado lenguaje S, desarrollado en los Laboratorios Bell por John Chambers et al.

10.1. Distribución binomial

- Probabilidad de obtener al menos 26 caras de 51 lanzamientos de una moneda:
  pbinom(26,51,0.5)
- Cuantas caras tendrán la probabilidad de $p = 0{,}25$ cuando una moneda se lance 51 veces.
  qbinom(0.25,51,1/2)
- Encuentra 8 valores aleatorios de una muestra de 150 con probabilidad de 0.4.
  rbinom(8,150,.4)

```
# Crea una lista de números en el eje x del 0 al 50.
x <- seq(0, 50  by  = 1)
# Crear los valores de la distribución binomial(y), para x = {0, ..,50}.
y <- dbinom(x, 50, 0,5).
# Graficar para n=50, p=0.5
plot(x,y,col="Salmon")
```



10.2. Gráficas

10.2.1. Pie

R es el lenguaje de programación tiene numerosas bibliotecas para crear tablas y gráficos. Un gráfico circular es una representación de valores como segmentos de un círculo con diferentes colores. Las secciones están etiquetadas y los números correspondientes a cada sección también se representan en la tabla.

En R, el gráfico circular se crea utilizando la función pie() que toma números positivos como una entrada vectorial. Los parámetros adicionales se utilizan para controlar las etiquetas, el color, el título, etc.

La sintaxis básica para crear un gráfico circular utilizando la R es

pie(x, labels, radius, main, col, clockwise)

- **x** es un vector que contiene los valores numéricos utilizados en el gráfico circular.
- **label** se utilizan para dar descripción a los ejes.
- **radius** indica el radio del círculo del gráfico circular (valor entre -1 y $+1$).
- **main** indica el título del cuadro.
- **col** indica la paleta de colores.
- **clockwise** es un valor lógico que indica si los cortes se dibujan en el sentido de las agujas del reloj o en sentido contrario a las agujas del reloj

```
# Crear los datos x de la gráfica y las etiquetas.
x <- c(21, 62, 10, 53)
labels <- c("London", "New York", "Singapore", "Mumbai")
# Graficar el pastel
pie(x,labels,col=rainbow(5))
```

Podemos cambiar las características de la gráfica agregando más parámetros a la función pie. Usaremos el parámetro **main** para agregar un título a la gráfica y otro parámetro es **col** que hará uso de la paleta de colores del **rainbow()**. La longitud de la paleta debe ser la misma que la cantidad de valores que tenemos para el gráfico. Por eso utilizamos la longitud (x).

```
# Crear los datos x de la gráfica y las etiquetas.
x <- c(21, 62, 10, 53)
labels <- c("London", "New York", "Singapore", "Mumbai")
```

# Graficar el pastel con título y paleta de colores
pie(x, labels, main = ”Gráfica pastel de la ciudad”, col = rainbow(length(x)))

Podemos agregar un porcentaje de sector y una leyenda del gráfico creando variables de gráfico adicionales.

# Crear los datos x de la gráfica y las etiquetas.
x <- $c(21, 62, 10, 53)$
labels <- c(”London”, ”New York”, ”Singapore”, ”Mumbai”)
# Graficar el pastel con título y paleta de colores
pie(x, labels = piepercent, main = ”Gráfica pastel de ciudades”,col = rainbow(length(x))) legend(”left”, c(”London”,”New York”,”Singapore”,”Mumbai”), cex = 0.5, fill = rainbow(length(x)))

Gráfico circular 3D

Se puede dibujar un gráfico circular con 3 dimensiones utilizando el paquete **plotrix** tiene una función llamada pie3D() que se usa para esto.

# Obtener la libreria graficas en 3D. library(plotrix) # Crear los datos x de la gráfica y las etiquetas.
x <- $c(21, 62, 10, 53)$
labels <- c(”London”, ”New York”, ”Singapore”, ”Mumbai”)

```
# Graficar el pastel con título y paleta de colores
pie3D(x,labels = labels,explode = 0.1, main = "Gráfica pastel de ciudades ")
```

[3]

10.2.2. Histograma

Un histograma representa las frecuencias de los valores de una variable agrupada en rangos. El histograma es similar al chat de barra, pero la diferencia es que agrupa los valores en rangos continuos. Cada barra en el histograma representa la altura del número de valores presentes en ese rango.

R crea un histograma usando la función **hist().** Esta función toma un vector de datos como entrada y usa algunos parámetros más para construir histogramas.

La sintaxis básica para crear un histograma utilizando la R es

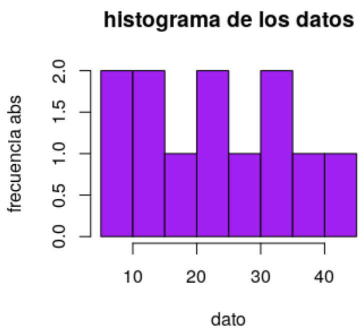
hist(v,main,xlab,xlim,ylim,breaks,col,border)

- **v** Es un vector que contiene valores numéricos utilizados en el histograma.
- **main** Indica el título del histograma.
- **col** Se utiliza para ajustar el color de las barras.
- **border** Se utiliza para establecer el color del borde de cada barra.
- **xlab** Se utiliza para dar descripción del eje x.
- **xlim** se utiliza para especificar el rango de valores en el eje x.
- **ylim** se utiliza para especificar el rango de valores en el eje y.
- **breaks** Se usa para mencionar el ancho de cada barra.

Ejemplo. Calcular la gráfica de los datos.

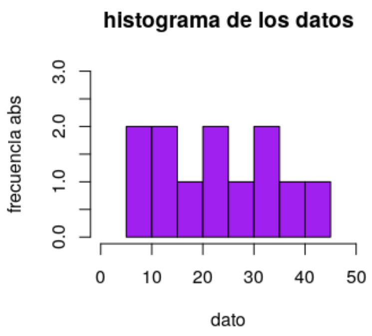
[3] explode=0.1, separa cada rebanada de la gráfica.

# Crear los datos x de la gráfica y las etiquetas.
v <- $c(9, 13, 21, 8, 36, 22, 12, 41, 31, 33, 19)$
# Graficar
hist(v,col = "purple",border = "black",xlab="dato",ylab="frecuencia abs",main = "histograma de los datos")



Cambiar rango en eje x y eje y: $x = \{0, ..., 50\}$, $y = \{0, ..., 3\}$

# Crear los datos x de la gráfica y las etiquetas.
v <- $c(9, 13, 21, 8, 36, 22, 12, 41, 31, 33, 19)$
# Graficar
hist(v,col = "purple",border = "black",xlab="dato",ylab="frecuencia abs",main = "histograma de los datos",xlim=c(0,50),ylim=c(0,3))



10.2.3. Gráfica de barras

Un gráfico de barras representa datos en barras rectangulares con la longitud de la barra proporcional al valor de la variable. R usa la función **barplot()** para crear gráficos de barras. R puede dibujar barras verticales y horizontales en el gráfico de barras. En el gráfico de barras se puede dar colores diferentes a cada una de las barras.

La sintaxis básica para crear una gráfica de barras utilizando R es

barplot(H,xlab,ylab,main, names.arg,col)

- **H** es un vector o matriz que contiene valores numéricos utilizados en el gráfico de barras.
- **xlab** Es la etiqueta para el eje x.
- **ylab** Es la etiqueta para el eje y.
- **main** Es el título del gráfico de barras.
- **names.arg** Es un vector de nombres que aparecen debajo de cada barra.
- **col** Se utiliza para dar colores a las barras en el gráfico.

Ejemplo. Crear un gráfico de barras simple utilizando solo el vector de entrada y el nombre de cada barra.

```
# Crear los datos x de la gráfica.
H <- c(7, 12, 28, 3, 41)
M <- c("Mar", "Apr", "May","Jun","Jul")
# Graficar
barplot(H,names.arg=M,xlab="Mes",ylab="Ingresos",col="Salmon",main="gráfica de ingresos"
,border="purple")
```

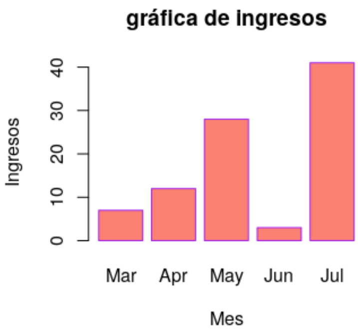
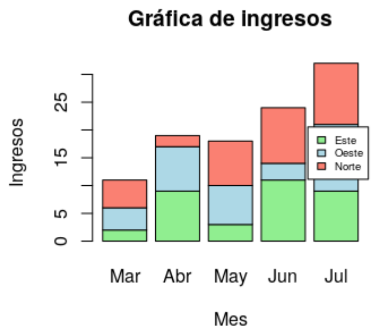


Gráfico de barras de grupo y gráfico de barras apiladas. Podemos crear un gráfico de barras con grupos de barras y pilas en cada barra utilizando una matriz como valores de entrada.

Más de dos variables se representan como una matriz que se utiliza para crear el gráfico de barras del grupo y el gráfico de barras apiladas.

```
# Crear los datos de la gráfica.
colores = c("lightgreen","lightblue","salmon")
meses <- c("Mar", "Abr","May","Jun","Jul")
region <- c("Este","Oeste","Norte")
# Ordenar los datos de la gráfica.
Valores <- matrix(c(2,9,3,11,9,4,8,7,3,12,5,2,8,10,11), nrow = 3, ncol = 5, byrow = TRUE)
```

\# Graficar
barplot(Valores, main = "Gráfica de Ingresos", names.arg = meses, xlab = "Mes", ylab = "Ingresos", col = colores)
Leyendas
legend(right", regions, cex = 0.6, fill = colors)



10.2.4. Gráfica de caja

Los diagramas de caja son una medida de cuán bien distribuidos están los datos en un conjunto de datos. Divide el conjunto de datos en tres cuartiles. Esta gráfica representa el mínimo, el máximo, la mediana, el primer cuartil y el tercer cuartil en el conjunto de datos. También es útil para comparar la distribución de datos entre conjuntos de datos al dibujar diagramas de caja para cada uno de ellos.
Los diagramas de caja se crean en R utilizando la función **boxplot()**.

La sintaxis básica para crear una gráfica de caja utilizando R es

boxplot(x, data, notch, varwidth, names, main)

- **x** es un vector de datos.
- **data** es el marco de datos
- **notch** Es un valor lógico. Establecer como VERDADERO para dibujar una muesca.
- **varwidth** Es un valor lógico. Establecer como verdadero para dibujar el ancho de la caja proporcional al tamaño de la muestra.
- **names** son las etiquetas de grupo que se imprimirán debajo de cada diagrama de caja.
- **main** Se utiliza para dar un título a la gráfica.

10.2.5. Ejemplo 1. Gráficas juntas.

El gobierno desea saber si el número promedio de hijos por familia ha descendido los últimos años. Para ello ha encuestado a 20 familias preguntando su número de hijos, y ha obtenido:
$2, 4, 5, 3, 1, 2, 4, 2, 3, 0, 3, 3, 4, 5, 5, 0, 3, 2, 1, 2$

```
hijos=c(2,4,5,3,1,2,4,2,3,0,3,3,4,5,5,0,3,2,1,2)
fabs<-table(hijos)
fabs
frel<- fabs/length(hijos)
frel
fabsacum<- cumsum(fabs)
fabsacum
frelacum<- fabsacum/length(hijos)
frelacum
tablafrecuencias<-cbind(fabs, frel,fabsacum,frelacum)
tablafrecuencias
```

Estadística

```
mean(hijos)
median(hijos)
var(hijos)
sd(hijos)
summary(hijos)
```

Gráficas

¿Cuántos hijos tiene el %25 de familias encuestadas?

```
quantile(hijos,.25)
```

¿Cuántos hijos tiene el %5 de familias encuestadas?

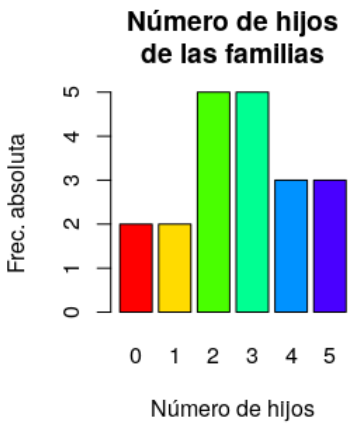
```
quantile(hijos,.05)
```

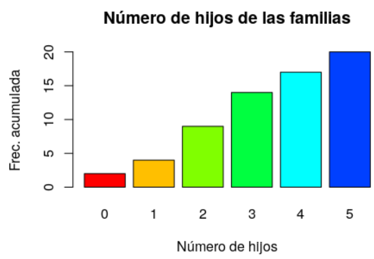
¿Cuántos hijos tiene el %75 de familias encuestadas?

```
quantile(hijos,.75)
```

```
la <– c("0-hijos","1-hijos","2-hijos","3-hijos","4-hijos","5-hijos")
barplot(fabs,ylab="Frec. absoluta", xlab="Número de hijos", main="Número de hijos de las familias",
col=rainbow(7))
pie(fabs, ylab="Frec. absoluta",xlab="Número de hijos", main="Número de hijos de las familias",
col=rainbow(7))
plot(fabs, ylab="Frec. absoluta", xlab="Número de hijos", main="Número de hijos de las familias",
col="blue",lwd=3,type="l")
```

```
library(plotrix)
pie3D(fabs,labels=la, main="Número de hijos de las familias", col=rainbow(7),explode=0.1)
```





10.2.6. Ejemplo 2. Gráficas juntas.

- i) Import Dataset
- ii) From Text(base)
- iii) \Carpeta\...\riesgos.csv
- iv) Import
- v) ?read.csv
- vi) head(riesgos)
- vii) tail(riesgos)

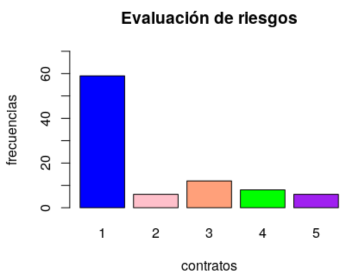
- viii) colnames(riesgos)
- ix) dim(riesgos)
- x) summary(riesgos)
- xi) attach(riesgos)
- xii) summary(edad)
- xiii) sd(edad)
- xiv) sum(edad)
- xv) summary(riesgos[, c("edad","peso","talla")])
- xvi) table(jornada)
- xvii) table(jornada) / margin.table(table(jornada))
- xviii) range(edad,na.rm=TRUE)

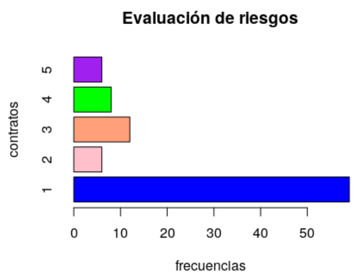
```
boxplot(edad, col = "red",notch=TRUE, main="Evaluación de riesgos laborales")
```

- i) table(contrato)
- ii) la<-c("1-contrato", "2-contrato", "3-contrato","4-contrato", "5-contrato")
  pie(table(contrato), la, col=rainbow(7), main="Evaluación de riesgos laborales")

```
library(plotrix)
la=c("1","2","3","4","5")
pie3D(table(contrato), labels=la, main="Evaluación de riesgos laborales", explode=0.1)
```

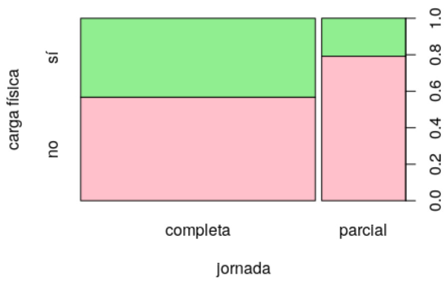
- i) table(contrato)
- ii) barplot(table(contrato))
- iii) barplot(table(contrato), xlab = "contratos", ylab ="frecuencias", main = "Evaluación de riesgos", col=c("blue","pink","lightsalmon","green","purple"), ylim=c(0,70))
- iv) barplot(table(contrato), xlab = "frecuencias", ylab = "contratos", main = "Evaluación de riesgos", col=c("blue","pink","lightsalmon", "green","purple"), horiz = TRUE)
- v) boxplot(edad ∼ carpsiqui, col = "red",notch=TRUE, main="Evaluación de riesgos laborales")

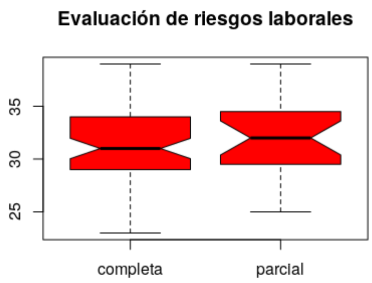


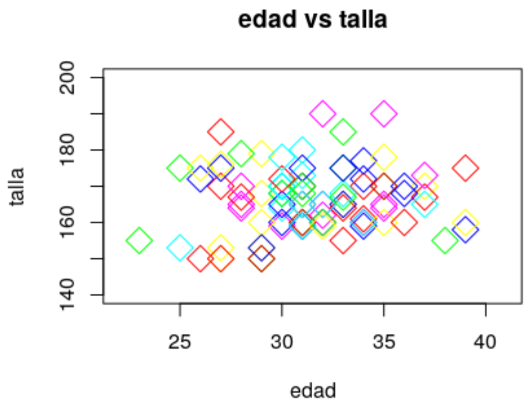


- i) table(jornada, carfisi, carpsiqui)
- ii) table(jornada, carfisi)

- iii) plot(jornada, carfisi, xlab="jornada", ylab="carga física", col=c("pink","lightgreen"))
- iv) boxplot(edad ∼ jornada, col = "red", notch=TRUE, main="Evaluación de riesgos laborales")
- v) plot(talla ∼ edad, pch=5, col =rainbow(6), main=" Evaluación de riesgos laborales", cex=2, ylim=c(140,200), xlim=c(22,41))







10.2.7. Ejemplo 3. Gráficas juntas.

Los siguientes datos corresponden a las notas obtenidas por 100 alumnos en un curso de Estadística:

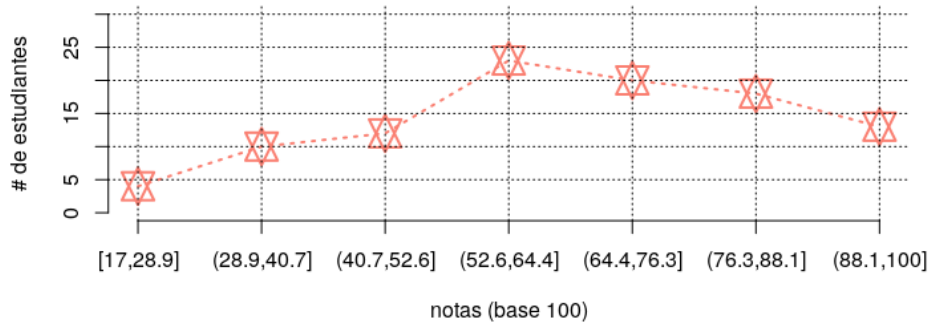
- library(readxl)
  intervalos <- read_excel("Dropbox/intervalos.xlsx")
  View(intervalos)
- intervalos
- range(intervalos)
- nclass.Sturges(intervalos)[a]
- interva=c(t(intervalos))

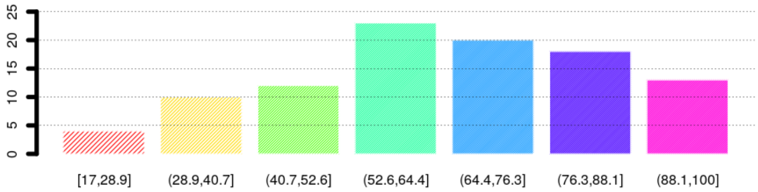
100	87	54	82	93	47	40	53	88	58
84	65	57	66	25	70	85	36	61	34
33	33	100	69	77	88	63	17	42	55
98	70	68	70	65	70	84	52	60	54
57	47	57	86	25	66	40	100	32	39
90	83	64	95	85	100	67	60	42	65
82	85	62	72	65	76	23	96	30	45
77	55	100	80	55	52	85	68	53	82
55	51	47	47	64	75	65	60	45	75
62	93	98	58	95	83	33	70	51	60

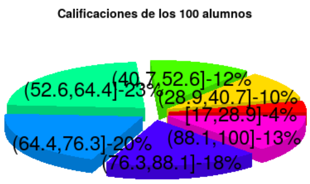
- interva
- nclass.Sturges(interva)
- range(interva)
- seq(17,100,length=nclass.Sturges(interva))
- intervalosno=cut(interva,breaks=seq(17,100,length=8),include.lowest=TRUE)
- intervalosno
- library(agricolae)
- tbFreqno=table.freq(hist(interva,plot=FALSE))
- tbFreqno. (Éstos 3 renglones es exclusivo para el área ágrícola.)
- notas<-table(intervalosno)
- notas
- frel<-notas/length(interva)
- frel
- facum<-cumsum(notas)
- facum
- frelacum<-facum/length(interva)
- frelacum
- tablatotal<-cbind(notas,frel,facum,frelacum)
- tablatotal
- mean(interva)

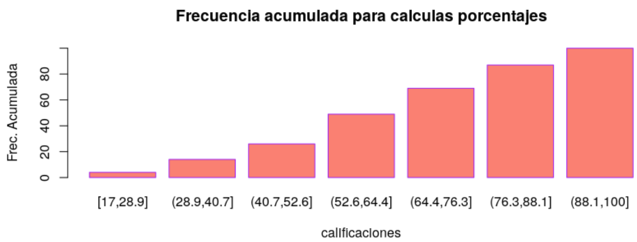
- median(interva)
- sd(interva)
- var(interva)
- plot(notas,type="o",pch=11,cex=2.5,col="salmon",ylim=c(0,30),lty=3,xlab="notas (base 100)", ylab ="♯ de estudiantes")
- abline(v=c(3,4,5,6,7),lty=3)
- abline(h=c(5,10,15,20,25,30),lty=3)
- barplot(notas,col=rainbow(7),border="white",lwd=5,ylim=c(0,25),density=c(25,35,45,55,65,75,80))
- abline(h=c(5,10,15,20,25,30),lty=3,lwd=0.5)
- la<- c("[17,28.9]-4 %,"(28.9,40.7]-10 %","(40.7,52.6]-12 %","(52.6,64.4]-23 %","(64.4,76.3]-20 %", "(76.3,88.1]-18 %","(88.1,100]-12 %")
- pie(notas,labels=la,col=rainbow(7),main="calificaciones de los 100 alumnos")
- library(plotrix)
- pie3D(notas,labels=la,col=rainbow(7),main="Calificaciones de los 100 alumnos",radius = 2,explode=0.2,border=NA,cex.lab=0.5,cex.main=0.9)
- barplot(facum,xlab="calificaciones",ylab="Frec. Acumulada",ylim=c(0,100),col="salmon", border="purple",main="Frecuencia acumulada para calculas porcentajes")

[a] Para construir los intervalos de edad podemos utilizar la regla de Sturges, que nos proporciona un número adecuado de intervalos en función del rango de los datos (ver help(nclass.Sturges); de modo alternativo podría utilizarse la regla de Scott o la de Freedman-Diaconis). Una vez que hemos decidido el número de intervalos, generamos una secuencia de valores con los límites de cada intervalo y utilizamos la función cut() para que nos construya una nueva variable que recodifica la edad en dichos intervalos.









10.2.8. Ejemplo 4. Gráficas juntas.

Una empresa que tiene 50 trabajadores se propone reestructurar las remuneraciones, se estudia los años de servicio de los trabajadores determinándose los siguientes resultados:

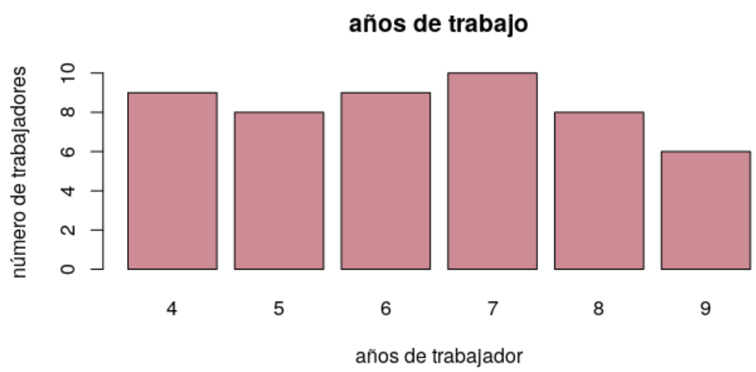
- library(readxl)
  anios <- read_excel("Dropbox/anios.xlsx")
  View(anios)

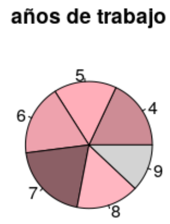
- vuno<-c(t(anios))
  vuno

- años<-table(vuno)
  años

4	5	4	6	7	9	7	7	5	8
8	7	6	7	7	4	6	8	8	9
6	8	9	5	6	5	4	7	9	6
7	6	5	4	4	4	6	8	8	7
8	9	5	5	4	6	7	9	5	4

- frel <- años/length(vuno)
  frel
- facum<-cumsum(años)
  facum
- frelacum<- facum/length(vuno)
  frelacum
- summary(vuno)
- sd(vuno)
- barplot(años,col="lightpink3",main="años de trabajo",xlab = "años de trabajador",ylab = "número de trabajadores")
- pie(años,col=c("lightpink3","lightpink1","lightpink2","lightpink4","lightpink","lightgrey"), main="años de trabajo")
- et<- c("18 %","16 %","18 %","20 %","16 %","12 %")
- pie(años,col=c("lightpink3","lightpink1","lightpink2","lightpink4","lightpink","lightgrey"), main="años de trabajo",label=et)
- library(plotrix)





10.2.9. Ejemplo 5. Gráficas juntas.

Un nuevo hotel va a abrir sus puertas en cierta ciudad. Antes de decidir el precio de sus habitaciones, el gerente investiga los precios por habitación de 40 hoteles de la misma categoría de esa ciudad. Los datos obtenidos en miles de pesetas fueron

3,9	4,7	3,7	5,6	4,3	4,9	5,0	6,1	5,1	4,5
5,3	3,9	4,3	5,0	6,0	4,7	5,1	4,2	4,4	5,8
3,3	4,3	4,1	5,8	4,4	4,8	6,1	4,3	5,3	4,5
4,0	5,4	3,9	4,7	3,3	4,5	4,7	4,2	4,5	4,8

10.3. PRUEBAS PARAMÉTRICAS

- Se conoce el modelo de distribución de la población objeto de estudio y se desconoce un número finito de parámetros de dicha distribución que hay que estimar con los datos de la muestra.
- Requieren conocer la distribución de la muestra para poder realizar realizar inferencias inferencias sobre la población población.

10.4. PRUEBAS NO PARAMÉTRICAS

- Son métodos de distribución libre. No requieren conocer la distribución de la muestra.
- Se utilizan estadísticos cuya distribución se determina con independencia de cuál sea la distribución de la población.

¿Para qué se utilizan?

i) Son una alternativa a las pruebas paramétricas cuando los datos no cumplen los requisitos de las pruebas paramétricas.

ii) Permiten conocer cómo es la forma de la distribución de la población de la que se ha extraído la muestra. Contrastes de Bondad de Ajuste para conocer la forma de la población que ha originado la muestra.

Referencias

[1] Vladimirovna Panteleeva, Olga Gutierrez Gonzalez, Eduardo. Probabilidad Y Estadistica. Aplicaciones A La Ingenieria y a las Ciencias (Patria, 2014).

[2] Murray R. Spiegel. Probabilidad y Estadística: 760 Problemas Resueltos (Schaum). 1ra Edición (McGraw-Hill 1982)

[3] http://www.dma.ulpgc.es/profesores/personal/stat/cursoR4ULPGC/8-estaDescriptiva.html

[4] https://www.tutorialspoint.com/r/index.htm

[5] http://biocosas.github.io/R/030_estadistica_descriptiva.html