Probabilidad y Estadística

Año de cursada 1° AÑO

Clase Nº8: Más distribuciones de Probabilidad.

Contenido: *Distribución binomial, Distribución de Poisson, Relación entre la distribución binomial y de Poisson. Características y parámetros.*

1. Presentación:

Bienvenidos y bienvenidas a nuestra clase n° 7 de Probabilidad y Estadística. Esta semana, estudiaremos otros modelos de distribuciones de probabilidad que amplían el campo de conocimiento para el análisis de situaciones prácticas.

Anteriormente, habíamos definido un modelo de distribución de probabilidades que involucra conceptos que han sido desarrollados en el tránsito por esta materia, y nos permitió calcular e interpretar probabilidades de algunos fenómenos en situaciones particulares.

En esta clase, como ya se mencionó, avanzaremos con el estudio de dos modelos de distribución de probabilidades que nos permitirán hacer predicciones en situaciones específicas, de acuerdo con las características del conjunto que estemos estudiando.

2. Desarrollo:

La distribución de Bernoulli.

Como ya sabemos, un experimento aleatorio es cualquier actividad que pueda arrojar un conjunto de soluciones que no pueden anticiparse, aunque se realicen bajo las mismas condiciones iniciales.

Ahora, en este caso vamos a considerar aquellos experimentos que pueden arrojar solo dos posibles resultados que son mutuamente excluyentes. Esto es que el conjunto de soluciones es dicotómico y la aparición de un resultado, anula la posibilidad de aparición del otro resultado. Cada lanzamiento o escogencia se denomina "prueba o intento". Por citar ejemplos, podemos considerar como experimento lanzar una moneda, donde solo se puede obtener cara o seca; o los resultados de un examen de Probabilidad de Estadística, cuyos resultados pueden ser aprobar o desaprobar. Estos se denominan **Ensayos o Pruebas de Bernoulli**, en honor al matemático Jakob Bernoulli fue quien desarrolló dicha teoría. Cada uno de estos posibles resultados, es denominado de forma general, como *éxito* (p) o *fracaso* (q), ello dependiendo de la situación de estudio.

Veamos la siguiente situación, supongamos que, en el examen parcial de Probabilidad y Estadística, se presentan situaciones donde las respuestas tienen solo dos opciones posibles, verdadero o falso. Cada una de las preguntas, es independiente de las demás que se incluyen en el examen. Para este caso, denominamos éxito a la selección correcta de la respuesta, y fracaso a la selección incorrecta de la respuesta. Además, se sabe que la probabilidad de responder correctamente es de 70%. Claramente, esta situación puede ser modelizada utilizando la Distribución de Bernoulli, donde $p = 0.7$ y para determinar la probabilidad de fracaso decimos que,

$q = 1 - 0.7 = 0.3$. Resulta entonces, que el fracaso en una prueba de Bernoulli, se determinar mediante la siguiente fórmula $q = 1 - p$

Como podemos ver, este modelo de distribución resulta bastante simple, pero es uno de los más importantes dentro del universo de las probabilidades.

Distribución Binomial

Este modelo de distribución de probabilidades podría considerarse una generalización del modelo de distribución de probabilidad de Bernoulli que hemos estudiado al inicio de la clase, o también podría considerarse que el modelo de distribución de Bernoulli es un caso particular del modelo de distribución de probabilidades Binomial.

Entonces, sabemos que *p* es la probabilidad de que un suceso ocurra en una sola prueba de Bernoulli (probabilidad de éxito). Entonces *q= 1 – p* es la probabilidad de que no ocurra en una sola prueba (probabilidad de fracaso). Así, la probabilidad de un evento ocurra *x* veces en *n* cantidad de pruebas es:

$$P(x) = P(X = x) = \binom{n}{x} . p^x . q^{n-x}$$

Se puede ver que en la fórmula aparece una expresión que podría llegar a ser des conocida, es $\binom{n}{x}$, lo que corresponde al número combinatorio de *n* elementos agrupados de a *x*.

Un experimento binomial presenta las siguientes propiedades:

1. Consiste en un número fijo, n, de pruebas idénticas.
2. Cada prueba resulta en uno de dos resultados: éxito, o fracaso.

3. La probabilidad de éxito en una sola prueba es igual a algún valor p y es el mismo de una prueba a la otra. La probabilidad de fracaso es igual a q = (1 – p).
4. Las pruebas son independientes.
5. La variable aleatoria de interés es Y, el número de éxitos observado durante las n *pruebas.*

Veamos a continuación, algunos ejemplos:

Ejemplo 1. Un sistema de detección de alarma temprana para aviones consta de cuatro unidades de radar idénticas que operan de manera independiente entre sí. Suponga que cada una tiene una probabilidad de 0.95 de detectar un avión intruso. Cuando un avión intruso entra en escena, la variable aleatoria de interés es Y que se define como el número de unidades de radar que no detecta el avión. ¿Es éste un experimento binomial?

Entonces, para decidir si éste es un experimento binomial, debemos determinar que los cinco requisitos enunciados anteriormente, se cumplan. Observemos que la variable aleatoria de interés Y, es el número de unidades de radar que no detecta un avión. La variable aleatoria de interés en un experimento binomial es siempre el número de éxitos; en consecuencia, este experimento puede ser binomial sólo si llamamos éxito al evento de no detectar. Ahora examinemos el experimento en cuanto a las cinco características del experimento binomial:

1. El experimento comprende cuatro pruebas idénticas; cada una de ellas consiste en determinar si una unidad particular de radar detecta (o no) el avión.
2. Cada prueba arroja uno de dos resultados. Como la variable aleatoria de interés es el número de éxitos.

3. Como todas las unidades de radar detectan el avión con igual probabilidad, la probabilidad de no detectar en cada prueba es la misma y p = P (no detectar) = 0.05.
4. Las pruebas son independientes porque las unidades operan de manera independiente.
5. La variable aleatoria de interés es Y, el número de éxitos en cuatro pruebas.

Entonces, el experimento es binomial con n = 4, p = 0.05 y q = 1 – 0.05 = 0.95.

Ejemplo 2. Supongamos que lanzamos una moneda equilibrada cinco veces consecutivas. Queremos calcular la probabilidad de obtener exactamente tres caras en esos cinco lanzamientos. Utilizaremos el modelo de distribución binomial para resolver este problema. Recordemos que, en cada lanzamiento de la moneda, tenemos dos resultados posibles: cara o cruz, y la probabilidad de obtener cara en un solo lanzamiento es del 50%. ¿Cuál es la probabilidad de que obtengamos exactamente tres caras en cinco lanzamientos de moneda?

Entonces, tenemos que $p = q = 0.5 = \frac{1}{2}$, luego aplicando la fórmula para la distribución binomial, se tiene que:

$$P(X = 3) = \binom{5}{3}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right)^{(5-3)} = \frac{5!}{3!\,(5-3)!} * \frac{1}{8} * \frac{1}{4} = \frac{10}{32} = 0.3125$$

Por lo tanto, la probabilidad de obtener tres caras en cinco lanzamientos es de 31,25%.

Ejemplo 3. Suponga que un lote de 5000 fusibles eléctricos contiene 5% de piezas defectuosas. Si se prueba una muestra de 5 fusibles, encuentre la probabilidad de hallar al menos uno defectuoso.

Es razonable suponer que Y, el número observado de defectuosos tiene una distribución binomial aproximada porque el lote es grande. Retirar unos cuantos fusibles no cambia lo suficiente la composición de los restantes como para preocuparnos. Entonces,

$$P(al\ menos\ uno\ defectuoso) = 1 - p(0) =$$

$$= 1 - \binom{5}{0} p^0 q^5 = 1 - 0.95^2 = 1 - 0.774 = 0.226$$

Observemos que hay una probabilidad más bien grande de ver al menos uno defectuoso, aun cuando la muestra sea muy pequeña.

Esperanza matemática y desvío estándar en la distribución binomial

Valiéndonos de algunos cálculos matemáticos y definiciones que hemos trabajado a lo largo del desarrollo de la presente cátedra, podemos obtener que la esperanza matemática **E(x)=*n*.p**, como ya sabemos, la esperanza matemática es el valor esperado al realizar el experimento (*n* veces para este caso). Además, se puede determinar que el desvío estándar, resulta del siguiente calculo, $\sigma = \sqrt{n.p.q}$

Ejercicios para practicar

1. Se construye un complejo sistema electrónico con cierto número de piezas de respaldo en sus subsistemas. Un subsistema tiene cuatro componentes idénticos, cada uno con una probabilidad de 0.2 de fallar en menos de 1000 horas. El subsistema va a operar si dos de los cuatro componentes están

operando. Suponga que los componentes operan de manera independiente. Encuentre la probabilidad de que:

   a. exactamente dos de los cuatro componentes duran más de 1000 horas.
   b. El subsistema opere más de mil horas.

2. La probabilidad de que un paciente se recupere de una enfermedad estomacal es 0.8. Suponga que se sabe que 20 personas han contraído la enfermedad. ¿Cuál es la probabilidad de que:
   a. exactamente 14 se recuperen?,
   b. al menos 10 se recuperen?,
   c. al menos 14 pero no más de 18 se recuperen?,
   d. a lo sumo 16 se recuperen?
3. Un examen de opción múltiple tiene 15 preguntas, cada una con cinco posibles respuestas, sólo una de las cuales es correcta. Suponga que uno de los estudiantes que hace el examen contesta cada una de las preguntas con una adivinación aleatoria independiente. ¿Cuál es la probabilidad de que conteste correctamente al menos diez preguntas?

La Distribución de Poisson

Hemos trabajado hasta ahora, tanto con variables aleatorias continuas (cuando trabajamos con la distribución normal y normal estandarizadas), como con variables aleatorias discretas (en la distribución de Bernoulli y la distribución Binomial). Ahora vamos a considera una variable discreta, pero con un soporte continuo. Esto es, por ejemplo, cuando queremos estudiar el número de accidentes automovilísticos ocurridos en la ruta 3 para el tramo Rio Grande – Ushuaia en la última semana de la temporada invernal 2023.

Claramente, se pueden distinguir la característica discreta y el soporte continuo en la definición de la situación de estudio.

A primera vista esta variable aleatoria, el número de accidentes, no parece estar relacionada con una variable aleatoria binomial, pero veremos que existe una relación interesante. Consideremos el periodo, una semana en este ejemplo, como dividido entre *n* subintervalos, cada uno de los cuales es tan pequeño que a lo sumo un accidente podría ocurrir en él con probabilidad diferente de cero. Denotando con *p* la probabilidad de un accidente en cualquier subintervalo, tenemos, para todos los fines prácticos,

- P(no ocurren accidentes en un subintervalo) = 1 – p,
- P(ocurre un accidente en un subintervalo) = p,
- P(ocurre más de un accidente en un subintervalo) = 0.

Entonces el número total de accidentes en la semana es precisamente el número total de subintervalos que contienen un accidente. Si la ocurrencia de accidentes puede ser considerada como independiente de un intervalo a otro, el número total de accidentes tiene una distribución binomial.

Aun cuando no hay una forma única de seleccionar los subintervalos y por tanto no conocemos ni *n* ni *p*, parece razonable que cuando dividimos la semana en un número mayor de n subintervalos, disminuye la probabilidad *p* de un accidente en uno de estos subintervalos más cortos.

Cuando en una distribución binomial de probabilidades, *n* es un número grande y la probabilidad del evento x es cercana a 0; de modo que q = 1 – p, se aproxima a 1, la distribución tiene un enorme sesgamiento. En estos casos de utiliza la distribución de Poisson, denominada también de los "casos raros". Fue desarrollada por el matemático francés Poisson, en el siglo XIX. Es por tanto también, una distribución para variable aleatoria discreta.

Sea χ una variable aleatoria discreta que puede tomar valores $0 \leq \chi$ tal que la probabilidad de un valor cualquiera x de la variable aleatoria χ está dada por:

$$P(x) = P(\chi = x) = \frac{\lambda^x e^{-x}}{x!}$$

Los elementos enunciados en la fórmula son:

- P(χ=x) significa la probabilidad de que ocurra un evento determinado x.
- λ es una constante positiva dada que resulta de $\lambda = n.p,$ es el promedio de la distribución.
- e es el número de Euler.
- $x!$ Representa el factorial de x, que está dado por el siguiente producto x!=$x.(x-1)(x-2). \ldots$ .3.2.1

En este modelo de distribución de probabilidades, la esperanza matemática coincide con la varianza de la distribución, es decir, $E(x) = V(x) = n.p = \lambda$.

Veamos algunos ejemplos donde resulta útil este modelo de distribución de probabilidades:

Ejemplo 1. Supongamos que se han realizado anotaciones sobre la cantidad de imperfecciones por panel de madera que debían repararse antes de fletarlos al comprador. La distribución de la cantidad de imperfecciones, (como rayaduras, superficies sin pulir adecuadamente, etc), se aproximaba a la distribución de Poisson. Es decir, una cantidad apreciable de paneles no tenían imperfecciones, algunos tenían una falla, muy pocos tenían dos, y así sucesivamente.

La cantidad media de imperfecciones por panel se calculó en 0,5. Determinemos la probabilidad de no encontrar imperfecciones en los paneles de madera:

$$P(0) = \frac{0.5^0 . e^{-0.5}}{0!} = \frac{1. e^{-0.5}}{1} = 0.6065 = 60{,}65\%$$

Así también, podemos determinar la probabilidad de encontrar una imperfección, de manera análoga:

$$P(1) = \frac{0.5^1 . e^{-0.5}}{1!} = \frac{0.5. e^{-0.5}}{1} = 0.3033 = 30{,}33\%$$

Si se solicita determinar la probabilidad de encontrar hasta tres imperfecciones, entonces, hay sumar las probabilidades de encontrar tres, dos, una y ninguna imperfección. Esto porque la condición menciona hasta tres imperfecciones. Sin embargo, puede darse el caso de que se solicite determinar la probabilidad de encontrar más de cinco imperfecciones, entonces, para resolver la situación podríamos hacer uso del complemento. Esto es, sabiendo que la condición nos dice más de cinco, entonces:

$$P(x > 5) = 1 - P(x \leq 5) =$$

$$= 1 - [P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4) + P(x = 5)]$$

Solo restaría utilizar la fórmula para determinar cada una de las probabilidades individuales y operar para responder al interrogante que hemos propuesto.

Ejemplo 2. Se ha recibido un pedido de compra por 1.000 paneles. El departamento de costos debe calcular el costo total de reparar los paneles, antes de comenzar la tarea.

La experiencia anterior indicaba que la reparación de cada imperfección costó 10 centavos de dólar. El número de casos esperados, con:

0 imperfecciones es $\cong 606$ paneles, lo que resulta de hacer 1000*P(0)

1 imperfección es $\cong 303$ paneles

2 imperfecciones es $\cong 75$ paneles

3 imperfecciones es $\cong 12$ paneles

4 imperfecciones es $\cong 1$ panel

Entonces, los paneles con 0 imperfecciones no implicarán gastos. La reparación de los paneles con 1 imperfección costará $303 * 0{,}10 = 30{,}30$. La reparación de los paneles con 2 imperfecciones costará $75 * 0{,}20 = 15$. En el caso los paneles que tienen 3 imperfecciones, su reparación costará $12 * 0{,}30 = 3{,}60$. Luego, para reparar los paneles que tiene 4 imperfecciones, se gastará $1 * 0{,}40 = 0{,}40$.

Ahora, sumando todos los valores obtenidos, se tiene que para reparar todos los paneles se deben destinar un total de $U\$D\ 49{,}30$

Ejemplo 3. En la fábrica LongChat, 10% de las herramientas producidas en un proceso de fabricación determinado resultan defectuosas. Se quiere estudiar la probabilidad de que, en una muestra de 10 herramientas seleccionadas, resulten exactamente 2 defectuosas.

Podemos tomar dos alternativas para resolver esta situación. La primera de ellas es usar el modelo distribución binomial y, en ese caso, tenemos que

$$P(2) = \binom{10}{2} . 0{,}1^2 . 0{,}9^{10-2} = 0{,}1937 = 19{,}37\%$$

Decimos entonces que la probabilidad de que en una muestra de 10 herramientas hayan 2 defectuosa, es de 19,37%.

Ahora, usando el modelo de distribución de probabilidades de Poisson, se tiene que $\lambda = n.p$ por lo tanto, $\lambda = (10).(0{,}1) = 1$, entonces:

$P(2) = \frac{1^2 . e^{-1}}{2!} = 0{,}1839.$

Entonces, la probabilidad que, en una muestra de 10 herramientas, resulten 2 defectuosas es del 18,39%.

Como podemos notar existe una diferencia de 0,98% entre ambas probabilidades, esta diferencia se debe a las aproximaciones y suposiciones que hemos hecho para poder usar la distribución de Poisson. Recordemos que la distribución Binomial y la de Poisson, se relacionan, pero con características particulares de los tamaños de la población. Además, hemos podido observar que la situación no mencionaba la media de productos defectuosos de acuerdo con algún intervalo de tiempo, y por ello decimos que hubo suposiciones a las que se le asigna esa leve diferencia.

3. Actividad:

Analice las siguientes situaciones y resuelva.

a. Denote con X una variable aleatoria que tenga una distribución de Poisson con media l = 2. Encuentre
   i. P(X= 4)
   ii. $P(X \geq 4)$
   iii. $P(X < 4)$

b. Llegan clientes a un mostrador de salida en una tienda de departamentos de acuerdo con una distribución de Poisson, a un promedio de siete por hora. Durante una hora determinada, ¿cuáles son las probabilidades de que:
   - no lleguen más de tres clientes?,
   - lleguen al menos dos clientes?,
   - lleguen exactamente cinco clientes?

c. La variable aleatoria X tiene una distribución de Poisson y es tal que p(0) = p(1). ¿Cuál es p(2)?

d. Aproximadamente 4% de las obleas de silicio producidas por un fabricante tienen menos de dos defectos grandes. Si Y, el número de defectos por oblea tiene una distribución de Poisson, ¿qué proporción

de las obleas tiene más de cinco defectos grandes?

4. Actividad Integradora.

Habilitamos en el campus un cuestionario con las respuestas a cada uno de los puntos, para que una vez que resuelvan las actividades, puedan responder el cuestionario.

5. Cierre:

Llegamos al final de la clase 8. Hemos trabajado sobre tres modelos de distribución de probabilidades que resultan muy importantes ya que aparecen constantemente en el trabajo cuando analizamos determinados fenómenos.

Les deseamos buenas tareas, cualquier consulta pueden hacerla por los medios habilitados en los foros habilitados para consultas.

Saludos cordiales.

6. **Bibliografía:**

- ✓ Ruiz Diaz, G. (2019). Estadística y Probabilidad. Notas de cátedra. Universidad Nacional de Formosa.
- ✓ García, J; López, N; Calvo, J. (2011). Estadísticas Básicas para Estudiantes de Ciencias. Facultad de Ciencias Físicas Universidad Complutense de Madrid. España.
- ✓ Wackerly, D; Mendenhall, W; Schaeffer, R. (2010). Estadística Matemática con Aplicaciones. 7ma Ed. Cengage Learning. Santa Fe, México.