Otras distribuciones de variables discretas: Poisson-Hipergeométrica. Aproximación

En esta lectura seguimos estudiando otros modelos de distribución de probabilidades para variables discretas. Ahora, es el turno de la distribución de Poisson y la hipergeométrica. Se trata de que diferencies entre todas las distribuciones hasta aquí estudiadas por sus características. También se presenta una aproximación cuando un experimento binomial es engorroso hacerlo o no está en la tabla. Esto significa que todos los problemas tienen resolución. Y esto es muy alentador cuando un investigador o la persona que vaya a tomar decisiones tiene que elegir, de entre los métodos, el que más se adapte y, al mismo tiempo, sea correcto porque respeta las características del experimento o situación planteada. De esta manera, los errores por aproximación sean los mínimos. Pero la estrella de todas las distribuciones es la normal, que verás en la próxima lectura, no sin antes aprender las que se presentan aquí, cuyos modelos también son muy utilizados. Todo depende del experimento que tengas que realizar para después decidir. ¡Adelante, entonces!

Distribución de Poisson

Caso: Hostelaustral.com

Los hoteles de San Carlos de Bariloche reciben una gran cantidad de huéspedes por año. El sitio web www.hostelaustral.com registra los alojamientos de todos los hoteles del sur, especialmente de Bariloche. Este sitio tiene un promedio aproximado de siete visitantes por minuto. Muchos hoteles acuden a este sitio por las derivaciones que les hacen según las preferencias de los huéspedes. Se quiere hacer un estudio sobre este sitio web y tienes que averiguar lo siguiente:

1. Calcule la probabilidad de que no haya ningún visitante al sitio web en un lapso de un minuto.
2. De que haya dos o más visitantes al sitio web en un lapso de un minuto.
3. De que haya uno o más visitantes al sitio web en un lapso de 30 segundos.
4. De que haya cinco o más visitantes al sitio web en un lapso de un minuto.

Características de la distribución de Poisson

En esta lectura estudiarás otra distribución de variable aleatoria discreta que se utiliza para estimar la cantidad de veces que se da un hecho determinado —que se llaman ocurrencias— en un intervalo de tiempo o de espacio.

Ejemplos generales:

Hay muchos ejemplos, pero para que vayas interiorizándote de sus características, citaremos algunos:

En todos los casos, tendremos que calcular alguna probabilidad de la variable aleatoria x de:

1. Número de clientes que concurren a una entidad bancaria por hora.
2. Número de errores tipográficos por página.
3. Número de defectos que, por metro cuadrado, acusa una fábrica de mosaicos.
4. Número de llamadas que por minuto llegan a un conmutador.
5. Cantidad de cabezas de ganado por Ha.
6. Número de accidentes que en una esquina céntrica se producen diariamente.

Observa que, si bien la variable es discreta, resulta de un conteo. Está en función del tiempo o del espacio (que es una variable continua).

Además, es imposible contar los fracasos. El número de accidentes que no se produjeron no se pueden contar, así como el número de llamadas que no llegaron a la central telefónica.

Muchos autores suelen definir la distribución de Poisson como propia de casos no muy comunes o raros.

Las condiciones que deben darse para que sea una distribución de Poisson

1. La probabilidad de ocurrencia de un evento es la misma para cualquier intervalo de la misma magnitud.
2. El número de ocurrencias/no ocurrencias en cualquier intervalo es independiente del número de ocurrencias/no ocurrencias en cualquier otro intervalo.
3. La probabilidad de que el evento ocurra es proporcional a la longitud del intervalo.

4. La probabilidad de que uno o más eventos se presenten en un intervalo muy pequeño, es tan pequeña que puede despreciarse.

5. En realidad, es una binomial para *n* que tiende a infinito y *p* a cero.

Expresión matemática. Parámetros

Esta distribución tiene como parámetro un solo valor, el promedio de ocurrencia, al que denominaremos como *Lambda: λ*

La expresión matemática (función de densidad) correspondiente a esta distribución es:

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

P(x): Es la probabilidad de que ocurra un evento x dentro de un intervalo (probabilidad de tener exactamente x ocurrencias).

x: Número de éxitos/ocurrencias

Parámetros

λ: Promedio de ocurrencia o valor esperado en un intervalo

e: número irracional (e = 2,717281…)

Observa que el número de ocurrencias x, no tiene límite superior. Esta es una variable aleatoria discreta que toma los valores de una sucesión infinita de números (x = 0, 1, 2,…).

En esta distribución debe cumplirse que:

$$\sum P(x_i) = 1$$

Donde *i* varía, al menos teóricamente, entre cero e infinito. Decimos *teóricamente*, ya que, por ejemplo, si se informa que en promedio en una pista de aterrizaje de aviones despegan por hora tres aviones de cabotaje; no tendría sentido pretender determinar la probabilidad de que en una hora aterricen en dicho aeropuerto 90 aviones de cabotaje.

Valor esperado y varianza de la distribución de Poisson

$$E(x) = \lambda$$

En la distribución de Poisson, la media, que es igual al valor esperado, E(x) es igual a lo que se espera que ocurra, es decir λ.

$$Var(x) = \sigma\text{^}2 = \lambda$$

La varianza también es igual al promedio de ocurrencia λ. Por lo tanto, la desviación estándar es:

$$\sigma = \sqrt{\lambda}$$

Resolución del caso de Hostelaustral.com

A continuación, estudiarás como utilizar las tablas de Poisson y la fórmula resolviendo el caso propuesto al comienzo de la lectura. Al finalizar la lectura 4 de este módulo, brindamos un artículo en el que podrás ver las formas de todas las distribuciones al variar sus parámetros. Te recomiendo que, cuando llegues, las veas. Es muy interesante el enfoque empírico que allí se muestra.

Resolución punto a) del caso hostelaustral.com

a) Calcule la probabilidad de que no haya ningún visitante al sitio web en un lapso de un minuto.

Dato: $\lambda=7$ visitantes por minuto.

En primer lugar, tienes que verificar que λ y x (la variable aleatoria x) se refieran a la misma unidad de tiempo en este caso. Como ambas se refieren a un minuto, aplicamos λ como está, no la cambiamos, ni transformamos a otra unidad.

Resolución por fórmula

x indica la cantidad de visitantes al sitio web. En nuestro caso: x=0 (recuerda que la probabilidad de éxito es la cantidad de visitantes que entran al sitio web), pues en este caso nos piden que cero visitantes entren al sitio.

$x=0$

λ=7

El parámetro es λ y el número e, que ya lo conocemos, lo podemos calcular por calculadora. Es irracional con infinitas cifras decimales no periódicas. Por eso, tendremos que aproximarlo. Mi consejo es que hagas todos los cálculos directamente sin poner valores parciales en las cuentas.

Entonces:

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

Sustituimos:

$$P(x = 0) = \frac{e^{-7}.7^{0}}{0!}$$

Recuerda que *7^0 = 1 y 0! = 1* y , entonces

P (x=0) = 0,0009 o 9 x 10-4 en notación científica (que te aconsejo que la practiques).

Nota: si tienes una calculadora científica simple el cálculo para lo haces así: SHIFT Ln -7

Eso te muestra e-7, la calculadora toma al -7 como exponente.

La probabilidad de que no haya ningún visitante al sitio web en un lapso de un minuto es 0,0009.

Resolución por tabla de probabilidad puntual de Poisson

Tabla 1: Porción de la tabla de Poisson puntual para el caso Hostelaustral.com

Descripción de la tabla 1: Esta tabla muestra el valor puntual de la probabilidad de Poisson para x=0 y λ=7.

	λ									
X	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	0.0022	0.0020	0.0018	0.0017	0.0015	0.0014	0.0012	0.0011	0.0010	0.0009
1	0.0137	0.0126	0.0116	0.0106	0.0098	0.0090	0.0082	0.0076	0.0070	0.0064
2	0.0417	0.0390	0.0364	0.0340	0.0318	0.0296	0.0276	0.0258	0.0240	0.0223
3	0.0848	0.0806	0.0765	0.0726	0.0688	0.0652	0.0617	0.0584	0.0552	0.0521
4	0.1294	0.1249	0.1205	0.1162	0.1118	0.1076	0.1034	0.0992	0.0952	0.0912
5	0.1579	0.1549	0.1519	0.1487	0.1454	0.1420	0.1385	0.1349	0.1314	0.1277
6	0.1605	0.1601	0.1595	0.1586	0.1575	0.1562	0.1546	0.1529	0.1511	0.1490
7	0.1399	0.1418	0.1435	0.1450	0.1462	0.1472	0.1480	0.1486	0.1489	0.1490
8	0.1066	0.1099	0.1130	0.1160	0.1188	0.1215	0.1240	0.1263	0.1284	0.1304
9	0.0723	0.0757	0.0791	0.0825	0.0858	0.0891	0.0923	0.0954	0.0985	0.1014

Fuente: elaboración propia.

Esta es la parte 3 de la tabla que está en el texto básico. Fíjate que para encontrarla tienes que buscar el valor de lambda que no siempre está en el primer cuerpo de la tabla. A la izquierda buscas el valor de la variable aleatoria, que llega hasta cierto valor de n. Dijimos que n tiende a infinito; es decir, el valor de λ va determinando la cantidad de ensayos, que vemos que son más a medida que lambda aumenta.

La probabilidad de que no haya ningún visitante al sitio web en un lapso de un minuto es 0,0009.

Puedes verlo por la tabla acumulada que da lo mismo. Aquí no tenemos que hacer ninguna transformación según la tabla que se use.

Resolución punto b) del caso de Hostelaustral.com

b) De que haya dos o más visitantes al sitio web en un lapso de un minuto.

Dato: λ=7 visitantes por minuto.

x=variable aleatoria que nos indica la cantidad de visitantes al sitio web. En este caso, nos piden en un lapso de un minuto.

Son coherentes x y λ, entonces, adelante.

Resolución por fórmula

En este caso, no se pide una probabilidad puntual, sino un intervalo que va de dos a más infinito; en realidad, no sabemos la cantidad de ensayos *n*.

Por lo que no podemos calcular todas las probabilidades mayores o iguales a 2. Lo más sencillo es hacerlo por complemento.

Lo que sí sabemos es que la sumatoria de todas las probabilidades de cada variable discreta es igual a 1. Entonces, podemos escribir:

$$PP\ (x\geq 2) = 1 - P(x<2) = 1 - (P(x=0) + P(x=1))$$

Por lo tanto, calcularemos la P(0) y P(1). Aplicamos la fórmula:

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

Sustituimos:

$$P(x=0) = \frac{e^{-7}.7^{0}}{0!} = 0{,}0009$$

calculada en el ítem a)

$$P(x=1) = \frac{e^{-7}.7^{1}}{1!} = \frac{e^{-7} \times 7}{1} = 0{,}0064$$

Entonces:

$$P(x \geq 2) = 1 - (0{,}0009 + 0{,}0064) = 1 - 0{,}0073 = 0{,}9927$$

Resolución por tabla de probabilidad puntual de Poisson

Si te fijas en la tabla 1 del punto anterior, tienes que sumar la probabilidad de x=1 y x=2 con un *lambda* de 7

$$P(x < 2) = 0{,}0009 + 0{,}0064 = 0{,}0073$$

No olvides que la tabla da las probabilidades puntuales y tú tienes que decidir qué sumar o restar para llegar a la probabilidad buscada.

Ahora, cómo nos piden:

$P(x \geq 2) = 1 - 0{,}0073 = 0{,}9927$, hay que restárselo a 1.

La probabilidad de que haya dos o más visitantes al sitio web en un lapso de un minuto es de 0,9927 que representa un 99,27 % de posibilidades. ¡Bastante bien!

Resolución punto c) del caso Hostelaustral.com

c) De que haya uno o más visitantes al sitio web en un lapso de 30 segundos.

Resolución por fórmula:

Dato: $\lambda=7$ visitantes *por minuto*.

La variable aleatoria x aquí tiene otro intervalo de tiempo, se pide $P(x \geq 1)$ en 30 segundos.

No hay coherencia entre x y λ, por lo que tenemos que cambiar lambda, pues sabemos que es proporcional al intervalo de tiempo en este caso.

Esta conversión es sencilla, porque justo es la mitad del tiempo y el resultado es de $\lambda=3,5$. Pero no siempre es así, por lo tanto, con una regla de tres simple, transformamos el λ. Si en 60 segundos entran 7 visitantes, en 30 segundos, entrarán *(30x7)/60 = 3,5* visitantes.

Antes de aplicar la fórmula pensemos qué tenemos que calcular:

$P(x \geq 1)$ Estamos en un caso parecido al del ítem b). Estarás de acuerdo con nosotros en que: $P(x \geq 1) = 1 - P(x=0)$

¡Cuidado!, no te tientes en decir que P(x=0). Ya la tienes calculada. Pues ahora lambda vale 3,5.

Sustituyamos en la fórmula:

$$P(x = 0) = \frac{e^{-3,5} . 3,5^0}{0!} = \frac{e^{-3,5} \times 1}{1} = 0,0302$$

Entonces:

$$P(x \geq 1) = 1 - P(x = 0)$$

$$P(x \geq 1) = 1 - 0,0302 = 0,9698$$

Resolución por tabla de probabilidad puntual de Poisson

Tabla 2: Porción de la tabla de Poisson puntual para el caso Hostelaustral.com

Descripción de la tabla 2: Esta tabla muestra el valor puntual de la probabilidad de Poisson para x=0 y λ=3,5.

	λ									
X	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	0.0450	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202	0.0183
1	0.1397	0.1304	0.1217	0.1135	0.1057	0.0984	0.0915	0.0850	0.0789	0.0733
2	0.2165	0.2087	0.2008	0.1929	0.1850	0.1771	0.1692	0.1615	0.1539	0.1465
3	0.2237	0.2226	0.2209	0.2186	0.2158	0.2125	0.2087	0.2046	0.2001	0.1954
4	0.1734	0.1781	0.1823	0.1858	0.1888	0.1912	0.1931	0.1944	0.1951	0.1954
5	0.1075	0.1140	0.1203	0.1264	0.1322	0.1377	0.1429	0.1477	0.1522	0.1563
6	0.0555	0.0608	0.0662	0.0716	0.0771	0.0826	0.0881	0.0936	0.0989	0.1042
7	0.0246	0.0278	0.0312	0.0348	0.0385	0.0425	0.0466	0.0508	0.0551	0.0595

Fuente: elaboración propia.

Observa que la probabilidad es la misma que obtuvimos por fórmula, pero ¡¡cuidado!! Esta probabilidad no es la que se pide; hay que calcular lo siguiente de acuerdo con la pregunta:

$$P(x \geq 1) = 1 - P(x = 0) = 1 - 0{,}0302 = 0{,}9698$$

La probabilidad de que por lo menos un visitante ingrese al sitio web en un lapso de 30 segundos es 0,9698.

Resolución punto d) del caso Hostelaustral.com

d) De que haya cinco o más visitantes al sitio web en un lapso de un minuto.

Resolución por fórmula

En este caso λ y x son coherentes, se refieren a visitantes en un minuto. Por lo que λ=7

Por otro lado, pensemos qué calcular.

Estarás de acuerdo en que:

$$P(x \geq 5) = 1 - P(x < 5)$$

pues este último término es lo que nos da la tabla y, además, es más sencillo para calcular, ya que son menos cuentas.

Ahora:

$$P(x < 5) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4)$$

En los ítems a y b ya calculamos algunas de las probabilidades. No podremos tomar nada del c), dado que lambda tiene otro valor.

Entonces para una lambda igual a 7:

$$P(x = 0) = \frac{e^{-7}.7^0}{0!} = 0{,}0009$$
$$P(x = 1) = \frac{e^{-7}.7^1}{1!} = \frac{e^{-7} \times 7}{1} = 0{,}0064$$

Nos faltaría calcular el resto:

$$P(x = 2) = \frac{e^{-7}.7^2}{2!} = \frac{e^{-7} \times 49}{2 \times 1} = 0{,}0223$$

$$P(x = 3) = \frac{e^{-7}.7^3}{3!} = 0{,}0521$$

$$P(x = 4) = \frac{e^{-7}.7^4}{4!} = 0{,}0912$$

$$P(x < 5) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4)$$
$$P(x < 5) = 0{,}0009 + 0{,}0064 + 0{,}0223 + 0{,}0521 + 0{,}0912 = 0{,}1729$$

Entonces: $P(x \geq 5) = 1 - P(x < 5) = 1 - 0{,}1729 = 0{,}8271$

Resolución por tabla de probabilidad puntual y acumulada de Poisson

Si nos fijamos en la tabla 1, allí están todas las probabilidades que necesitamos. Tienes que sumar las probabilidades de la columna de λ=7, desde x=0 hasta x=4 inclusive.

Lo que te va a dar esa suma, ya sabes que es la probabilidad de que x sea menor a 5.

$$P(x < 5) = 0{,}0009 + 0{,}0064 + 0{,}0223 + 0{,}0521 + 0{,}0912 = 0{,}1729$$

Luego:

$$P(x \geq 5) = 1 - P(x < 5) = 1 - 0{,}1729 = 0{,}8271$$

Nota: Si aprendes a utilizar la tabla acumulada de Poisson, estos cálculos son mucho más fáciles. Claro que tienes que saber qué buscar, como en este caso. Si buscas x=4, para λ=7 te dará directamente:

$$P(x < 5) = P(x \leq 4) = 0{,}17$$

Hay tablas que solo aproximan a tres cifras decimales. Observa que el resultado es el mismo; solo tienes que restarlo a 1 para obtener el resultado final.

Tabla 3: Porción de la tabla de Poisson acumulada para el caso Hostelaustral.com

Descripción de la tabla 3: Esta tabla muestra el valor acumulado de la probabilidad de Poisson para x=4 y λ=7

X	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	0.0022	0.0020	0.0018	0.0017	0.0015	0.0014	0.0012	0.0011	0.0010	0.0009
1	0.0159	0.0146	0.0134	0.0123	0.0113	0.0103	0.0095	0.0087	0.0080	0.0073
2	0.0577	0.0536	0.0498	0.0463	0.0430	0.0400	0.0371	0.0344	0.0320	0.0296
3	0.1425	0.1342	0.1264	0.1189	0.1119	0.1052	0.0988	0.0928	0.0871	0.0818
4	0.2719	0.2592	0.2469	0.2351	0.2237	0.2127	0.2022	0.1920	0.1823	0.1730
5	0.4298	0.4141	0.3988	0.3837	0.3690	0.3547	0.3407	0.3270	0.3137	0.3007
6	0.5902	0.5742	0.5582	0.5423	0.5265	0.5108	0.4953	0.4799	0.4647	0.4497
7	0.7301	0.7160	0.7018	0.6873	0.6728	0.6581	0.6433	0.6285	0.6136	0.5987
8	0.8367	0.8259	0.8148	0.8033	0.7916	0.7796	0.7673	0.7548	0.7420	0.7291
9	0.9090	0.9016	0.8939	0.8858	0.8774	0.8686	0.8596	0.8502	0.8405	0.8305
10	0.9531	0.9486	0.9437	0.9386	0.9332	0.9274	0.9214	0.9151	0.9084	0.9015
11	0.9776	0.9750	0.9723	0.9693	0.9661	0.9627	0.9591	0.9552	0.9510	0.9467

Fuente: elaboración propia.

$$P(x \geq 5) = 1 - P(x < 5) = 1 - 0{,}173 = 0{,}827$$

El resultado es el mismo que el obtenido por fórmula.

La probabilidad de que haya cinco o más visitantes al sitio web en un lapso de un minuto es de 0,8271.

Lectura obligatoria: ¿Cómo utilizar la tabla de la distribución binomial?

Indicaciones sobre la lectura: Este artículo explica cómo manejar la tabla de la distribución de Poisson de probabilidades acumuladas y lleva a reflexionar sobre algunos casos extremos en los cuales no se encuentra la probabilidad directa. Se recomienda la lectura de las siguientes páginas: 4-6 (inclusive). También explica la aproximación de la binomial a Poisson, que explicaremos con más detalle al final de la lectura.

PDF

Taules.pdf

46.2 KB

Fuente: ¿Cómo utilizar la tabla de la distribución binomial? (s. f.). Recuperado de http://www-eio.upc.edu/teaching/estad/MC/taules/com-usar-taules.pdf

Lectura obligatoria: Distribución de Poisson

Indicaciones sobre la lectura: En esta publicación se explican los conceptos y características principales de la distribución de Poisson. Se resuelven problemas tipo. Es muy útil para afianzar los conceptos desarrollados hasta aquí. Además, tiene un programa interactivo para ver las distintas formas de la distribución cuándo se cambian los parámetros.

PDF

Distribución de Poisson.pdf
245.6 KB

Fuente: Cañas Escamilla, J. J. y Galo Sánchez, J. R. (s. f). Distribución de Poisson. Recuperado de https://proyectodescartes.org/iCartesiLibri/materiales_didacticos/EstadisticaProbabilidadInferencia/VAdiscreta/4_2DistribucionPoisson.html

Tablas

Lectura obligatoria: Apéndice tabla 4(b). Valores directos para determinar probabilidades de Poisson

Indicaciones sobre la lectura: Te facilitamos las tablas de probabilidades puntuales de las distribuciones de Poisson. Te serán de utilidad para rendir los exámenes. Puedes bajarlas y fotocopiarlas. Estas están en el texto básico.

Apéndice Tabla 4(b).pdf
1.6 MB

Fuente: Levin, R. y Rubin, D. (2012). Estadística para Administración y Economía. Ciudad de México, MX: Pearson.

Lectura obligatoria: Tablas de la función de distribución de Poisson acumuladas

Indicaciones sobre la lectura: Aquí te facilitamos las tablas de probabilidades acumuladas de las distribuciones de probabilidades de Poisson. Te serán de utilidad para rendir los exámenes. Puedes bajarlas y fotocopiarlas.

Tablas de la función de distribución de Poisson acumuladas.pdf
32.2 KB

Fuente: Cátedra: Probabilidad y estadística. Facultad Regional de Mendoza. UTN. (s. f.). Tabla D.4 Distribución de Poisson. Recuperado de http://www1.frm.utn.edu.ar/estadistica/documentos/td4_poissonacumulada.pdf

Los parámetros de la función de densidad de Poisson son λ y ____.

Escriba su respuesta aquí

SUBMIT

Distribución hipergeométrica

Caso: Máquinas de coser Romero S. A.

La empresa de máquinas de coser Romero S. A. es proveedora de máquinas de varios comercios de la zona centro de la provincia de Mendoza. En la última semana, entregó veinte máquinas. El departamento de Producción y Calidad da aviso de que hay dos máquinas que tienen un defecto en la bobina. Aunque las máquinas funcionan bien, estas bobinas no cumplen con las normas de calidad establecidas. El departamento de comercialización tiene que identificar las máquinas que ya fueron entregadas. Lo que hace es realizar las siguientes pruebas basadas en el cálculo de distribución de probabilidades para poder tomar una decisión, para lo cual se pregunta:

1. Si tomo una muestra aleatoria de cuatro, ¿cuál es la probabilidad de que todas las máquinas cumplan con los requisitos?
2. Si tomo una muestra aleatoria de cuatro, ¿cuál es la probabilidad de encontrar a lo sumo una máquina que no cumpla los requisitos?
3. Si la muestra es de cinco máquinas, ¿cuál es la probabilidad de encontrar a lo sumo una máquina que no cumpla los requisitos?

Distribución hipergeométrica. Características

Introducción

Hasta ahora hemos analizado distribuciones consistentes en experimentos en los cuales la probabilidad de éxitos/fracasos se mantiene constante. Además, la ocurrencia de un suceso no interfería en la ocurrencia de otro. Por eso, esta distribución viene a dar solución a los casos en los que no hay reposición de elementos en un experimento y por lo tanto la probabilidad de ocurrencia no se mantiene constante.

Cuando estudiemos las características de esta distribución y la ejemplifiquemos, notaremos que la población no se considera infinita frente a la muestra. Es más, población y muestra están cercanas en tamaño.

Ejemplo

Consideremos un curso de 40 alumnos de los cuales 15 de ellos han aprobado el parcial de Estadística. Si elegimos al azar un alumno, puedo tener dos situaciones: que el alumno haya aprobado o que el alumno no haya aprobado. La probabilidad de que haya aprobado Estadística es P (A) = 15/40

Ahora bien, si de la población se extrae una muestra de 10 alumnos y el muestreo se realiza sin reposición, la probabilidad de elegir un nuevo alumno, azarosamente, de la división y que haya aprobado el parcial de Estadística, ya no será P (A) = 15/40, pues esta probabilidad se ha visto modificada por el muestreo previo realizado.

Pero, aparentemente, el fenómeno presenta características similares a la de una binomial, en la que hay dos opciones: éxito/fracaso.

La diferencia está en lo siguiente:

1. Que la población es finita y el tamaño de la muestra es grande, significativo, respecto al tamaño poblacional y, por lo tanto, modifica la probabilidad de elegir nuevamente a un alumno que aprobó el parcial.

2. El muestreo se realiza sin reposición y, al ser significativo el tamaño de la muestra con respecto al tamaño de la población, modifica la probabilidad de acierto.

Debemos recordar que, en la distribución binomial, la probabilidad de lo que consideramos acierto se mantiene constante y esto solo se logra si el muestreo se realiza con reposición o la población es infinita o muy grande.

Retomemos el caso de los 40 alumnos. Si se toma una muestra aleatoria de 10 alumnos, puede ocurrir que en la muestra se tenga alguno de los siguientes: ninguno aprobado, un aprobado, dos aprobados… Hasta podrían estar los 10 alumnos elegidos aprobados. Estas consideraciones las vamos a ir tomando para definir las características de una distribución hipergeométrica, su variable aleatoria, su expresión matemática (función de densidad) y sus parámetros.

Características de una distribución hipergeométrica

- El proceso consta de n pruebas separadas, de entre un conjunto de N pruebas posibles.
- En cada una de las pruebas se dan dos resultados posibles mutuamente excluyentes: éxito (E) o fracaso (F).
- Los ensayos son dependientes.
- Las probabilidades de obtener un éxito o un fracaso varían en las sucesivas pruebas, dependiendo de los resultados anteriores.
- La población es finita y el tamaño de la muestra es grande; para establecer un límite, se aplica la distribución hipergeométrica si la muestra es mayor al 5 % de la población: $n > 5\% \, N$

Expresión matemática. Parámetros

Si denotamos con:

N: número de elementos de la población (finita)

k: número de éxitos en la población

n: número de ensayos (muestra)

x: variable aleatoria, éxitos que se pretende encontrar en la muestra

Entonces:

N-k: n.º de fracasos de la población

n-x: n.º de fracasos en la muestra o en los ensayos.

La probabilidad de que en una muestra aleatoria de n elementos, seleccionados sin reemplazo, se tengan x éxitos de los k éxitos que hay en la población y n-x fracasos de los N-k fracasos de la población, se calcula con la siguiente fórmula:

$$P(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

que proporciona la probabilidad de tener x éxitos en una muestra de tamaño n (utilizaremos la notación de número combinatorios:

$$C_{m,n} = \binom{m}{n}$$

que significa lo mismo: combinación de m elementos tomados de a n).

Ahora volvamos al ejemplo:

N=40 alumnos

k=15 aprobados

n=10 alumnos

y supongamos que queremos calcular la probabilidad de que en la muestra haya exactamente 2 alumnos que hayan aprobado el parcial, entonces:

$$P(x=2)=\frac{\binom{15}{2}\binom{40-15}{10-2}}{\binom{40}{10}}=\frac{\binom{15}{2}\binom{25}{8}}{\binom{40}{10}}=0{,}134$$

Existe una probabilidad de 0,134 de que, en la muestra, haya exactamente 2 alumnos que hayan aprobado el parcial de estadística.

Parámetros

N: número de elementos de la población (finita)

k: número de éxitos en la población

n: número de ensayos (muestra)

Valor esperado (media) y varianza de la distribución hipergeométrica

$$E(x) = \mu = n.\left(\frac{k}{N}\right)$$

$$Var(x) = \sigma^2 = n.\left(\frac{k}{N}\right)\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Tablas

- Existen tablas para calcular la distribución de una variable aleatoria en un experimento hipergeométrico, pero no vamos a necesitarlas en este curso.
- Existen tablas para calcular la distribución de una variable aleatoria en un experimento hipergeométrico, pero no vamos a necesitarlas en este curso.
    - La sumatoria de todas las probabilidades puntuales de cada variable aleatoria es igual a 1.
    - Para calcular la probabilidad de una o varias variables aleatorias puede recurrirse al complemento si es necesario.
- Usar calculadora científica para resolver las combinaciones.

Lectura obligatoria: Distribución hipergeométrica

Indicaciones sobre la lectura: En esta publicación, se explican los conceptos y características principales de la distribución hipergeométrica. Se resuelven problemas tipo. Es muy útil para afianzar los conceptos desarrollados hasta aquí. Además, tiene un programa interactivo para ver las distintas formas de la distribución cuándo se cambian los parámetros.

Distribución Hipergeométrica.pdf
234.6 KB

Fuente: Cañas Escamilla, J. J. y Galo Sánchez, J. R. (s. f). Distribución de Poisson. Recuperado de https://proyectodescartes.org/iCartesiLibri/materiales_didacticos/EstadisticaProbabilidadInferencia/VAdiscreta/4_1DistribucionHipergeometrica/index.html

Resolución del caso máquinas de coser Romero S. A.

Resolución del caso punto a)

a) Si tomamos una muestra aleatoria de 4, ¿cuál es la probabilidad de que todas las máquinas cumplan con los requisitos?

DATOS:

N= 20

k= 2 con defectos

n=4

La variable aleatoria x podría ser 4, pero ¡¡cuidado!! Significa que las 4 bobinas de la muestra cumplen con los requisitos; por lo tanto, estamos hablando de la distribución de las bobinas que cumplen con los requisitos; sin embargo, el k (éxito de la población) se refiere a las bobinas que no cumplen con los requisitos.

Como tiene que haber coherencia con la variable aleatoria x y el éxito de la población k, tenemos dos opciones:

1. Tomar x=4 (bobinas de la muestra que cumplen) con k= 18 (bobinas de la población que cumplen).
2. Tomar x= 0 (cero bobinas no cumplen) con k= 2 (bobinas de la población que no cumplen).

Ambos resultados dan la misma probabilidad.

Tomemos el caso 2), que x=0 no cumplen. Utilizo x=0 con k=2

$$P(x=0)=\frac{\binom{2}{0}\binom{20-2}{4-0}}{\binom{20}{4}}$$

$$P(x=0)=\frac{\binom{2}{0}\binom{18}{4}}{\binom{20}{4}}=0{,}6316$$

que es la probabilidad de que todas las bobinas cumplan con los requisitos (es decir, si tomo una al azar tengo una probabilidad de 0,6316 o bien un 63,16 % de que cumplan con los requisitos o que ninguna (cero) cumpla con los requisitos.

La probabilidad de ninguna de las máquinas cumpla con los requisitos o que todas las máquinas cumplan con los requisitos es de 0,6316.

Resolución del caso punto b)

Si tomo una muestra aleatoria de 4, ¿cuál es la probabilidad de encontrar a lo sumo una máquina que no cumpla los requisitos?

DATOS:

N= 20

k= 2 con defectos

n=4

x≤2 no cumpla, sea defectuoso.

P (x ≤ 1) = P (x = 0) + p (X = 1) con k=2

P (x = 0) ver ítem a)

$$P(x=1)=\frac{\binom{2}{1}\binom{20-2}{4-1}}{\binom{20}{4}}=\frac{\binom{2}{1}\binom{18}{3}}{\binom{20}{4}}=0{,}3368$$

$$P(x\leq 1)=P(x=0)+P(x=1)=0{,}6316+0{,}3368=0{,}9684$$

La probabilidad de que encontremos —como máximo— una máquina que no cumpla con los requisitos es de 0,9684.

Es una probabilidad alta, ya que es muy probable encontrar en la muestra de 4 máquinas que haya 1 o ninguna máquina que no cumplan con las condiciones.

Resolución del Caso punto c)

c) Si la muestra es de 5 máquinas, ¿cuál es la probabilidad de encontrar a lo sumo una máquina que no cumpla los requisitos?

Datos:

N= 20

k= 2 con defectos

n=5

x≤1 no cumpla, sea defectuoso.

$$P(x \leq 1) = P(x = 0) + P(x = 1)$$

$$P(x = 0) = \frac{\binom{2}{0}\binom{20-2}{5-0}}{\binom{20}{5}} = \frac{\binom{2}{0}\binom{18}{5}}{\binom{20}{5}} = 0{,}5526$$

$$P(x = 1) = \frac{\binom{2}{1}\binom{20-2}{5-1}}{\binom{20}{5}} = \frac{\binom{2}{1}\binom{18}{4}}{\binom{20}{5}} = 0{,}3947$$

$$P(x \leq 1) = P(x = 0) + P(x = 1) = 0{,}5526 + 0{,}3947 = 0{,}9473$$

La probabilidad de que encontremos —como máximo— una máquina que no cumpla con los requisitos en una muestra de 5 es de 0,9473.

Aproximación de la binomial a Poisson

PRINTCITY es una empresa que realiza impresiones y grabados en todo el país. Sus clientes la eligen siempre por el bajo porcentaje de errores en sus trabajos. Los errores de impresión en los trabajos que les piden los alumnos de arquitectura son del 0,1 %.

Actualmente tiene un pedido de 1500 trabajos. Calcula las siguientes probabilidades:

1. Ninguno de los trabajos presente errores.
2. Exactamente 10 trabajos presenten errores
3. Como máximo 2 trabajos presenten errores.

Resolución de la binomial por aproximación a la de Poisson

¿Cuándo aproximamos y por qué?

En muchas oportunidades, nos encontramos con distribuciones que responden al modelo binomial, con un p (probabilidad de éxito) pequeño y un n (número de ensayos) grande. En estos casos, el cálculo de las probabilidades de éxito se puede volver tedioso, especialmente, cuando n o p no se encuentra en la tabla. A los efectos de su resolución, puede aplicarse la distribución de Poisson.

Condiciones

Para poder aproximar mediante la distribución de Poisson una distribución binomial, se deben cumplir las siguientes condiciones:

- Ser un experimento binomial
- Con $n \geq 20$ y $p \leq 0{,}05$

Parámetros

λ=n.p pues es la media de la binomial.

Resolución del caso

Resolución caso punto a)

Datos: n=1500 y p=0,001 con errores

a) Ninguno de los trabajos presenta errores.

$\lambda = n.p = 1500 \times 0{,}001 = 1{,}5$ con errores. Con $x=0$ con errores

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

$$P(x=0) = \frac{e^{-1{,}5}.1{,}5^{0}}{0!} = 0{,}2231$$

La probabilidad de que ninguno de los trabajos presente errores es 0,2231.

Resolución caso punto b)

b) Exactamente 10 trabajos presentan errores

n=1500 *p*=0,001 con errores. Pregunta x=10 con errores.
Por lo que λ = n × q = 1500 × 0,001 = 1,5

$$P(x = 10) = \frac{e^{-1,5}.\,1,5^{10}}{10!} = 0$$

La probabilidad de elegir exactamente 10 trabajos sin errores es 0.

Resolución caso punto c)

c) Como máximo 2 trabajos tengan errores.

n=1500 *p*=0,001 con errores. Pregunta x ≤ 2 con errores.
Por lo que λ = n × p = 1500 × 0,001 = 1,5

La probabilidad de que como máximo haya 2 trabajos con errores es:

$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$
$P(x = 0) = 0{,}2231$ ver punto a)

$$P(x = 1) = \frac{e^{-1,5}.\,1,5^{1}}{1!} = 0{,}3347$$

$$P(x = 2) = \frac{e^{-1,5}.\,1,5^{2}}{2!} = 0{,}2510$$

$$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2) = 0{,}2231 + 0{,}3347 + 0{,}2510 = 0{,}8088$$

La probabilidad de que como máximo haya 2 trabajos con errores es de 0,8088. Por supuesto, que esto cambiaría si la muestra fuera menor.

Lectura obligatoria: Problemas resueltos de variables aleatorias discretas

Indicaciones sobre la lectura: En esta publicación se plantean ejercicios de variables aleatorias discretas con sus soluciones, puede ser un buen recurso para practicar

Problemas resueltos de variables aleatorias discretas.pdf
434.4 KB

Fuente: Cañas Escamilla, J. J. y Galo Sánchez, J. R. (s. f). Problemas resueltos. Recuperado de https://proyectodescartes.org/iCartesiLibri/materiales_didacticos/EstadisticaProbabilidadInferencia/VAdiscreta/5PROBLEMAS.html

Considere una distribución de Poisson en que la media es de dos ocurrencias por un período de tiempo.

a) ¿Cuál es el número esperado de ocurrencias en tres periodos de tiempo?

b) Calcule la probabilidad de dos ocurrencias en un periodo de tiempo.

c) Calcule la probabilidad de 6 ocurrencias en 3 periodos de tiempo.

- ○ a) E(x)=2 b) P(x=2)=0,2707 c)P(x=6)=0,1606
- ○ a) E(x)=6 b) P(x=2)=1 c)P(x=6)=0,1606
- ○ a) E(x)=6 b) P(x=2)=0,2707 c)P(x=6)=0,1606
- ○ a) E(x)=6 b) P(x=2)=0,2707 c)P(x=6)=0,4462

SUBMIT

Referencias

Anderson, David R., Sweeney, Dennis J. y Williams, Thomas A. (2008). *Estadística para Administración y Economía*. Ciudad de México, MX: Cengage Learning Editores, S. A.

Dirección General de Estadística y Censos-Ministerio de Economía y Finanzas GCBA. (s. f.) ETOI. 3.o trimestre de 2019. Recuperado de https://www.estadisticaciudad.gob.ar/eyc/wp-content/uploads/2019/12/ir_2019_1415.pdf

Levin, R. y Rubin, D. (2012). *Estadística para Administración y Economía*. Ciudad de México, MX: Pearson.

Morettini, M. (2013). Aproximaciones de distribuciones de probabilidad: enfoque empírico. Recuperado de http://nulan.mdp.edu.ar/2040/1/morettini.2013.pdf

Sangaku S. L. (2020). La distribución normal o Gaussiana. Recuperado de https://www.sangakoo.com/es/temas/la-distribucion-normal-o-gaussiana

Wiper, M. (2016). Variables continuas: la distribución normal. Recuperado de http://halweb.uc3m.es/esp/Personal/personas/mwiper/docencia/Spanish/GC/guardia%20civil/lecture%20notes/class15.pdf