Trabajo práctico 7: Variables aleatorias y distribuciones

AL TERMINAR ESTA GUÍA

- Calcular la esperanza, la varianza y la función de distribución acumulada de una variable aleatoria discreta dada por su tabla.
- Reconocer cuándo una variable tiene distribución binomial o de Poisson y calcular probabilidades puntuales y acumuladas.
- Tipificar y usar la tabla de la normal estándar para calcular proporciones y percentiles.
- Aproximar la binomial con la normal, con corrección por continuidad, y saber cuándo la aproximación falla.

PARA EL DOCENTE

Los ejercicios 1 a 4 son para el encuentro práctico, con la tabla de la normal impresa; el 4 es el más importante y conviene dibujar la campana en cada consigna. El 5 necesita computadora: sirve para un encuentro virtual o como tarea.

El 6 es para discutir en grupos al final, o para abrir la clase siguiente: repasa las condiciones de cada modelo antes de empezar con muestreo.

EJERCICIO 1 Materias en el turno de examen

A mano

Sea X la cantidad de materias que rinde un estudiante de segundo año, elegido al azar, en el turno de examen de diciembre. Según los registros de la facultad, su función de probabilidad es:

Materias (x)	P(X = x)
0	0,14
1	0,30
2	0,29
3	?
4	0,10

a) ¿X es una variable aleatoria discreta o continua? Calculá la probabilidad que falta en la tabla.

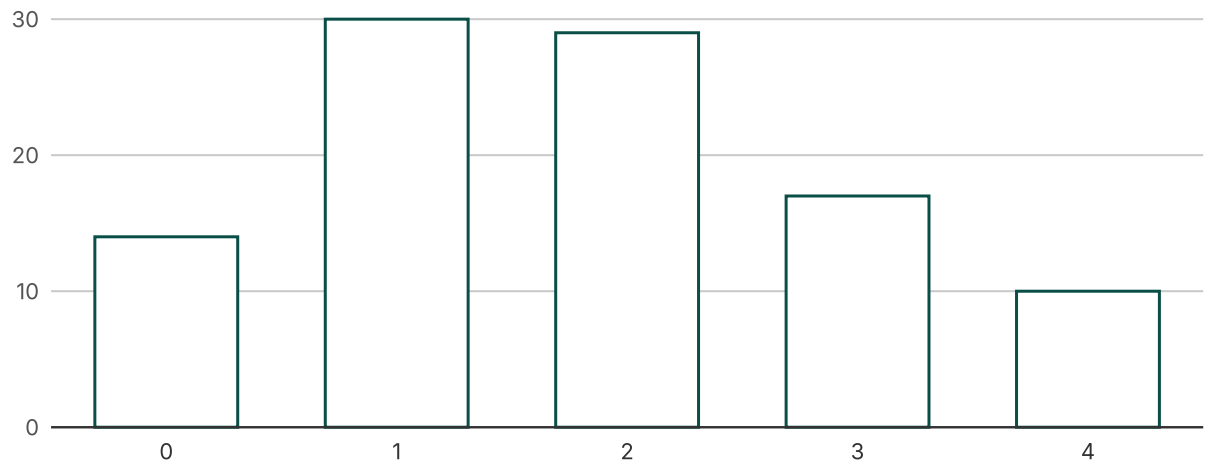
RESOLUCIÓN

Discreta: toma valores aislados que se pueden contar (0, 1, 2, 3 o 4 materias). Las probabilidades de todos los valores posibles tienen que sumar 1, así que la que falta sale por diferencia:

$$P(X = 3) = 1 - (0{,}14 + 0{,}30 + 0{,}29 + 0{,}10) = 0{,}17$$

b) Graficá la función de probabilidad y armá la función de distribución acumulada F(x) = P(X ≤ x). ¿Cuánto vale P(X ≤ 2)? ¿Y la probabilidad de que rinda dos materias o más?

RESOLUCIÓN



Eje vertical: probabilidad, en %.

x	P(X = x)	F(x)
0	0,14	0,14
1	0,30	0,44
2	0,29	0,73
3	0,17	0,90
4	0,10	1,00

$$P(X \leq 2) = F(2) = 0{,}73 \qquad P(X \geq 2) = 1 - F(1) = 1 - 0{,}44 = 0{,}56$$

La acumulada va sumando de izquierda a derecha y termina en 1. Para «dos o más» conviene restar del total lo contrario —cero o una materia— en lugar de sumar tres términos. Con una variable discreta, P(X ≥ 2) y P(X > 2) son distintas: la segunda deja afuera al 2.

c) Calculá la esperanza E(X). ¿Qué significa en este contexto, si nadie puede rendir una cantidad no entera de materias?

RESOLUCIÓN

$$E(X) = \sum x\, P(X = x) = 0 \cdot 0{,}14 + 1 \cdot 0{,}30 + 2 \cdot 0{,}29 + 3 \cdot 0{,}17 + 4 \cdot 0{,}10 = 1{,}79$$

Es el promedio de largo plazo: si se miraran muchísimos estudiantes, rendirían en promedio 1,79 materias cada uno. Como cualquier media, no tiene por qué ser un valor posible de la variable. Es el centro de gravedad del gráfico de barras.

d) Calculá la varianza y el desvío estándar de X.

RESOLUCIÓN

$E(X^2) = 0 \cdot 0{,}14 + 1 \cdot 0{,}30 + 4 \cdot 0{,}29 + 9 \cdot 0{,}17 + 16 \cdot 0{,}10 = 4{,}59$

$V(X) = E(X^2) - [E(X)]^2 = 4{,}59 - 1{,}79^2 = 1{,}3859 \qquad \sigma = \sqrt{1{,}3859} = 1{,}177$

La fórmula abreviada evita calcular cinco desvíos al cuadrado. El desvío está en materias, como X, y es lo que se interpreta: cuánto se aparta típicamente un estudiante del promedio.

e) La inscripción al turno cuesta $5.000 fijos más $2.000 por cada materia. ¿Cuánto paga en promedio un estudiante? ¿Y cuál es el desvío estándar de lo que paga?

RESOLUCIÓN

$Y = 5.000 + 2.000\,X \qquad E(Y) = 5.000 + 2.000 \cdot 1{,}79 = 8.580 \qquad \sigma_Y = 2.000 \cdot 1{,}177 = 2.354$

Con una transformación lineal no hace falta armar la tabla de Y: la esperanza se transforma igual que la variable, y el desvío solo se multiplica por la pendiente. El monto fijo corre todo el gráfico, pero no cambia cuánto se dispersa.

EJERCICIO 2 Turnos que se pierden

A mano

En un centro de salud, el 15 % de los pacientes con turno no se presenta. Un consultorio da 10 turnos por mañana, a pacientes que deciden si van o no independientemente unos de otros.

Sea X la cantidad de pacientes que faltan en una mañana.

a) ¿Por qué X tiene distribución binomial? ¿Cuáles son sus parámetros?

RESOLUCIÓN

Se cumplen las cuatro condiciones: hay una cantidad fija de ensayos (10 turnos); cada uno tiene dos resultados (falta o va); la probabilidad de faltar es la misma para todos (0,15); y los pacientes deciden independientemente. X cuenta los «éxitos» —acá, las ausencias—, así que X ~ B(10; 0,15). Cada turno, por separado, es una variable de Bernoulli.

b) ¿Cuál es la probabilidad de que no falte nadie? ¿Y de que falten exactamente dos?

RESOLUCIÓN

$P(X = 0) = 0{,}85^{10} = 0{,}1969$

$P(X = 2) = \binom{10}{2} 0{,}15^2\, 0{,}85^8 = 45 \cdot 0{,}0225 \cdot 0{,}2725 = 0{,}2759$

El coeficiente binomial cuenta de cuántas maneras se pueden elegir los dos turnos que fallan entre los diez; cada una de esas maneras tiene la misma probabilidad.

c) ¿Cuál es la probabilidad de que falten como mucho dos pacientes? ¿Y de que falten tres o más?

RESOLUCIÓN

$P(X \leq 2) = P(0) + P(1) + P(2) = 0{,}1969 + 0{,}3474 + 0{,}2759 = 0{,}8202$

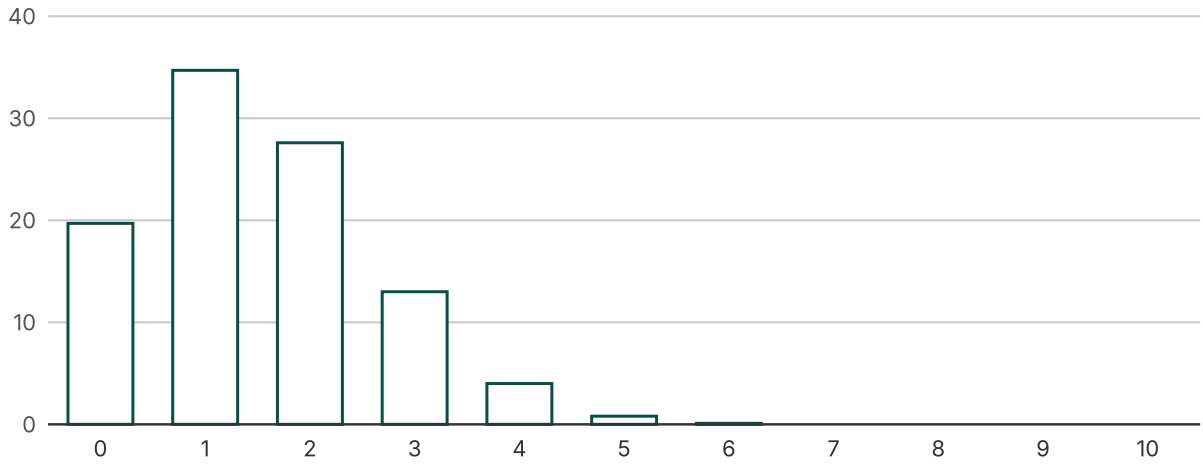
$P(X \geq 3) = 1 - P(X \leq 2) = 0{,}1798$

Para «tres o más» habría que sumar ocho términos; el complemento resuelve con tres. Los mismos valores se leen en la tabla de la binomial acumulada, en la fila n = 10.

d) ¿Cuántos pacientes se espera que falten por mañana? ¿Con qué desvío estándar?

RESOLUCIÓN

$E(X) = np = 10 \cdot 0{,}15 = 1{,}5 \qquad \sigma = \sqrt{np(1-p)} = \sqrt{10 \cdot 0{,}15 \cdot 0{,}85} = 1{,}129$



Eje vertical: probabilidad, en %.

En promedio faltan 1,5 pacientes por mañana. El gráfico muestra la distribución completa: asimétrica a la derecha, porque p es chico y X no puede bajar de cero.

e) Para aprovechar las ausencias, el consultorio decide dar 11 turnos aunque solo puede atender a 10. ¿Cuál es la probabilidad de que vengan los 11 y alguien se quede sin atención?

RESOLUCIÓN

$P(\text{vienen los 11}) = 0{,}85^{11} = 0{,}1673$

Ahora son 11 ensayos y el problema aparece solo si no falta nadie. Pasa en 16,7 % de las mañanas: es la cuenta que hacen las aerolíneas cuando venden más pasajes que asientos. Si el ausentismo bajara, ese riesgo subiría.

EJERCICIO 3 Llegadas a la guardia

A mano

A la guardia de un hospital llegan, entre las 2 y las 6 de la mañana, 2,5 pacientes por hora en promedio. Las llegadas son independientes entre sí y el promedio se mantiene estable en esas cuatro horas.

Sea X la cantidad de pacientes que llegan en una hora.

a) ¿Por qué es razonable usar la distribución de Poisson? ¿Cuánto valen la esperanza y la varianza de X?

> RESOLUCIÓN
>
> Se cuentan sucesos que ocurren al azar en un intervalo continuo (una hora), de a uno, independientes y con un ritmo promedio constante. No hay un número fijo de ensayos ni un máximo, como en la binomial: en teoría podrían llegar 20. En la Poisson la esperanza y la varianza son iguales al parámetro: E(X) = V(X) = λ = 2,5.

b) ¿Cuál es la probabilidad de que en una hora no llegue nadie? ¿Y de que lleguen exactamente tres pacientes?

> RESOLUCIÓN
>
> $$P(X = 0) = e^{-2,5} = 0,0821 \qquad P(X = 3) = \frac{e^{-2,5}\, 2,5^3}{3!} = 0,2138$$

c) ¿Cuál es la probabilidad de que lleguen como mucho dos pacientes? ¿Y más de cuatro?

> RESOLUCIÓN
>
> $$P(X \leq 2) = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right) = 0,5438$$
>
> $$P(X > 4) = 1 - P(X \leq 4) = 1 - 0,8912 = 0,1088$$
>
> «Más de cuatro» no incluye al 4: el complemento es «cuatro o menos». Los valores acumulados también se leen en la tabla de Poisson, en la columna λ = 2,5.

d) ¿Cuál es la probabilidad de que pasen dos horas seguidas sin que llegue ningún paciente?

> RESOLUCIÓN
>
> $$\lambda_{2\,\mathrm{h}} = 2 \cdot 2,5 = 5,0 \qquad P(Y = 0) = e^{-5,0} = 0,0067$$
>
> El parámetro es proporcional al largo del intervalo: en dos horas se esperan 5,0 llegadas. Da lo mismo que multiplicar 0,0821 · 0,0821, porque lo que pasa en horas distintas es independiente.

e) La guardia quiere tener personal para atender, en el 95 % de las horas, a todos los que llegan. ¿Para cuántos pacientes por hora tiene que estar preparada?

> RESOLUCIÓN

k	P(X ≤ k)
0	0,0821
1	0,2873
2	0,5438
3	0,7576
4	0,8912
5	0,9580

> Se busca el menor k con P(X ≤ k) ≥ 0,95: k = 5. Con 5 pacientes por hora alcanza en el 95,8 % de las horas; con 4, solo en el 89,1 %. Planificar para el promedio, 2,5, dejaría a la guardia desbordada muy seguido.

EJERCICIO 4 El peso al nacer

A mano

El peso de los recién nacidos a término en una maternidad sigue aproximadamente una distribución normal con media 3.320 g y desvío estándar 450 g.

Usá la tabla de la normal estándar: redondeá cada z a dos decimales.

a) Se considera bajo peso al nacer a menos de 2.500 g. ¿Qué proporción de los recién nacidos tiene bajo peso?

> RESOLUCIÓN
>
> $$z = \frac{2500 - 3320}{450} = -1{,}82 \qquad P(X < 2500) = P(Z < -1{,}82) = 0{,}0344$$
>
> Tipificar es contar a cuántos desvíos de la media está el valor: 2.500 g está 1,82 desvíos por debajo. Así cualquier normal se resuelve con una sola tabla, la de Z ~ N(0; 1). Un 3,4 % de los recién nacidos pesa menos de 2.500 g.

b) ¿Qué proporción pesa más de 4.000 g?

> RESOLUCIÓN
>
> $$z = \frac{4000 - 3320}{450} = 1{,}51 \qquad P(X > 4000) = 1 - P(Z < 1{,}51) = 1 - 0{,}9345 = 0{,}0655$$
>
> La tabla da el área a la izquierda; para la cola de la derecha se resta de 1. En una variable continua da igual «más de» que «4.000 o más»: la probabilidad de un valor exacto es cero.

c) ¿Qué proporción pesa entre 3.000 y 3.500 g?

> RESOLUCIÓN
>
> $$P(3000 < X < 3500) = P(-0{,}71 < Z < 0{,}40) = 0{,}6554 - 0{,}2389 = 0{,}4165$$
>
> Un intervalo es la diferencia entre dos áreas acumuladas. Dibujar la campana y sombrear el área antes de buscar en la tabla evita la mayoría de los errores de signo.

d) ¿Por debajo de qué peso está el 10 % de los recién nacidos más livianos? ¿Y a partir de qué peso está el 10 % más pesado?

> RESOLUCIÓN
>
> $$x_{0{,}10} = \mu + z\,\sigma = 3320 + (-1{,}28) \cdot 450 = 2.744{,}0 \qquad x_{0{,}90} = 3320 + 1{,}28 \cdot 450 = 3.896{,}0$$
>
> Ahora el camino es el inverso: se busca adentro de la tabla el área 0,10 (o la más cercana, 0,1003), se lee z = −1,28 en el margen y se destipifica. Por la simetría de la normal, el percentil 90 está a la misma distancia de la media, del otro lado.

e) En la maternidad nacen 1.200 bebés a término por año. ¿Cuántos se espera que nazcan con bajo peso?

> RESOLUCIÓN
>
> $1.200 \cdot 0{,}0344 = 41{,}3$
>
> Unos 41 por año. Es la esperanza de una binomial con n = 1.200 y p = 0,0344: la proporción de la normal pasa a ser la probabilidad de cada nacimiento. En la realidad la distribución del peso tiene la cola izquierda más pesada que la normal —por los prematuros y las complicaciones—, y por eso el modelo normal se usa para los nacidos a término.

EJERCICIO 5 Cuándo la binomial se parece a la normal

Con software

Una vacuna produce una reacción leve en el brazo (dolor o enrojecimiento) en el 25 % de las personas que la reciben. En una jornada de vacunación se aplican 120 dosis, a personas que reaccionan independientemente unas de otras.

Sea X la cantidad de personas con reacción leve. Se quiere la probabilidad de que haya como mucho 35. Se puede usar R, Python o Excel: en R, pbinom(k, n, p) da la binomial acumulada, pnorm(x, media, desvío) la normal y rbinom(r, n, p) simula.

a) ¿Qué distribución tiene X? Calculá su esperanza y su desvío, y verificá si se cumplen las condiciones habituales para aproximarla con la normal.

> RESOLUCIÓN
>
> $X \sim B(120;\ 0{,}25) \qquad \mu = n\,p = 30{,}0 \qquad \sigma = \sqrt{n\,p\,(1-p)} = 4{,}743$
>
> Las condiciones de manual piden n·p ≥ 5 y n·(1 − p) ≥ 5 (algunos libros piden 10): acá son 30,0 y 90,0, de sobra. Con tantos ensayos y p no demasiado cerca de 0 ni de 1, la binomial es casi simétrica.

b) Calculá con software la probabilidad exacta P(X ≤ 35).

> RESOLUCIÓN
>
> En R: pbinom(35, 120, 0.25). En Python: scipy.stats.binom.cdf(35, 120, 0.25). En Excel: =DISTR.BINOM.N(35; 120; 0,25; VERDADERO).
>
> $P(X \le 35) = 0{,}8759$
>
> A mano serían 36 términos con combinatorios enormes: por eso, antes de las computadoras, la aproximación normal era la única forma práctica de calcularla.

c) Aproximala con la normal, primero sin corrección por continuidad y después con ella. ¿Cuál se acerca más a la exacta?

> RESOLUCIÓN
>
> Sin corrección: $P\left(Z \le \dfrac{35 - 30{,}0}{4{,}743}\right) = 0{,}8541 \qquad$ Con corrección: $P\left(Z \le \dfrac{35{,}5 - 30{,}0}{4{,}743}\right) = 0{,}8769$

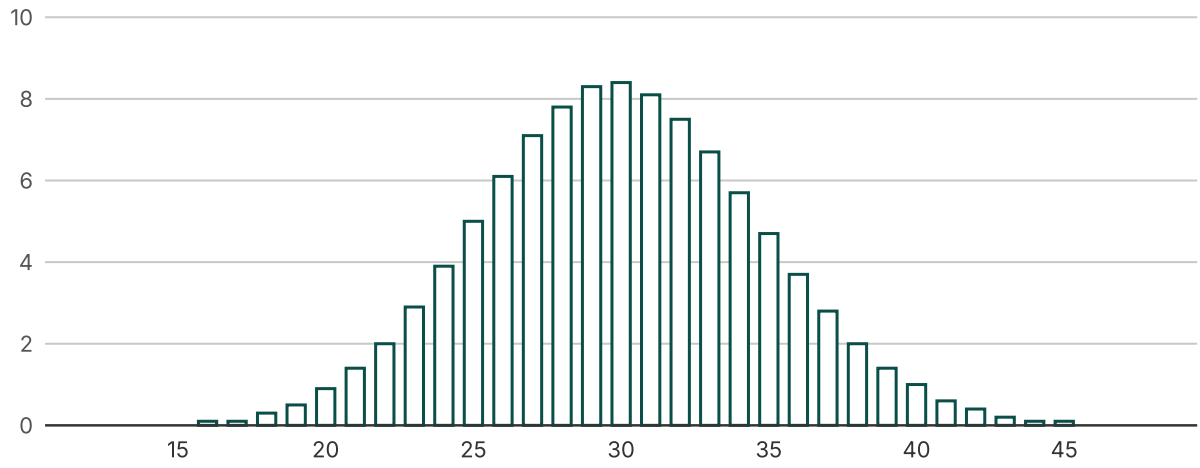
La exacta es 0,8759: sin corrección el error es 0,0218; con corrección, 0,0010. La binomial pone toda la probabilidad del valor 35 en una barra que en la normal va de 34,5 a 35,5; si se corta en 35, se pierde la mitad de esa barra. Por eso la corrección suma 0,5 al límite.

d) Simulá 10.000 jornadas de vacunación y calculá en qué proporción hubo como mucho 35 reacciones. Graficá la distribución de las 10.000 cantidades simuladas.

RESOLUCIÓN

En R: x ← rbinom(10000, 120, 0.25); mean(x <= 35); barplot(table(x)). En Python: x = rng.binomial(120, 0.25, 10000); np.mean(x <= 35).

La proporción cambia en cada corrida, pero debe quedar a una o dos centésimas de 0,876. El gráfico se tiene que parecer a la distribución exacta:



Eje vertical: probabilidad, en %.

Es una campana casi simétrica alrededor de 30,0: la razón por la que la normal aproxima bien. Es el teorema central del límite en acción, porque X es una suma de 120 variables de Bernoulli.

e) Repetí la comparación para otra vacuna, con reacción en el 5 % de las personas y solo 20 dosis: calculá P(X = 0) exacta y aproximada con la normal (con corrección). ¿Qué pasó?

RESOLUCIÓN

$P(X = 0) = 0{,}95^{20} = 0{,}3585 \qquad P(-0{,}5 < Y < 0{,}5) = 0{,}2421 \quad \text{con } Y \text{ normal, } \mu = 1,\ \sigma = 0{,}975$

La aproximación falla por mucho: n·p = 1, lejos de 5. La binomial queda muy asimétrica, pegada al cero, y la campana normal le asigna probabilidad a valores negativos que X no puede tomar. Con p chico y n moderado, la que aproxima bien es la Poisson con λ = n·p; con software, lo mejor es calcular la binomial exacta.

EJERCICIO 6 ¿Qué distribución corresponde?

Para discutir

Para cada variable, decidí si se puede modelar con una binomial, una Poisson, una normal o ninguna de las tres. Justificá con las condiciones de cada modelo y, si corresponde, decí cuáles serían los parámetros.

a) La cantidad de llamadas que recibe el servicio de emergencias en una hora.

> RESOLUCIÓN
>
> Poisson: se cuentan sucesos en un intervalo de tiempo, sin un máximo fijo, que llegan de a uno y con un ritmo promedio estable. El parámetro es el promedio de llamadas por hora, que se estima con los registros. Si el ritmo cambia mucho a lo largo del día, hay que usar un λ distinto para cada franja.

b) De 15 pacientes tratados con un medicamento que cura al 70 %, la cantidad que se cura.

> RESOLUCIÓN
>
> Binomial, B(15; 0,7): ensayos fijos, dos resultados, probabilidad constante y pacientes independientes. Los 15 ensayos son pocos para aproximarla con la normal, porque $n \cdot (1 - p) = 4{,}5$.

c) La estatura de las mujeres adultas de la provincia.

> RESOLUCIÓN
>
> Normal, aproximadamente: es una variable continua que resulta de muchos factores que se suman, simétrica y sin colas largas. Los parámetros —la media y el desvío— se estiman con una muestra. Ningún dato real es exactamente normal: el modelo es una buena descripción, no una ley.

d) El ingreso mensual de los hogares.

> RESOLUCIÓN
>
> Ninguna de las tres. Es continua, pero muy asimétrica a la derecha: la mayoría de los hogares se concentra en ingresos bajos y medios, y unos pocos tienen ingresos muy altos (el histograma de la EPH de la semana 2 lo muestra). Usar la normal daría probabilidad a ingresos negativos y subestimaría la cola de arriba.

e) Se sacan cinco cartas de un mazo español de 40, sin reponer, y se cuenta cuántas son de oro.

> RESOLUCIÓN
>
> No es binomial, aunque lo parezca: como las cartas no se reponen, la probabilidad de sacar oro cambia en cada extracción (10/40 la primera; 9/39 la segunda si salió oro). Los ensayos no son independientes. La distribución correcta es la hipergeométrica. Si la población fuera muy grande comparada con la muestra —elegir 5 personas de una ciudad—, la binomial sería una muy buena aproximación.

f) La cantidad de errores de tipeo por página en un libro de 300 páginas.

> RESOLUCIÓN
>
> Poisson: los errores son sucesos raros que pueden aparecer en cualquier lugar de la página, y se cuentan por unidad de «espacio» en lugar de tiempo. También se puede pensar como una binomial con muchísimos caracteres y una probabilidad de error muy chica, que es justamente el caso en que la binomial se parece a la Poisson.

---

estadistica.ar/docentes/trabajos-practicos/semana-7 · Las lecturas de cada tema y la hoja de fórmulas están en el sitio.