ESTADISTICA.AR

Variables aleatorias y distribuciones

Estadística · Semana 7 · Unidad 4: Variables aleatorias y distribuciones

1 / 30

NOTAS · 1

Pensada para 90 a 120 minutos. Es la clase que une las dos mitades del curso: las distribuciones de hoy son los modelos con los que, desde la semana que viene, se razona sobre las muestras.

PREGUNTA PARA LA CLASE

En Gran Catamarca, el 55,6 % de las personas de 14 años o más está ocupado. Si se eligen 10 al azar, ¿cuántas se espera que estén ocupadas? ¿Qué tan probable es que sean justo esas?

2 / 30

NOTAS · 2

Se esperan 10 · 0,556 = 5,56: es la esperanza de una binomial. Pero el valor más probable, 6, sale apenas el 24,1 % de las veces. Que se espere un valor no quiere decir que sea probable: el resto de la clase es para describir todo lo que puede salir.



NOTAS · 3

La semana 2 describió datos con tablas y gráficos; hoy se hace lo mismo con lo que todavía no pasó: la distribución de una variable aleatoria es la tabla de frecuencias de la población, o de infinitas repeticiones.



SEMANA 7 · VARIABLE ALEATORIA

Una variable aleatoria asigna un número a cada resultado

Discreta

- Toma valores que se pueden contar.
- Personas ocupadas en un hogar: 0, 1, 2...
- Se describe con la **función de probabilidad**, P(X = x).

Continua

- Toma cualquier valor en un intervalo.
- Un tiempo de espera, un peso, una estatura.
- Se describe con la **función de densidad**: la probabilidad es un área.

5 / 30

NOTAS · 5

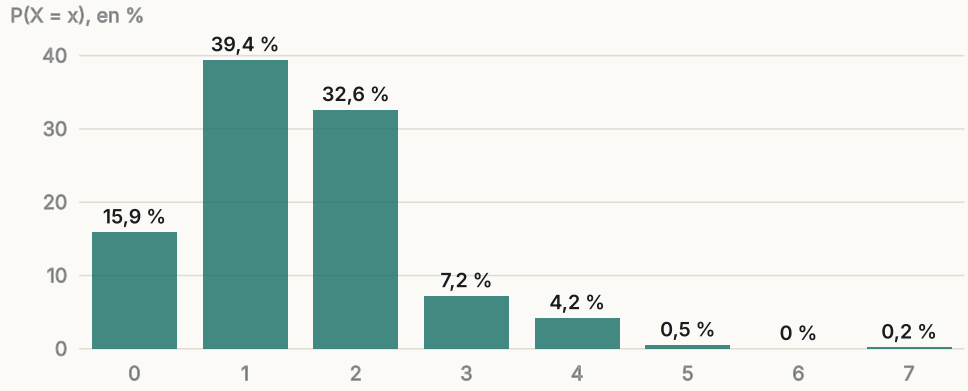
Se escribe con mayúscula, X, y sus valores con minúscula, x. El experimento es elegir un hogar al azar; la variable, cuántas personas ocupadas tiene.

En una continua, la probabilidad de un valor exacto es cero: lo que tiene probabilidad es un intervalo. Por eso P(X ≤ a) y P(X < a) son iguales en una continua, y no en una discreta.

SEMANA 7 · VARIABLE ALEATORIA

La función de probabilidad

X: cantidad de personas ocupadas en un hogar elegido al azar entre los 429 encuestados.



Fuente: INDEC, Encuesta Permanente de Hogares, primer trimestre de 2026. Gran Catamarca.

6 / 30

NOTAS · 6

Cada barra es una probabilidad: P(X = 1) = 169 / 429. Son todas mayores o iguales que cero y suman 1: las dos condiciones de cualquier función de probabilidad.

Ningún hogar de la muestra tiene 6 ocupados, y uno tiene 7: en la población seguramente hay hogares con 6. Esta distribución es la de los hogares encuestados, tomada como si fuera la población.

SEMANA 7 · VARIABLE ALEATORIA

La función de distribución acumulada

$F(x) = P(X \leq x)$: la probabilidad de que la variable valga x o menos.

Como mucho un ocupado

$F(1) = P(X = 0) + P(X = 1) = (68 + 169) / 429 =$ **55,2 %**

Dos o más

$P(X \geq 2) = 1 - F(1) =$ **44,8 %**

Por el complemento, como en la semana 5.

7 / 30

NOTAS · 7

Es la frecuencia relativa acumulada de la semana 2, ahora como probabilidad. Las tablas de la binomial y de la normal dan justamente F.

2

Esperanza y varianza

El centro y la dispersión de una distribución

8 / 30

SEMANA 7 · ESPERANZA Y VARIANZA

La esperanza y la varianza a mano

Ocupados, x	Hogares	x · hogares	x^2 · hogares
0	68	0	0
1	169	169	169
2	140	280	560
3	31	93	279
4	18	72	288
5	2	10	50
7	1	7	49
Suma	429	631	1.395

$E(X) = \Sigma\, x \cdot P(X = x) = 631 / 429$. $E(X^2) = 1.395 / 429$.

9 / 30

NOTAS · 9

Con las probabilidades se suma $x \cdot P(X = x)$; con las cantidades de hogares, se suma $x \cdot$ hogares y se divide por el total al final. Es lo mismo y evita los redondeos.

El valor 6 no aparece porque ningún hogar tiene seis ocupados: su término es cero.

SEMANA 7 · ESPERANZA Y VARIANZA

El centro y la dispersión

$E(X) = 1{,}471$
ocupados por hogar, en promedio

$V(X) = 1{,}088$
$E(X^2) - E(X)^2 = 3{,}252 - 1{,}471^2$

$\sigma = 1{,}043$
el desvío estándar, en ocupados

10 / 30

NOTAS · 10

La esperanza es la media de la población: si se repitiera muchas veces el experimento de elegir un hogar, el promedio de ocupados se acercaría a 1,471. Que no sea un número entero no es un problema: es un promedio, no un hogar.

La fórmula $E(X^2) - E(X)^2$ es la más cómoda a mano. La varianza también se puede calcular como $\Sigma\, (x - \mu)^2 \cdot P(X = x)$, como en la semana 3.

SEMANA 7 · ESPERANZA Y VARIANZA

Propiedades que se usan todo el tiempo

Un cambio de escala

$E(a + bX) = a + b \cdot E(X)$

$V(a + bX) = b^2 \cdot V(X)$: sumar una constante no cambia la dispersión.

Una suma

$E(X + Y) = E(X) + E(Y)$

Si X e Y son independientes, también $V(X + Y) = V(X) + V(Y)$.

11 / 30

NOTAS · 11

Si cada ocupado aporta un ingreso fijo de b pesos, el ingreso del hogar tiene esperanza $b \cdot E(X)$ y desvío $b \cdot \sigma$.

La segunda es la base de la binomial, que viene ahora: una suma de n ensayos independientes.

3

Binomial y Poisson

Contar éxitos y contar sucesos

12 / 30

SEMANA 7 · BINOMIAL Y POISSON

Bernoulli y binomial

Bernoulli

Un solo ensayo con dos resultados: éxito (1) o fracaso (0). Elegir una persona y ver si está ocupada: p = 0,556.

$E(X) = p$ y $V(X) = p(1 - p)$.

Binomial

La cantidad de éxitos en **n ensayos de Bernoulli independientes** con la misma p.

$E(X) = np$ y $V(X) = np(1 - p)$.

13 / 30

NOTAS · 13

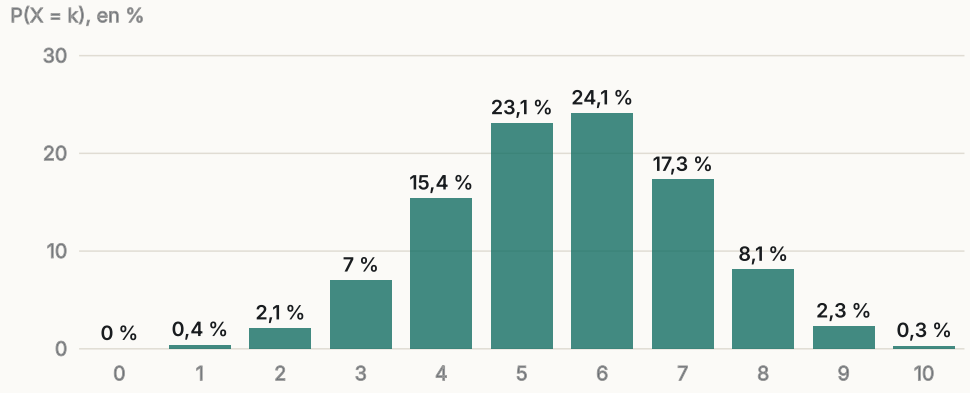
Las condiciones de la binomial: un número fijo de ensayos, dos resultados en cada uno, la misma probabilidad de éxito en todos e independencia entre ellos. Si alguna falla, no es binomial.

Elegir 10 personas de toda la ciudad es casi lo mismo que con reposición: la población es tan grande que sacar una no cambia la proporción de las demás.

SEMANA 7 · BINOMIAL Y POISSON

La binomial con n = 10 y p = 0,556

Cantidad de personas ocupadas entre 10 elegidas al azar.



$P(X = k) = C(10, k) \cdot p^k \cdot (1 - p)^{10-k}$: las maneras de elegir cuáles son los k éxitos, por la probabilidad de cada una.

14 / 30

NOTAS · 14

Por ejemplo, $P(X = 6) = C(10, 6) \cdot 0{,}556^6 \cdot 0{,}444^4 = 210 \cdot \ldots = 0{,}241$. La tabla de la binomial y la calculadora evitan la cuenta.

Con p cerca de 0,5 la distribución es casi simétrica; con p chica se amontona a la izquierda.

SEMANA 7 · BINOMIAL Y POISSON

Qué se puede responder con ella

5,56

ocupados esperados: E(X) = np = 10 · 0,556

1,57

de desvío: √(np(1 − p))

10,6 %

de probabilidad de que haya 8 o más: P(X ≥ 8) = 1 − F(7)

15 / 30

NOTAS · 15

Y la probabilidad de que haya como mucho 4: F(4) = 24,9 %. «Como mucho», «al menos» y «más de» son las expresiones en que más se equivocan los estudiantes: conviene traducir cada una a ≤ o ≥ antes de calcular.

SEMANA 7 · BINOMIAL Y POISSON

La Poisson: las coces de caballo

Muertes por coz de caballo en un cuerpo de caballería durante un año.

Muertes en el año	Cuerpos-año observados	Esperados con Poisson
0	109	108,7
1	65	66,3
2	22	20,2
3	3	4,1
4 o más	1	0,7
Total	200	200

Fuente: Bortkiewicz (1898), Das Gesetz der kleinen Zahlen. Diez cuerpos de caballería durante veinte años.

16 / 30

NOTAS · 16

La media es λ = 0,61 muertes por cuerpo y por año. Con ese único número, la Poisson predice casi exactamente cuántos cuerpos-año tuvieron 0, 1, 2… muertes. Bortkiewicz lo publicó como «la ley de los números pequeños».

Los esperados son 200 · P(X = k). Por ejemplo, 200 · e^(−0,61) = 108,7.

SEMANA 7 · BINOMIAL Y POISSON

La distribución de Poisson

$$P(X = k) = e^{(-\lambda)} \cdot \lambda^k / k!, \text{ para } k = 0, 1, 2, \ldots$$

Cuenta sucesos

En un intervalo de tiempo, de espacio o de volumen: llamadas por hora, errores por página.

Raros e independientes

A una tasa constante, sin que uno cambie la probabilidad de otro.

Un solo parámetro

λ es la esperanza y también la varianza: $E(X) = V(X) = \lambda$.

17 / 30

NOTAS · 17

La Poisson es el límite de una binomial con n muy grande y p muy chica: muchos soldados, cada uno con una probabilidad mínima de morir por una coz. Por eso sirve para sucesos raros.

Que la varianza sea igual a la media es una prueba práctica: si en los datos la varianza es mucho mayor, la Poisson no es el modelo.

4

La distribución normal

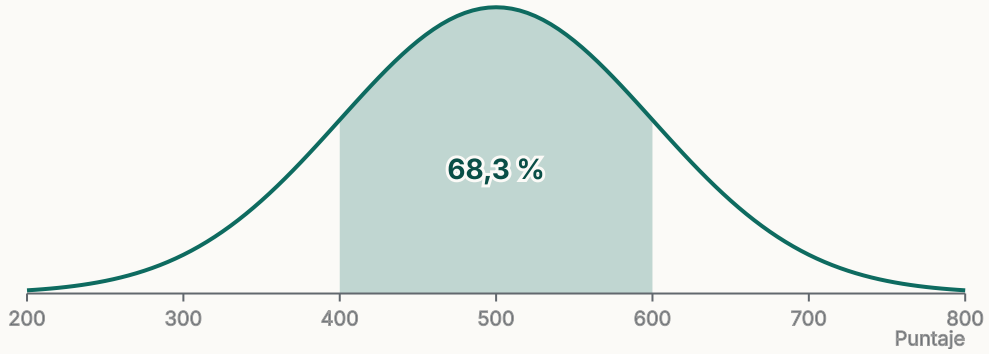
La campana, la tipificación y la tabla

18 / 30

SEMANA 7 · LA DISTRIBUCIÓN NORMAL

La distribución normal

Puntajes con distribución normal de media μ = 500 y desvío σ = 100, como la escala de las pruebas PISA.



La escala de PISA se fijó con media 500 y desvío 100 entre los países de la OCDE.

19 / 30

NOTAS · 19

La campana es simétrica alrededor de μ, que es a la vez media, mediana y moda. σ es la distancia de μ a los puntos donde la curva cambia de curvatura. El área total bajo la curva es 1.

Entre μ − σ y μ + σ, de 400 a 600 puntos, queda el 68,3 % del área: es la probabilidad de que un estudiante elegido al azar tenga un puntaje en ese intervalo.

SEMANA 7 · LA DISTRIBUCIÓN NORMAL

La regla 68 – 95 – 99,7

En cualquier normal, la probabilidad de caer a cierta cantidad de desvíos de la media es siempre la misma.

68,3 %

entre μ − σ y μ + σ: de 400 a 600

95,4 %

entre μ − 2σ y μ + 2σ: de 300 a 700

99,7 %

entre μ − 3σ y μ + 3σ: de 200 a 800

20 / 30

NOTAS · 20

Es lo que permite usar una sola tabla para todas las normales: se mide en desvíos. El 95 % entre ±2σ (más exactamente ±1,96σ) reaparece en los intervalos de confianza.

SEMANA 7 · LA DISTRIBUCIÓN NORMAL

La tipificación

$z = (x - \mu) / \sigma$: cuántos desvíos está x por encima o por debajo de la media.

La normal estándar

Si X es normal, $Z = (X - \mu) / \sigma$ es normal con media 0 y desvío 1.

Su función acumulada se llama Φ y es la de la tabla.

Un ejemplo

Un puntaje de 650: $z = (650 - 500) / 100 =$ **1,50**.

21 / 30

NOTAS · 21

Es el mismo puntaje z de la semana 3, ahora con μ y σ de la población. La tipificación convierte cualquier pregunta sobre una normal en una pregunta sobre Z.

SEMANA 7 · LA DISTRIBUCIÓN NORMAL

Cómo se lee la tabla de la normal

Φ(z) = P(Z ≤ z). La fila da el entero y el primer decimal de z; la columna, el segundo decimal.

z	0,00	0,02	0,04	0,06	0,08
1,2	0,8849	0,8888	0,8925	0,8962	0,8997
1,3	0,9032	0,9066	0,9099	0,9131	0,9162
1,4	0,9192	0,9222	0,9251	0,9279	0,9306
1,5	0,9332	0,9357	0,9382	0,9406	0,9429

Un extracto: la tabla completa va de −3,49 a 3,49.

22 / 30

NOTAS · 22

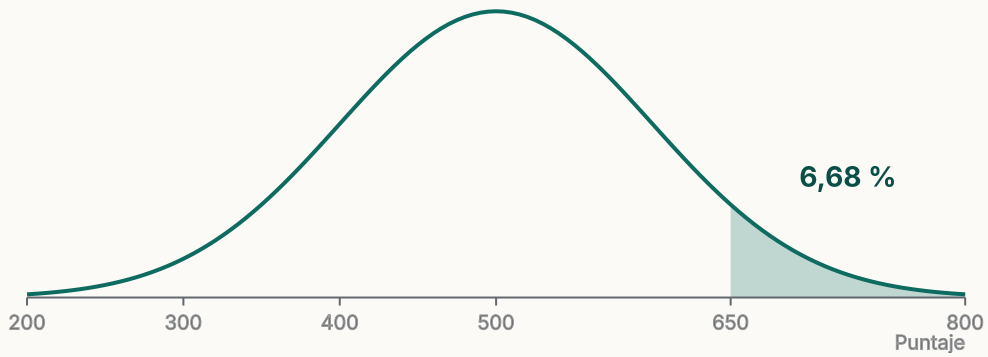
Para z = 1,50: fila 1,5, columna 0,00, Φ = 0,9332. Para z = 1,28: fila 1,2, columna 0,08, Φ = 0,8997.

Para un z negativo se usa la simetría: $\Phi(-z) = 1 - \Phi(z)$. Algunas tablas traen también los negativos.

SEMANA 7 · LA DISTRIBUCIÓN NORMAL

Una probabilidad con la tabla

$P(X > 650) = P(Z > 1{,}50) = 1 - \Phi(1{,}50) = 1 - 0{,}9332 = 0{,}0668$



23 / 30

NOTAS · 23

La tabla da el área a la izquierda; la pregunta pide la de la derecha, y por eso se resta de 1. Dibujar la curva y sombrear lo que se pide antes de buscar en la tabla evita la mitad de los errores.

Entre dos valores se restan dos áreas: $P(400 < X < 650) = \Phi(1{,}50) - \Phi(-1)$.

PREGUNTA PARA LA CLASE

¿Qué puntaje hay que superar para estar en el 10 % más alto?

24 / 30

NOTAS · 24

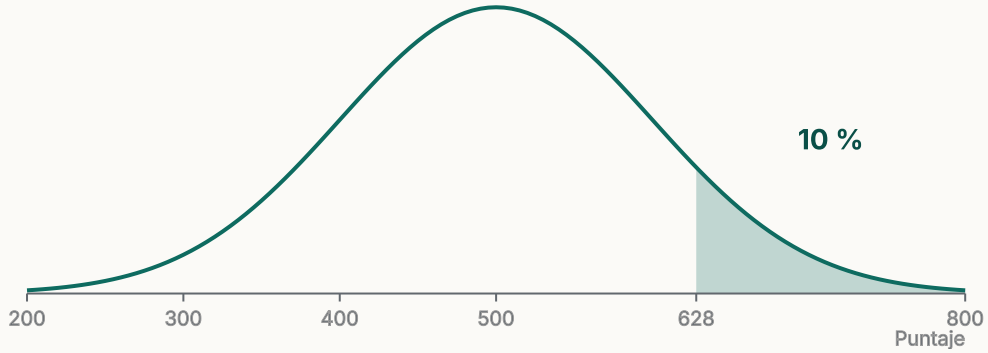
Es la pregunta al revés: se conoce el área y se busca el valor. En el cuerpo de la tabla se busca 0,90; el más cercano es 0,8997, en $z = 1{,}28$. Se destipifica: $x = \mu + z \cdot \sigma = 500 + 1{,}28 \cdot 100 = 628$.

Con la calculadora, que no redondea z, da 628,2. La diferencia es el redondeo de la tabla.

SEMANA 7 · LA DISTRIBUCIÓN NORMAL

El percentil 90

$x = \mu + z \cdot \sigma = 500 + 1{,}28 \cdot 100 = 628$



25 / 30

NOTAS · 25

El percentil 90 deja el 90 % del área a su izquierda. El mismo razonamiento da cualquier percentil, y es el que se usa para buscar los valores críticos de la inferencia.

5

Aproximación normal y elección

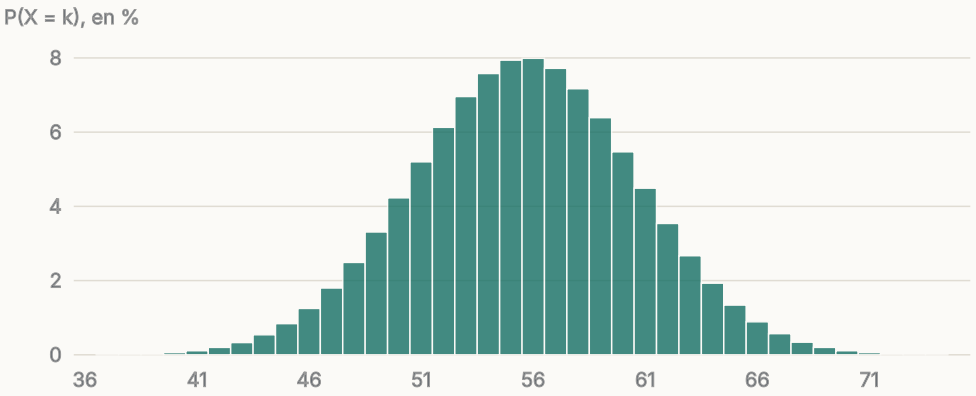
Cuándo usar cada distribución

26 / 30

SEMANA 7 · APROXIMACIÓN NORMAL Y ELECCIÓN

La binomial con n = 100 ya parece una normal

Cantidad de ocupados entre 100 personas elegidas al azar, con p = 0,556. De 36 a 75: afuera la probabilidad es casi cero.



27 / 30

NOTAS · 27

Media np = 55,6 y desvío √(np(1 − p)) = 4,97. La forma de campana aparece porque la binomial es una suma de muchos ensayos independientes: es el teorema central del límite, que viene la semana que viene.

SEMANA 7 · APROXIMACIÓN NORMAL Y ELECCIÓN

La aproximación normal

P(X ≤ 50) con X binomial, n = 100 y p = 0,556: μ = 55,6 y σ = 4,97.

0,1524
exacta, con la binomial

0,1292
con la normal: Φ((50 − 55,6) / 4,97) = Φ(−1,13)

0,1515
con corrección por continuidad: Φ(−1,03), usando 50,5

La corrección por continuidad: cada valor entero k de la binomial ocupa en la normal el intervalo de k − 0,5 a k + 0,5.

28 / 30

NOTAS · 28

Las condiciones habituales para aproximar: np ≥ 5 y n(1 − p) ≥ 5. Acá 55,6 y 44,4: se cumplen con margen.

Sin la corrección, el error es de 0,0232; con ella, de 0,0009. Hoy la calculadora da la exacta, pero la aproximación explica por qué la normal aparece en toda la inferencia sobre proporciones.

SEMANA 7 · APROXIMACIÓN NORMAL Y ELECCIÓN

Cuál usar

Distribución	Describe	Parámetros	Esperanza	Varianza
Bernoulli	Un ensayo con dos resultados	p	p	p(1 − p)
Binomial	Éxitos en n ensayos independientes	n, p	np	np(1 − p)
Poisson	Sucesos en un intervalo	λ	λ	λ
Normal	Una variable continua en campana	μ, σ	μ	σ^2

Antes de calcular: ¿la variable cuenta o mide? Si cuenta, ¿hay un número fijo de ensayos o un intervalo?

29 / 30

NOTAS · 29

Que una variable sea continua no la hace normal: el ingreso, por ejemplo, es muy asimétrico, como se vio en la semana 3. El práctico tiene un ejercicio para discutir cuál corresponde en cada caso.

SEMANA 7 · VARIABLES ALEATORIAS Y DISTRIBUCIONES

Para llevarse

- Una variable aleatoria se describe con su función de probabilidad (discreta) o de densidad (continua).
- La esperanza es la media de la población; la varianza, su dispersión.
- Binomial: éxitos en n ensayos independientes. Poisson: sucesos raros en un intervalo.
- En la normal, la probabilidad es un área: se tipifica y se lee Φ en la tabla.
- Con n grande, la binomial se aproxima con la normal, mejor con corrección por continuidad.

Guía de trabajo práctico 7
estadistica.ar/docentes/trabajos-practicos/semana-7

Tabla de la normal
estadistica.ar/herramientas/tabla-normal

Simulador de distribuciones
estadistica.ar/herramientas/sim-distribuciones

Ejercicios: normal
estadistica.ar/ejercicios/normal

30 / 30

NOTAS · 30

Para el práctico: la tabla de la normal impresa y calculadora. El simulador de distribuciones sirve para ver cómo cambia cada una con sus parámetros.