TEMA 6. Distribuciones

Alicia Nieto Reyes

BIOESTADÍSTICA

Probabilidad= Distribución= Distribución de Probabilidad

Cuando queremos conocer la probabilidad asociada a una variable (aleatoria) se habla de la distribución de la variable

Las variables las denotamos por letras como: X, Y, Z ...

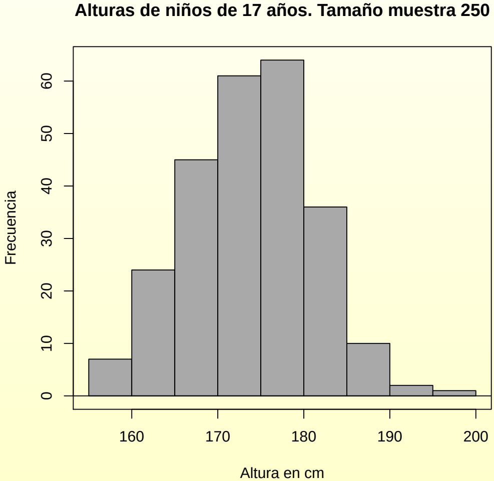
Distribución de variables

- continuas
  - Normal: $N(\mu, \sigma)$
  - t de Student: t_n
  - ji-cuadrado: χ^2_n
  - ...
- discretas
  - Binomial: $B(n, p)$
  - Poisson: $Poi(\lambda)$
  - ...

$X \sim N(\mu, \sigma)$ significa que la variable X sigue una distribución Normal de media μ y varianza σ^2

Normal o Gaussiana

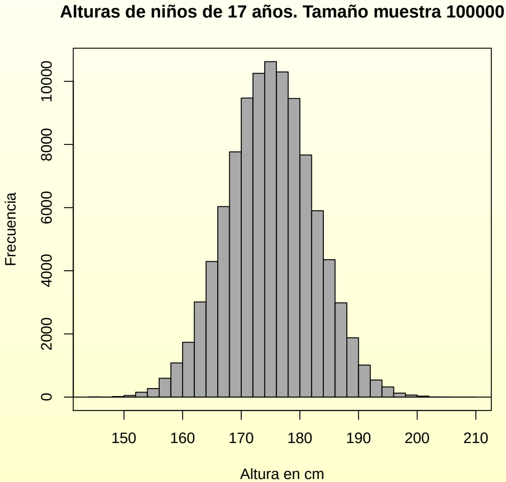
Ejemplo 1: Queremos estudiar la variable altura, X, de la población: niños de 17 años
Histograma de una muestra de tamaño 250



Normal o Gaussiana

Población: niños de 17 años Variable: Altura, X

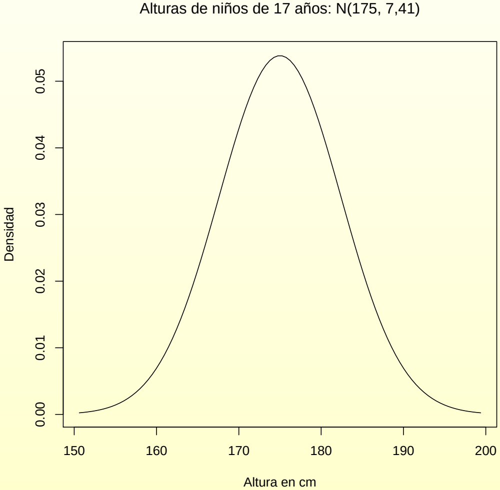
Histograma de una muestra de tamaño 100.000 (la muestra anterior de tamaño 250 está incluida)



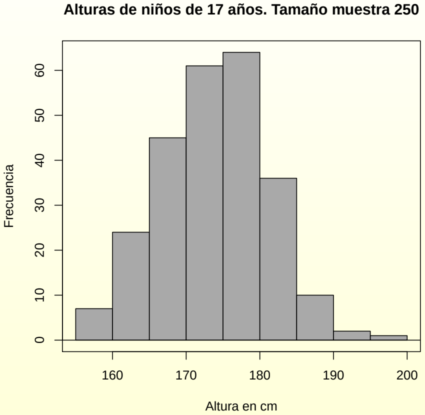
Normal o Gaussiana

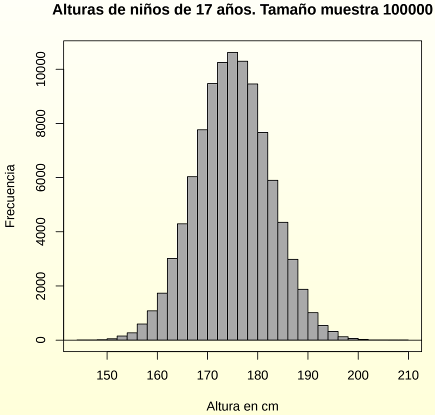
Si el tamaño de la muestra se acerca al de la población, el contorno del histograma pasaría a ser esta curva

Pero, hemos ajustado el eje de ordenadas para que el area bajo la curva sea 1



Normal o Gaussiana

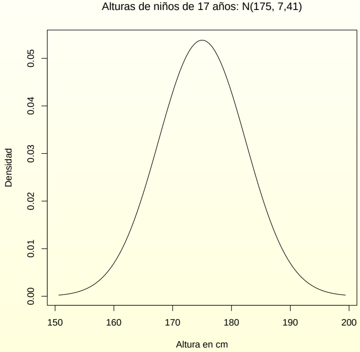




$$P_n(X <= 160) = \frac{7}{250} = 0,028 \text{ con } n = 250$$

$$P_n(X <= 160) = \frac{2.177}{100.000} = 0,02177 \text{ con } n = 100.000$$

Normal o Gaussiana

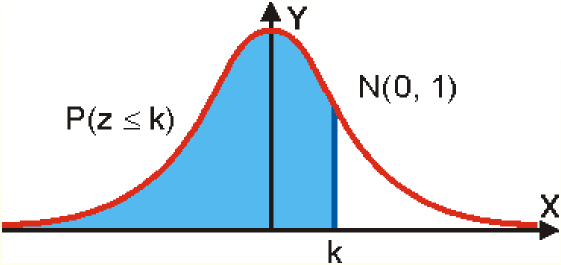


$$P(X \leq 160) =$$
$$P\left(\frac{X-175}{7.41} \leq \frac{160-175}{7.41}\right) =$$
$$P(Y \leq -2.02) =$$
$$1 - P(Y \leq 2.02) =$$
$$1 - 0,9783 =$$
$$0,0217$$

Resultado

Sea X una variable aleatoria que sigue una distribución $N(\mu, \sigma)$. Entonces la variable aleatoria $Y = \dfrac{X-\mu}{\sigma}$ sigue una distribución $N(0,1)$.

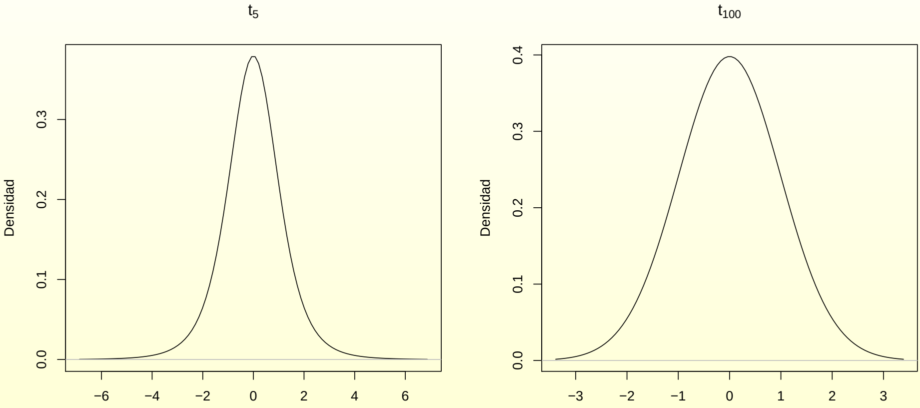
ÁREAS BAJO LA DISTRIBUCIÓN DE PROBABILIDAD NORMAL ESTÁNDAR, N(0, 1)



z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
…	…	…	…	…	…	…	…	…	…	…
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

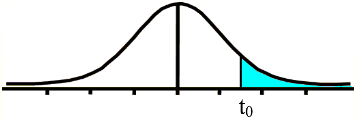
t de Student

t_n: t de Student con n grados de libertad



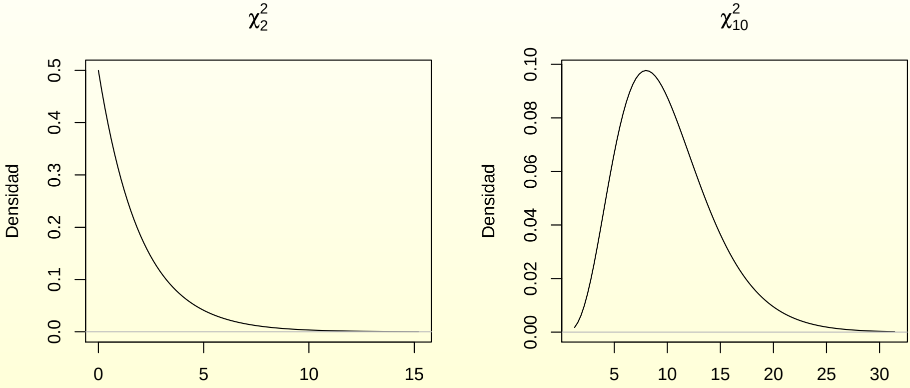
Se calcula con una tabla del tipo de la Normal ó, al igual que esta, con un programa informático adecuado

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609

χ_n^2: ji-cuadrado con n grados de libertad



- $\mu = n \qquad\qquad \sigma = 2n$
- El tipo de gráfica con $n = 1, 2$ es diferente a con $n \geq 3$

Tabla de distribución ji-cuadrado

Grados de
Libertad Probabilidad acumulada

	0.0005	0.001	0.0025	0.005	0.010	0.025	0.050	0.100	0.150	0.200	0.250
1	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.016	0.036	0.064	0.102
2	0.001	0.002	0.005	0.010	0.020	0.051	0.103	0.211	0.325	0.446	0.575
3	0.015	0.024	0.045	0.072	0.115	0.216	0.352	0.584	0.798	1.005	1.213
4	0.064	0.091	0.145	0.207	0.297	0.484	0.711	1.064	1.366	1.649	1.923
5	0.158	0.210	0.307	0.412	0.554	0.831	1.145	1.610	1.994	2.343	2.675
6	0.299	0.381	0.527	0.676	0.872	1.237	1.635	2.204	2.661	3.070	3.455
7	0.485	0.598	0.794	0.989	1.239	1.690	2.167	2.833	3.358	3.822	4.255
8	0.710	0.857	1.104	1.344	1.646	2.180	2.733	3.490	4.078	4.594	5.071
9	0.972	1.152	1.450	1.735	2.088	2.700	3.325	4.168	4.817	5.380	5.899
10	1.265	1.479	1.827	2.156	2.558	3.247	3.940	4.865	5.570	6.179	6.737
11	1.587	1.834	2.232	2.603	3.053	3.816	4.575	5.578	6.336	6.989	7.584
12	1.934	2.214	2.661	3.074	3.571	4.404	5.226	6.304	7.114	7.807	8.438
13	2.305	2.617	3.112	3.565	4.107	5.009	5.892	7.042	7.901	8.634	9.299
14	2.697	3.041	3.582	4.075	4.660	5.629	6.571	7.790	8.696	9.467	10.17
15	3.108	3.483	4.070	4.601	5.229	6.262	7.261	8.547	9.499	10.31	11.04
16	3.536	3.942	4.573	5.142	5.812	6.908	7.962	9.312	10.31	11.15	11.91

Binomial: $B(n,p)$
Se lee Binomial de parámetros n y p

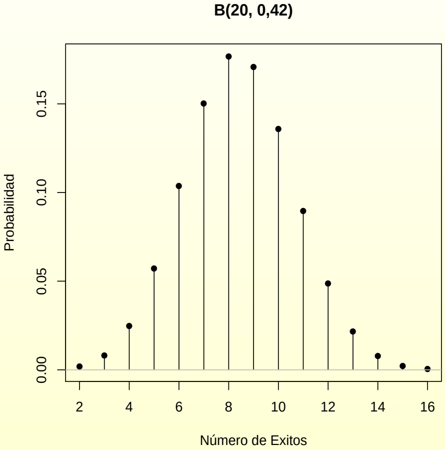
Si $X \sim B(n,p)$ entonces $P(X=k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}$

- Mide el número k de exitos en una secuencia de n ensayos con una porbabilidad p de exito
- Ejemplo: La probabilidad de que 5, k, cantabros/as fumen de entre 20, n, elegidos al azar, sabiendo que el 42% de la población Cantabra es fumadora, es

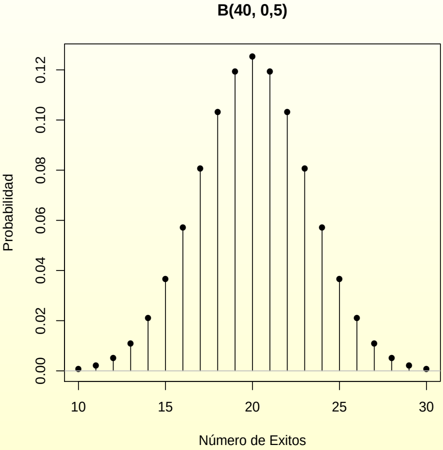
$$\begin{aligned} P(X=5) = & \left(\begin{array}{c} 20 \\ 5 \end{array} \right) 0,42^5 (1-0,42)^{20-5} \\ = & \frac{20!}{5!(20-5)!} 0,42^5 (0,58)^{15} = 0,057 \end{aligned}$$

donde $X \sim B(20, 0, 42)$

Binomial



B(20, 0,42)
Probabilidad
0.00
0.05
0.10
0.15
2
4
6
8
10
12
14
16
Número de Exitos



B(40, 0,5)
Probabilidad
0.00
0.02
0.04
0.06
0.08
0.10
0.12
10
15
20
25
30
Número de Exitos

Poisson: $Poi(\lambda)$

Se lee Poisson de parámetro λ

Si $X \sim Poi(\lambda)$ entonces $P(X = k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$

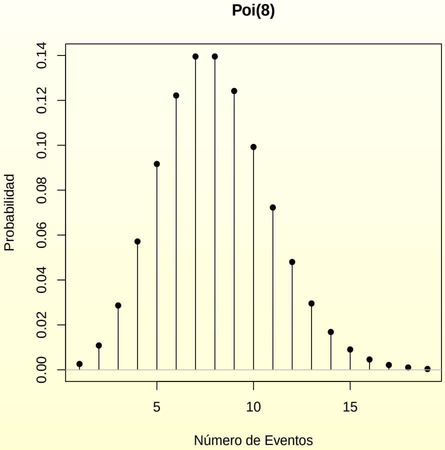
- Expresa la probabilidad de un número k de eventos ocurriendo en un tiempo fijo si estos eventos ocurren con una frecuencia media conocida, λ, y son independientes del tiempo discurrido desde el último evento
- Ejemplo: La probabilidad de que 5, k, jeringuillas de una tirada de 400 esten defectuosas sabiendo que el 2% lo estan es

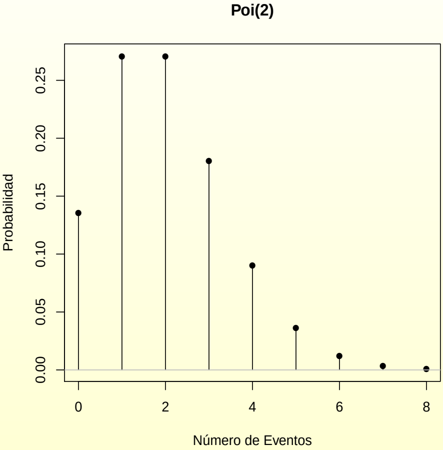
$$P(X = 5) = \frac{e^{-8}8^5}{5!} = 0,0916$$

donde $X \sim Poi(8)$

Nota que λ es el valor esperado de jeringuillas defectuosas: 2% de 400

Poisson


Poi(8)
Probabilidad
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
5
10
15
Número de Eventos


Poi(2)
Probabilidad
0.00
0.05
0.10
0.15
0.20
0.25
0
2
4
6
8
Número de Eventos