

U.D. 8: Distribuciones de probabilidad. Inferencia estadística

Índice

1. Distribuciones de probabilidad	2
1.1. Distribuciones de probabilidad discreta	2
1.1.1. Experimento de Bernoulli	2
1.1.2. Distribución binomial	2
1.2. Distribuciones de probabilidad continua	3
1.2.1. Distribución normal	3
1.3. Aproximación de la distribución binomial a una distribución normal	6
2. Muestreo	6
2.1. Tipos de muestreo aleatorio	7
3. Inferencia estadística	7
3.1. Distribuciones muestrales	7
3.1.1. Distribución de la media	7
3.1.2. Distribución de la proporción	8
3.1.3. Distribución de la diferencia de medias	8
3.2. Intervalos de confianza	8
3.2.1. Intervalos de confianza para la media	9
3.2.2. Intervalos de confianza para la proporción	9
3.2.3. Intervalos de confianza para la diferencia de medias	10

1. Distribuciones de probabilidad

1.1. Distribuciones de probabilidad discreta

Una **distribución de probabilidad de variable discreta** es el resultado de asignar a cada valor de la variable su probabilidad.

x_i	$P[x_i] = p_i$
x_1	p_1
x_2	p_2
x_3	p_3
\vdots	\vdots

A la vista de esta distribución, nos podemos plantear la media (o esperanza), varianza y desviación típica (como en el caso de variables estadísticas). Ahora, usaremos la siguiente notación: μ , esperanza; σ^2 , varianza; σ , desviación típica. El modo de calcular estos parámetros es igual que en estadística, pero teniendo en cuenta que las sumas de todas las probabilidades es 1, i.e., $N = 1$.

1.1.1. Experimento de Bernouilli

Si en una experiencia aleatoria destacamos un suceso A y prestamos atención, exclusivamente, a si ocurre A o su contrario, se trata de una experiencia de Bernouilli. Al suceso A se le denomina **éxito**. Si p es la probabilidad de éxito, entonces la de su contrario es $q = 1 - p$.

1.1.2. Distribución binomial

Se repite n veces una misma experiencia de Bernouilli y nos preguntamos por el número, x , de éxitos. La variable x es discreta y puede tomar los valores $0, 1, 2, 3, \dots, n$. La distribución de probabilidad de la variable x se llama **distribución binomial**, $B(n, p)$, siendo n el número de veces que se repite la experiencia y p la probabilidad de éxito en cada una de las experiencias. En una distribución binomial $B(n, p)$ la probabilidad de k éxitos viene dada por:

$$p[x = k] = \binom{n}{k} p^k \cdot q^{n-k}$$

Y los parámetros de esta distribución son:

$$\mu = np$$
$$\sigma = \sqrt{npq}$$

1.2. Distribuciones de probabilidad continua

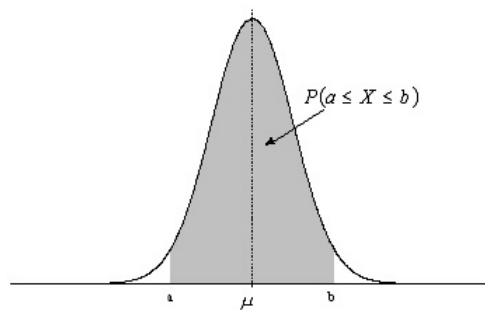
Las variables aleatorias continuas toman un número infinito de valores dentro de un intervalo de la recta real.

Las probabilidades, en este caso, vienen dadas por el área bajo una curva. A la función $f(x)$ que determina esta curva se le llama **función de densidad**. Con lo cual:

- f no puede ser negativa.
- El área bajo toda la curva es 1.
- Para hallar la probabilidad $P[a \leq x \leq b]$, obtendremos el área que hay bajo la curva en el intervalo $[a, b]$
- Las probabilidades de los sucesos puntuales son cero: $P[x = a] = 0$. Por tanto, $P[a \leq x \leq b] = P[a < x < b]$

1.2.1. Distribución normal

La campana de Gauss, o **curva normal**, es una función de densidad continua, simétrica, cuyo máximo coincide con la esperanza μ

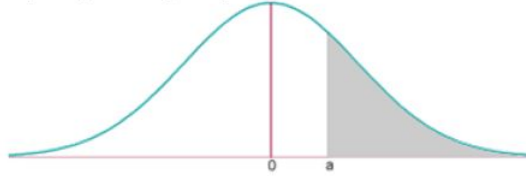


Cuando tengamos una función de densidad de este tipo, hablaremos de distribución normal: $N(\mu, \sigma)$.

Para calcular probabilidades en este tipo de distribuciones usamos la tabla de la $N(0, 1)$, como se indica en los siguientes ejemplos:

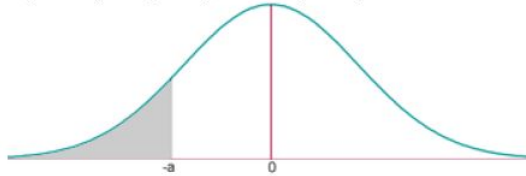
z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99995	0.99995	0.99995	0.99995	0.99995	0.99995	0.99995	0.99995
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

$$P(Z > a) = 1 - P(Z \leq a)$$



Ej: $P(Z > 1.47) = 1 - P(Z \leq 1.47) = 1 - 0.9292 = 0.0708$

$$P(Z \leq -a) = P(Z > a) = 1 - P(Z \leq a)$$



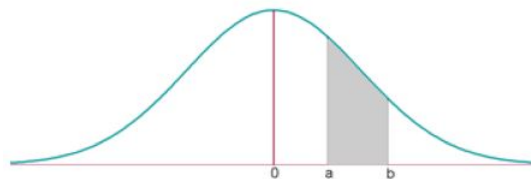
Ej: $P(Z \leq -1.47) = 1 - P(Z \leq 1.47) = 1 - 0.9292 = 0.0708$

$$P(Z > -a) = P(Z \leq a)$$



Ej: $P(Z > -1.47) = P(Z \leq 1.47) = 0.9292$

$$P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$$



Ej: $P(0.45 \leq Z \leq 1.47) = P(Z \leq 1.47) - P(Z \leq 0.45) = 0.9292 - 0.6736 = 0.2556$

Si x es $N(\mu, \sigma)$, para calcular la probabilidad $P[a < x < b]$ se procede del siguiente modo:

$$P[a < x < b] = P\left[\frac{a - \mu}{\sigma} < x < \frac{b - \mu}{\sigma}\right]$$

Este cambio se llama **tipificación de la variable**. Y la variable tipificada z sigue una distribución $N(0, 1)$.

Ejemplos: 

Ejercicio 1 Sea X una distribución normal de media 7 y desviación típica 2. Calcula las siguientes probabilidades:

- a) $P(x < 7, 2)$
- b) $P(6, 5 < x < 7, 5)$
- c) $P(x > 7, 1)$
- d) $P(x < 6, 5)$

1.3. Aproximación de la distribución binomial a una distribución normal

En general, una binomial $B(n, p)$ se puede aproximar a una distribución normal si np y nq son ambos mayores que 5.

En este caso, los parametros de la distribución normal son: $\mu = np$ y $\sigma = \sqrt{npq}$.

2. Muestreo

La estadística es la ciencia que se ocupa de recoger y ordenar datos referidos a diversas características, para su posterior análisis e interpretación.

- **Población:** es el conjunto de elementos sobre el que se realiza el estudio.
- **Muestra:** es la parte de la población que estudiamos.

El **muestreo** es el conjunto de técnicas aplicadas para extraer muestras representativas de una población. Puede ser:

- **Aleatorio:** cuando todos los elementos tienen la misma probabilidad de ser seleccionados como elementos de la muestra.
- **No aleatorio:** cuando todos los elementos que van a ser incluidos en la muestra no tienen la misma probabilidad de ser seleccionados.

2.1. Tipos de muestreo aleatorio

Se dice que un muestreo aleatorio es **con reposición** cuando, tras elegir cualquier elemento, este puede volver a ser elegido.

Si cada elemento de la población puede ser escogido una sola vez para la muestra, el muestreo es **sin reposición**.

Un muestreo se llama:


a) **Aleatorio simple**, cuando cada individuo de la población tiene la misma probabilidad de ser elegido.

b) **Sistemático**, cuando se elige al azar a un individuo de la población y, a partir de este, se eligen los demás a intervalos constantes.

c) **Estratificado**, cuando se divide a la población en clases o estratos y se escogen aleatoriamente un número de individuos de cada estrato mediante muestreo aleatorio simple o sistemático. Puede ser:

- Con afijación igual si se toma el mismo número de individuos de cada estrato sin tener en cuenta su tamaño.
- Con afijación proporcional, si se toma el número de individuos de cada estrato proporcionalmente al tamaño del mismo.

d) **Por conglomerados**, si se divide la población en conjuntos o conglomerados, se eligen al azar algunos de estos conglomerados y se escoge una muestra únicamente de estos conglomerados por muestreo aleatorio simple.

Ejemplos: 

Ejercicio 2 *Se dispone de cuatro tornillos de 1, 2, 3 y 4 gramos de peso respectivamente.*

- a) *Mediante muestreo aleatorio simple, exprese todas las muestras posibles de tamaño 2.*
- b) *Determine la media y la varianza de los pesos medios muestrales*

Ejercicio 3 *En un centro docente hay 160 alumnos matriculados en 1º de ESO, 120 en 2º, 120 en 3º, 80 en 4º, 240 en 1º de Bachillerato y 200 en 2º. Se quiere constituir una comisión en la que todos los cursos estén representados de forma proporcional.*

- a) ¿Cuántos alumnos debe haber en la comisión y cuántos de cada curso si dicha comisión está formada por el 5 % del total del alumnado?
- b) ¿Cuál sería la composición de la comisión si queremos que haya 9 alumnos de 2º de ESO?

3. Inferencia estadística

3.1. Distribuciones muestrales


3.1.1. Distribución de la media

Partimos de una población de tamaño N . Obtenemos k muestras (todas las posibles de tamaño n) y de cada una de ellas se calcula una medida obteniendo k valores. La variable aleatoria que a cada muestra le hace corresponder su media se llama **media muestral**, y se escribe \bar{X} .

Si se cumplen alguna de estas condiciones:

- Las muestras son aleatorias y la población de donde se extraen las muestras sigue una distribución normal $X \rightarrow N(\mu, \sigma)$.
- Las muestras son aleatorias, el tamaño de cada muestra es suficientemente grande ($n > 30$), y la población de donde se extraen las muestras tiene una media μ y desviación típica σ .

Entonces, la distribución de la media muestral, \bar{X} , de las muestras aleatorias de tamaño n que se pueden extraer de una población de media μ y desviación típica σ , sigue una distribución normal $\bar{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$.

Ejemplos: 

Ejercicio 4 La altura de los estudiantes de 2º de Bachillerato de un centro sigue un ley Normal de media 165 cm y desviación típica 10 cm.

- a) ¿Qué distribución sigue la altura media de las muestras de tamaño 25?
- b) Se elige al azar una muestra de 25 estudiantes y se le mide la altura. ¿Cuál es la probabilidad de que la altura media de esa muestra supere los 160 cm?

3.1.2. Distribución de la proporción

Cuando se estudia si los individuos de una población tienen o no cierta característica, habrá una proporción p de individuos que sí la tienen y el resto de individuos, una proporción $q = 1 - p$, que no la tienen. Si se consideran todas las muestras aleatorias de tamaño n que se pueden extraer de dicha población, la variable aleatoria que a cada muestra le hace corresponder la proporción de individuos con esa característica se llama **proporción muestral**, y se denota por \hat{P} .

Si la muestra es aleatoria y su tamaño es suficientemente grande, $n > 30$ o $np > 5$ y $nq > 5$, entonces $\hat{P} \rightarrow N(p, \sqrt{\frac{pq}{n}})$.

3.1.3. Distribución de la diferencia de medias

La distribución de la diferencia de medias muestrales, $\bar{X}_1 - \bar{X}_2$, de las muestras aleatorias de tamaños n_1 y n_2 que se pueden extraer de dos poblaciones de medias μ_1 y μ_2 y desviaciones típicas σ_1 y σ_2 es $\bar{X}_1 - \bar{X}_2 \rightarrow N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$, siempre que las distribuciones sean normales o las muestras suficientemente grandes.

3.2. Intervalos de confianza

Un **intervalo de confianza** con un **nivel de confianza** $1 - \alpha$ es un intervalo que contiene al parámetro desconocido con una probabilidad $1 - \alpha$.


El **error máximo admisible**, E , que se puede cometer al tomar el punto medio del intervalo de confianza como estimación del parámetro desconocido es el radio del intervalo.

3.2.1. Intervalos de confianza para la media

Un **intervalo de confianza para la media** poblacional de una distribución normal con desviación típica σ , con un nivel de confianza $1 - \alpha$ construido a partir de una muestra de tamaño n , es:

$$\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Por tanto, el error máximo admisible es: $E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$. A partir de este error se puede determinar el **tamaño mínimo de la muestra**.

Ejemplos: 

Ejercicio 5 Los directivos de una empresa desean estimar el tiempo medio que tardan los empleados en llegar al puesto de trabajo desde sus domicilios. Admitimos que dicho tiempo sigue una distribución Normal de desviación típica 8 minutos. Se elige al azar una muestra de 9 empleados de esa empresa, obteniéndose los siguientes resultados, expresados en minutos:

10 17 8 27 6 9 32 5 21

- Determine un intervalo de confianza al 92 %, para la media poblacional.
- Con una confianza del 95'5 %, ¿Qué tamaño muestral mínimo sería necesario para estimar el tiempo medio con un error inferior a 1'5 minutos?

3.2.2. Intervalos de confianza para la proporción

Un **intervalo de confianza para la proporción** de individuos que cumplen una característica en una población, con un nivel de confianza $1 - \alpha$, construido a partir de una muestra de tamaño n , es:

$$\left(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}, \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

Por tanto, el error máximo admisible es: $E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$. A partir de este error se puede determinar el **tamaño mínimo de la muestra**.

Ejemplos: 

Ejercicio 6 Se desea estimar la proporción de individuos que piensan votar a un cierto partido político en una determinada ciudad. Para ello se toma una muestra aleatoria de 300 individuos de la ciudad, resultando que 135 de ellos piensan votar a ese partido.

- Calcule un intervalo de confianza al 97 % para la proporción de individuos que piensan votar a ese partido en dicha ciudad.
- Suponiendo que se mantiene la misma proporción muestral y el mismo nivel de confianza del apartado anterior, determine el tamaño mínimo de la muestra para estimar la proporción con un error inferior al 2 %.

3.2.3. Intervalos de confianza para la diferencia de medias

Un **intervalo de confianza para la diferencia de medias** poblacionales de dos distribuciones normales con desviaciones típicas σ_1 y σ_2 con un nivel de confianza $1 - \alpha$, construido a partir de dos muestras de tamaño n_1 y n_2 , es:

$$\left((\bar{X}_1 - \bar{X}_2) - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{X}_1 - \bar{X}_2) + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

Por tanto, el error máximo admisible es: $E = z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.