Tema 5

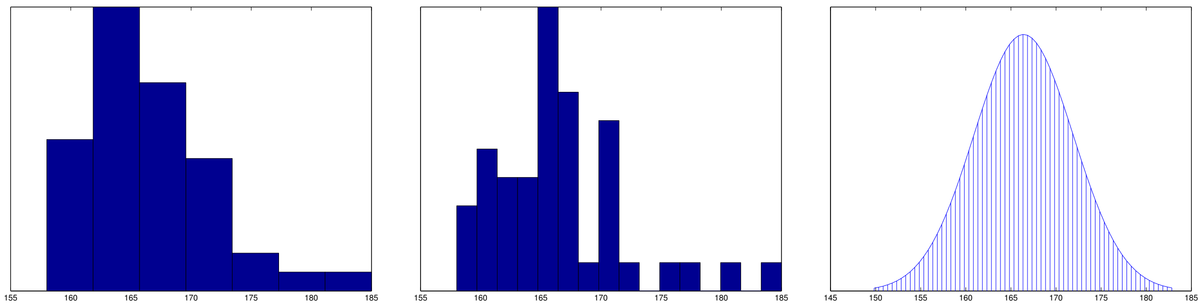
Variables aleatorias continuas

- 5.1 Introducción
- 5.2 Distribuciones continuas
    - 5.2.1 Uniforme
    - 5.2.2 Exponencial
    - 5.2.3 Weibull
    - 5.2.4 Normal
- 5.3 Muestras aleatorias. Otros tipos de muestreo
- 5.4 Teorema central del límite
- 5.5 Distribuciones en el muestreo

5.1. Introducción

En este tema y en la práctica 4 con el Statgraphics, se presentarán diversos modelos teóricos de distribuciones de probabilidad para variables continuas (recuerda en el tema 1 el apartado 1.1.). En el tema 1 (el apartado 1.5.), vimos como la distribución de la población de una variable aleatoria **continua** X podría describirse mediante una curva de densidad (como un histograma idealizado), que representaba frecuencias relativas como áreas bajo la curva. Si en un histograma hacemos tender la amplitud del intervalo de clase a cero tendremos un número infinito de intervalos, convirtiéndose el histograma en un número infinito de barras de grosor infinitesimal, dispuestas de modo continuo (histograma idealizado). De esta forma, llegaríamos a la que llamamos en el tema 1 **curva (o función) de densidad**, y que denotaremos como $f(x)$. Recordemos esto rápidamente usando el siguiente ejemplo:

Ejemplo 5.1.: Los siguientes histogramas representan los datos recogidos en clase de la variable aleatoria X = "altura en centímetros de las chicas (*¡qué son muy grandes!*) que estudian estadística en Diseño en la UJI este curso":

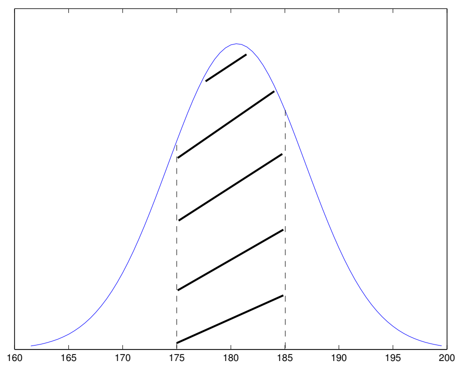


(a) (b) (c)

Figura 5.1: Histogramas de frecuencias relativas del ejemplo 5.1., variando la amplitud de los intervalos de clase, hasta llegar a un histograma idealizado, la curva de densidad (c)

Cada uno de los modelos que veremos (y los que no veremos) tiene asociado su función de densidad y a través de ella podremos calcular probabilidades de distintos sucesos. La forma de calcular probabilidades para variables continuas difiere de la que usamos en el tema anterior para variables discretas. Ahora para calcular la probabilidad de un suceso deberíamos calcular el área comprendida entre el eje x y la función de densidad (o sea, integrar), para los valores señalados por el suceso.

Ejemplo 5.2.: Si quisiéramos conocer la probabilidad de que un chico de la clase (*¡qué también son muy grandes!*) midiera entre 175 y 185 cm, P($175 \leq X \leq 185$), debemos calcular el área rallada, es decir, integrar la función de densidad entre 175 y 185 cm.



Según las reglas de probabilidad, tendremos que el **área total bajo la función de densidad es siempre 1**. Además, puesto que la integral de un punto al mismo punto vale cero (el área de una barra con grosor un punto es nula, mira el gráfico 5.1 (c) y recuerda la última observación del punto 1.5.), se tiene que para **variables continuas, la probabilidad de que una variable aleatoria continua tome un valor puntual es cero**. Así, en el **ejemplo 5.1.**, P($X = 168.96$) = 0, por ejemplo. Por esta razón, para cualquier variable continua X se cumple: P(a ≤ X ≤ b) = P(a < X ≤ b) = P(a < X < b) = P(a ≤ X < b), o sea, PARA VARIABLES CONTINUAS ÚNICAMENTE, la probabilidad será la misma tanto si la desigualdad es o no estricta.

[Fíjate que esta última propiedad no se cumple para las variables discretas (que se estudiaron en el tema anterior).]

Existe gran cantidad de modelos para variables continuas. En clase veremos este año la Normal. En prácticas se introducirán otros modelos: uniforme, exponencial, Weibull. Estas dos últimas distribuciones tienen aplicación en fiabilidad. Además en el tema siguiente usaremos otras como la t de Student, χ^2 Chi-cuadrado y F de Snedecor. Cada una de ellas tiene una curva de densidad y viene caracterizada por uno/s parámetros.

Como ya hemos dicho, para conocer la probabilidad de sucesos para variables continuas deberíamos integrar, sin embargo, para algunos modelos es posible expresar de forma analítica el valor de la integral mediante **la función de distribución acumulada** que denotaremos $F(x)$ y que nos proporcionará $P(X \leq x)$, es decir, para cada x, la función F nos devolverá la probabilidad de que la variable X tome un valor menor o igual que x. A veces, no existe tal expresión explícita y es preciso recurrir a tablas.

[A modo de resumen aclaratorio: cada modelo continuo viene determinado por su función

de densidad, f. Hay que tener claro que la función de densidad, f, NO da probabilidades, sino el área bajo dicha función. Para calcular probabilidades hay que usar F, la función de distribución acumulada.]

.

Observación: *el fin de esta aclaración es tratar de dar una visión general y localizar en que punto del temario nos encontramos, para no perder de vista el objetivo final. En el* **ejemplo 1.3** *(el del ratón ergonómico para niños), nos interesaba estudiar TODA la población de niños. Como eso es inviable, extraeremos una muestra (representativa) de la población, por ejemplo, N = 100 niños. A partir de esa muestra estudiaremos la variable X = "longitud del dedo índice" en la que estábamos interesados. Esta variable es cuantitativa y continua. (Podía habernos intersado más variables continuas como Y = "longitud entre dos puntos determinados de la mano", u otro tipo de variables, como Z = "satisfacción con un determinado juguete").*

Los datos (100 en este caso) que habríamos obtenido, primeramente los podríamos describir haciendo uso de las técnicas vistas en el tema 1: tablas de frecuencias (apartado 1.2), gráficas (histogramas, diagramas de cajas, etc.) (apartado 1.3.) y medidas descriptivas (apartado 1.4): media ($\overline{x}$), mediana, varianza (s^2), desviación típica (s), percentiles, etc. Pero como ya sabemos, no estamos interesados en eso 100 niños concretos, sino en TODOS los niños, toda la POBLACIÓN. Para poder extraer conclusiones (INFERIR) acerca de la población (esto se verá en el tema 6), "necesitamos" asumir que nuestros datos provienen de una población que sigue un determinado modelo teórico (en este punto del temario es en el que nos encontramos actualmente). A veces podría no asumirse un modelo parámetrico pero la estadística no paramétrica queda fuera de nuestro alcance. También existen tests para probar si nuestros datos provienen de un determinado modelo, esto de nuevo queda fuera del alcance de este curso.

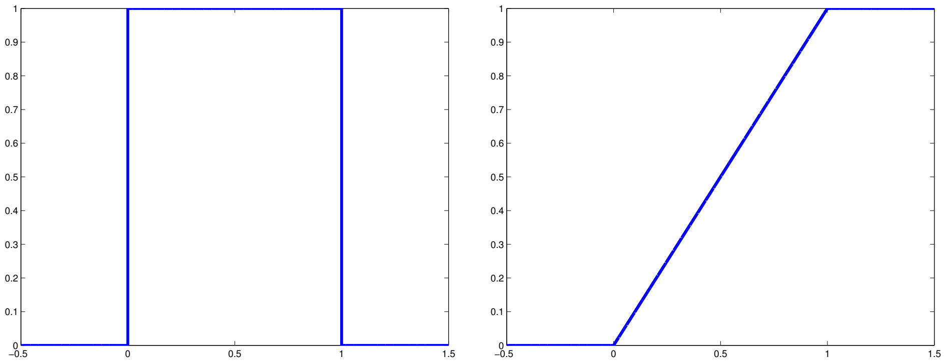
Las conclusiones que obtendremos vendrán dadas en términos probabilísticos (por ejemplo, el intervalo de confianza al 95 % para μ es ...) y serán conclusiones sobre descriptores de la población: media (μ), varianza (σ^2), etc., que en realidad, muy difícilmente se conocen.

En fotocopias tenéis algunas definiciones y propiedades.

5.2. Distribuciones continuas

5.2.1. Distribución uniforme(a,b)

Es la distribución que sigue una variable aleatoria X que toma valores en un intervalo [a,b] con la misma probabilidad. Por ejemplo, las calculadoras científicas con la tecla $\boxed{RAN\sharp}$ o $\boxed{Rnd}$ generan valores aleatorios de una variable uniforme entre 0 y 1. Su función de densidad y su función de distribución tienen la siguiente forma:



(a) (b)

Figura 5.2: (a) Función de densidad, f, de la Uniforme(0,1); (b) Función de distribución, F, de la Uniforme(0,1)

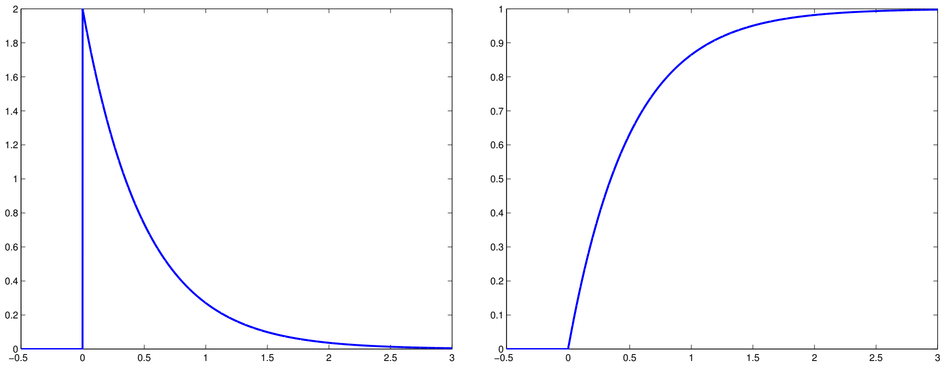
La función de densidad, de distribución acumulada, la media y varianza vienen dadas para una variable Uniforme(a,b) por:

$$f(x;a,b) = \begin{cases} \frac{1}{b-a} & \text{si } a < x < b \\ 0 & \text{en otro caso} \end{cases} \; ; \; \mu = \frac{a+b}{2}, \; \sigma^2 = \frac{(b-a)^2}{12}$$

$$F(x;a,b) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } x > b \end{cases}$$

5.2.2. Distribución exponencial(λ)

Es usada muchas veces para modelizar el comportamiento de variables aleatorias del tipo ”tiempo transcurrido hasta el fallo de un componente industrial” o ”tiempo que se tarda en completarse un proceso determinado”. La función de densidad y función de distribución de una exponencial de parámetro λ tienen la siguiente forma:



(a) (b)

Figura 5.3: (a) Función de densidad, f, de la Exponencial(0.5); (b) Función de distribución, F, de la Exponencial(0.5)

La función de densidad, de distribución acumulada, la media y varianza vienen dadas para una variable Exponencial(λ) por:

$$f(x;\lambda) = \begin{cases} 0 & \text{si } x \leq 0 \\ \lambda e^{-\lambda x} & \text{si } x > 0 \end{cases} \;;\; \mu = \frac{1}{\lambda},\; \sigma^2 = \frac{1}{\lambda^2}$$

$$F(x;\lambda) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

La distribución exponencial está relacionada con la Poisson de la siguiente forma: si el número de ocurrencias de un determinado fenómeno es una variable con distribución Poisson, el tiempo que pasa entre dos ocurrencias sucesivas es una variable con distribución exponencial.

La distribución Exponencial carece de memoria, se cumple $P(X > s+t|X > s) = P(X > t)$, en el contexto de "tiempos de vida" esto quiere decir que la probabilidad de fallar es independiente del pasado, el sistema no envejece. Aunque pueda parecer algo irreal, no es descabellado por ejemplo suponer que un fusible es "tan bueno como nuevo" mientras esté funcionado.

5.2.3. Distribución Weibull(α,β)

Otra de las distribuciones que se aplica además de la Exponencial a problemas de fiabilidad y "tiempos de vida de componentes - equipos", es la Weibull(α,β). De hecho, para $\beta = 1$, la Weibull se reduce a la Exponencial.

La función de densidad para Weibull(1,β) y distintos valores de β puede verse en el siguiente gráfico, $\beta > 0$ es un parámetro de forma y $\alpha > 0$ de escala.

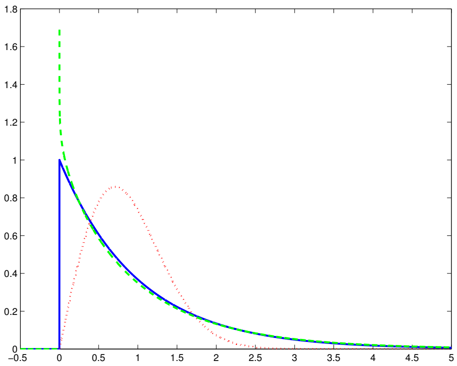


Figura 5.4: En azul y continua: Weibull(1,1), en rojo y puntos: Weibull(1,2), en verde y rayas: Weibull(1,0.95)

A continuación, aparece la expresión de su función de densidad:

$$f(x;\alpha,\beta) = \begin{cases} 0 & \text{si } x \leq 0 \\ \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} & \text{si } x > 0 \end{cases}$$

Como ya se ha dicho, la distribución Weibull puede emplearse para modelar el tiempo hasta presentarse un fallo en muchos sistemas físicos diferentes. Los parámetros de esta distribución permiten gran flexibilidad para modelizar sistemas en los que el número de fallos aumenta con el tiempo (por ejemplo, el desgaste), disminuye con el tiempo (algunos semiconductores) o permanece constante (fallos provocados por causas externas al sistema). En la siguiente página *http : //www.itl.nist.gov/div898/handbook/apr/apr.htm* podréis encontrar un capítulo dedicado a la fiabilidad.

5.2.4. Distribución Normal(μ,σ^2)

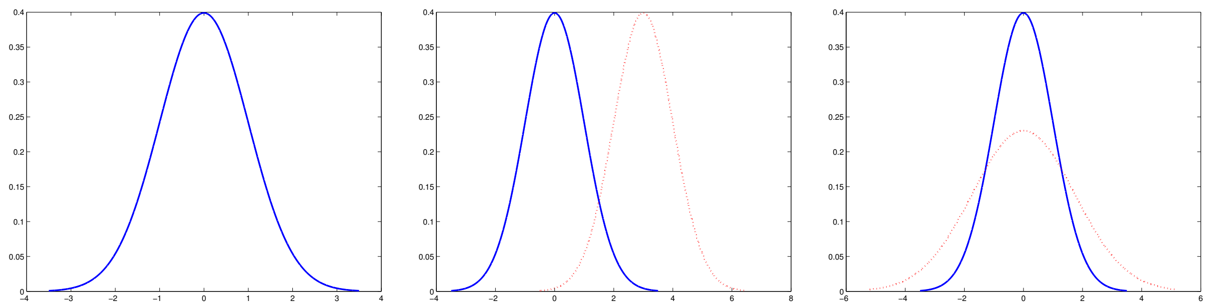
La distribución Normal o Gaussiana es muy importante puesto que se utiliza para modelar muchísimos fenómenos aleatorios; además incluso se usa para aproximar otras distribuciones. La distribución Normal aproxima lo observado en muchos procesos de medición sin errores sistemáticos, por ejemplo medidas físicas del cuerpo humano (X = "altura de las jóvenes españolas"**ejemplo 5.1.**, X = "altura de los jóvenes españoles"**ejemplo 5.2.**, X = "longitud del dedo índice de los niños"**ejemplo 1.3.**) (el próximo curso en Ergonomía lo comprobaréis nuevamente), medidas de calidad en muchos procesos industriales (**ejercicio 1 de la práctica 3**), etc. Una justificación de la frecuente aparición de la distribución Normal es el teorema central del límite (que veremos en este tema): cuando los resultados de un experimento son debidos a un conjunto muy grande de causas independientes que actúan sumando sus efectos, cada uno de ellos de poca importancia respecto al conjunto, es esperable que los resultados sigan una distribución Normal.

Ejemplo 1.3.: (ratón ergonómico). Este ejemplo, que nos ha ido *persiguiendo* durante todo el curso, nos va a permitir ver varios ejemplos más, de variables que podrían suponerse Normales.

Figura 5.5: Ratón óptico ergonómico 3M

Para comprobar científicamente las ventajas del ratón ergonómico frente al tradicional, se han realizado diversos estudios (en esta página, *http://www.animax.no/study*, puedes encontrar parte del trabajo). En esos estudios comparativos algunas de las variables empleadas y que podemos suponer Normales son: tiempo de movimiento de cada ratón, actividad eléctrica de varios músculos del antebrazo durante la utilización de cada ratón, intensidad del dolor medida en una cierta escala (VAS).

La función de densidad de una Normal de parámetros μ (media de la población) y σ^2 (varianza de la población, siempre positiva), que denotaremos N(μ,σ^2) (a veces, como en el Statgraphics , se denota N(μ,σ)), tiene la forma siguiente:



(a) Normal (0,1) (b) (c)

Figura 5.6: (a) Normal(0,1); (b) Un cambio en la media, supone una traslación: Normal(0,1) en azul y continua, Normal(3,1) en rojo y punteada; (c) Un cambio en la varianza, supone un cambio en la variabilidad, pero el área bajo la curva sigue siendo 1, por ello tienen distinta altura: Normal(0,1) en azul y continua y Normal(0,3) en rojo y punteada

Como puede apreciarse, la Normal (campana de Gauss) es simétrica respecto de la media (que en este caso coincide con la mediana y la moda), o sea, el coeficiente de asimetría valdrá cero y además el coeficiente de curtosis es 3 (recuerda apartado 1.4. del tema 1).

La función de densidad es:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , \quad x \in \mathbb{R}$$

La función de distribución acumulada es:

$$F(x;\mu,\sigma^2) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

La dejamos de esta forma, ya que un integrando de la forma e^{-z^2} no tiene primitiva. Por tanto, para calcularla o bien se emplea algún método numérico o se usan tablas, que es lo que haremos nosotros. Para ello necesitamos presentar la:

Distribución normal estándar: es aquella distribución normal con media 0 y varianza 1. La denotaremos mediante la letra Z.

Los valores que se recogen en las tablas (las tablas están en fotocopiadora) son para $N(0,1)$, además algunas calculadoras también permiten calcular probabilidades de una Normal estándar. La tabla nos proporciona:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}} dx \quad , \quad Z \sim N(0,1)$$

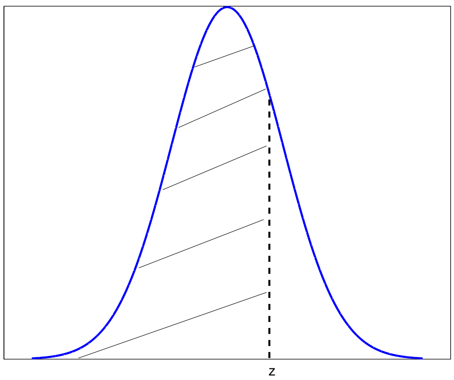


Figura 5.7: $\Phi(z)$

Ejemplo 5.3.: uso de la tabla de N(0,1)

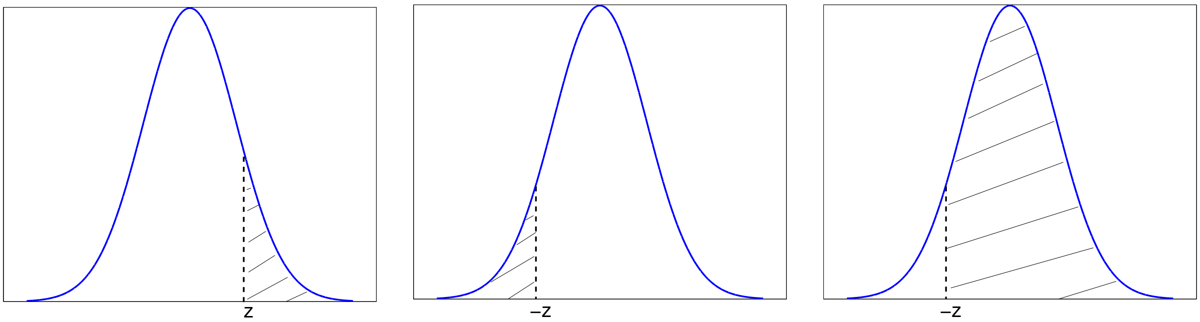
$P(Z \leq 1{,}96) =$

$P(Z \geq 0{,}53) =$

$P(Z \leq -2{,}33) =$

$P(Z \geq -2{,}33) =$

O sea, $P(Z \geq z)$ = 1 - $P(Z \leq z)$, $P(Z \leq -z)$ = 1 - $P(Z \leq z)$, $P(Z \geq -z) = P(Z \leq z)$. Ayúdate de un gráfico si lo necesitas.



(a) (b) (c)

Figura 5.8: (a) $P(Z \geq z)$; (b) $P(Z \leq -z)$; (c) $P(Z \geq -z)$

Veamos como con esa tabla podemos calcular cualquier probabilidad de cualquier Normal, con cualquier media μ y varianza σ^2, no necesariamente N(0,1).

Estandarización: Sea $X \sim$ N(μ,σ^2), podemos estandarizarla (o tipificarla) y convertirla en una N(0,1) de la siguiente forma:

$$Z = \frac{X - \mu}{\sigma}$$

(recuerda problema 9 del tema 1, hecho en clase). O sea, si $X \sim$ N(μ,σ^2), $P(a < X < b)$ $= P(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}) = P(Z < \frac{b-\mu}{\sigma})$ - $P(Z < \frac{a-\mu}{\sigma})$

[Fíjate que para estandarizar, dividimos por la desviación típica σ, NO por la varianza σ^2].

Ejemplo 5.4.: Sea $X \sim$ N(5,2). Calcula:

$P(X \leq 8) =$

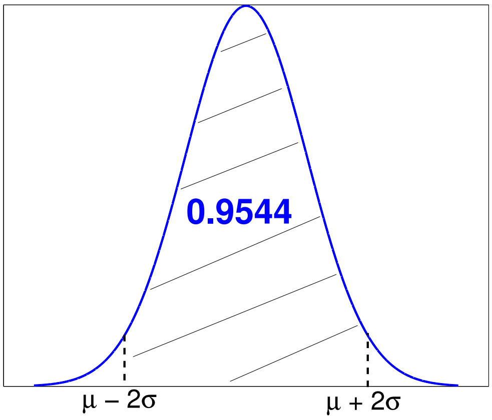
$P(4 \leq X \leq 8) =$

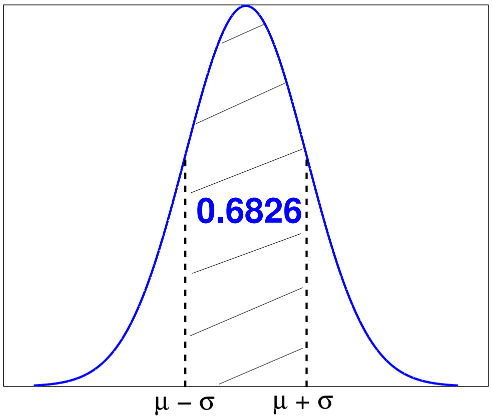
[Fíjate que como la Normal es simétrica respecto su media μ, para $X \sim$ N(μ,σ^2): $P(X \leq \mu)$ $= P(X \geq \mu) = 0.5$. Además, si $x \geq \mu$, $P(X \leq x) \geq 0.5$. También, si $x \leq \mu$, $P(X \leq x) \leq 0.5$. Siempre que tengas dudas, recurre a hacer una representación gráfica].

Ejemplo 5.5.: Si $X \sim$ N(μ,σ^2), la fracción (proporción) de números que están a 3 desviaciones de la media es 0.9972, no importa el valor de μ, ni σ^2:

$P(\mu - 3\sigma < X < \mu + 3\sigma) = P(\frac{\mu-3\sigma-\mu}{\sigma} < Z < \frac{\mu+3\sigma-\mu}{\sigma}) = P(-3 < Z < 3) = P(Z < 3)$ - $P(Z < -3) = 0.9986$ - (1 - $P(Z < 3)$) = 0.9986 - (1 - 0.9986) = 0.9972

Puedes comprobar que la fracción de números que están a 2 desviaciones de la media es 0.9544 y la fracción de números que están a 1 desviación de la media es 0.6826.





✍ **Ejemplo 5.6.:** Una compañía vende sacos de pienso de 5 kg. (según la etiqueta). Por cada saco que venda conteniendo menos de 5 kg., ha de pagar una multa. Las máquinas empaquetadoras no pueden garantizar un peso exacto de pienso por cada saco, así que el peso de los sacos se distribuye Normalmente con una media de 5.13 kg. y una desviación típica $\sigma =$ 0.08kg.

1. ¿Qué proporción de sacos pesa menos de 5kg.?

2. ¿Cuál es el peso w tal que el 80 % de los sacos pesa menos de w? (w es el 80-ésimo percentil)

3. La máquina puede ajustarse para cambiar el peso medio μ. Suponiendo que la desviación típica no cambia, ¿qué valor debería adoptar μ para que únicamente el 1 % de las bolsas pesen menos de 5 kg.?

Observación: *Aunque teóricamente la curva normal representa una distribución continua, a veces se usa para aproximadamente describir la distribución de una variable discreta. En esos casos, podría aplicarse la corrección de continuidad que veremos posteriormente en este tema, para así obtener una mayor precisión.*

5.3. Muestras aleatorias. Otros tipos de muestreo

Recordemos que nuestro objetivo es inferir sobre la POBLACIÓN. Nosotros sólo contamos con una muestra de la población. ¿Cómo generalizar más allá de un conjunto de datos particular? El primer paso para el desarrollo de una base para la inferencia estadística es encontrar un modelo probabilístico de las muestras que nos permita utilizarlas para inferir información sobre la población de la que se han extraído: el muestreo aleatorio simple.

Existen diversas técnicas de extracción de muestras de una población (como vereremos seguidamente). Nosotros nos centraremos en la más simple:

Muestreo aleatorio simple: se caracteriza por:

i) cada miembro de la población tiene la misma probabilidad de ser seleccionado

ii) las selecciones son independientes las unas de las otras.

Ejemplo 5.7.: Imaginemos que deseamos conocer el gasto en ocio (en un mes) de los jóvenes (18-30 años) españoles. Para ello extraemos una muestra de tamaño N (por ejemplo $N = 100$) por muestreo aleatorio simple (*pregunto el gasto a N jóvenes completamente al azar*). Si cada estudiante de la clase repitiera el experimento, tendríamos tantas muestras de tamaño N como estudiantes en la clase.

Por tanto, podemos considerar las variables aleatorias X_1, X_2, ..., X_N donde X_1 representa el valor (gasto) de la primera persona elegida (que variará de una muestra a otra), X_2 el valor de la segunda persona, ..., X_N el valor de la N-ésima persona.

Por la condición *i*), la distribución de cada X_i, $1 \le i \le N$, es la misma que la de la población (todas las variables X_i siguen la misma distribución). Por *ii*) X_1, X_2, ..., X_N son independientes (el conocimiento de una variable no aporta información acerca de los valores de la otra variable).

En consecuencia, X_1, X_2, ..., X_N, son independientes e idénticamente distribuidas (i.i.d) y constituyen una muestra aleatoria de tamaño N.

Estadístico: es cualquier función de las variables X_1, X_2, ..., X_N que constituyen una muestra aleatoria. Algunos ejemplos son:

Media de muestreo:

$$\overline{X} = \frac{X_1 + X_2 + ... + X_N}{N}$$

Varianza de muestreo:

$$S^2 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}$$

Un estadístico es una variable aleatoria por ser función de variables aleatorias, por lo cual tiene una distribución que se llama **distribución de muestreo**.

[Nota: denotamos con mayúsculas los estadísticos de muestreo por ser variables aleatorias, de esta forma se distinguen de las cantidades muestrales ($\overline{x}$ y s^2, por ejemplo) que vimos en el tema 1, que corresponden a una muestra concreta y tienen un valor numérico concreto.]

Aunque a lo largo de este curso siempre supondremos que nuestra muestra se ha obtenido por muestreo aleatorio simple, existen otros tipos de muestreo.

Un objetivo primordial de los procedimientos de muestreo es conseguir que la muestra sea representativa de la población ("como la población, pero en tamaño reducido"). Acabamos de presentar el muestreo aleatorio simple, que se usará cuando los elementos de la población sean homogéneos respecto a la característica a estudiar. Pero si disponemos de algún tipo de información sobre la población sería coveniente emplearla a la hora de seleccionar la muestra. Un ejemplo clásico son las encuesta de opinión (en "Diseño conceptual": obtención de información conocida a nivel personal), donde los elementos (personas) de la población son (o pueden serlo) heterogéneas en razón a su sexo, edad, profesión, etc. En estos casos interesaría que la muestra tuviera una composición análoga a la población, lo cual se conseguiría mediante muestreo estratificado.

Muestreo estratificado: los elementos de la población se dividen en clases o estratos. La muestra se toma asignando un número de miembros a cada estrato (pueden usarse distintos criterios: proporcional al tamaño relativo del estrato en la población, proporcional a la variabilidad del estrato, considerando costes, ...) y escogiendo los elementos por muestreo aleatorio simple dentro de cada estrato.

Ejemplo 5.7.: en este ejemplo, estaría bien dividir los elementos según su nivel económico, y por ejemplo dividirlos según la zona de la ciudad en que habiten: zona centro (clase alta), zona intermedia (clase media), barrios periféricos (clase baja).

Ejemplo 5.8.: queremos conocer la resistencia de los plásticos que hay en un almacén. Los plásticos provienen de dos fabricantes distintos. Sería mejor considerar dos estratos (cada fabricante), que los plásticos como un todo y muestrear sin distinción, porque puede que la distribución sea diferente en cada estrato.

Muestreo por conglomerados: se utiliza si la población se encuentra de manera natural agrupada en conglomerados, que podemos considerar como una muestra representativa de la población. La muestra se toma seleccionando algunos conglomerados al azar y dentro de ellos analizando todos sus elementos o una muestra aleatoria simple.

Ejemplo 5.7.: siguiendo con este ejemplo, dentro de cada estrato (zona de la ciudad) podemos hacer divisiones en calles, las calles serían conglomerados ya que podemos considerarlas homogéneas respecto a la característica a estudiar.

Ejemplo 5.9.: supongamos que queremos analizar el diámetro de unas tuercas que tenemos almacendas en cajas. Sería más conveniente seleccionar al azar unas cajas y dentro de ellas realizar un muestreo aleatorio simple que llevar a cabo un muestreo aleatorio simple, pues esto implicaría seguramente abrir muchas más cajas.

Las ideas de estratificación y conglomerado son opuestas: la estratificación funciona tanto mejor cuanto mayor sean las diferencias entre los estratos y más homogéneos sean éstos internamente; los conglomerados funcionan si hay muy pocas diferencias entre ellos y son muy heterogéneos internamente.

Muestreo sistemático: cuando los elementos de la población están ordenados en listas, se usa el muestreo sistemático. Si la población es de tamaño N y la muestra deseamos que sea de tamaño n, tomaremos k como el entero más próximo a N/n, elegiremos un elemento al azar entre los k primeros, por ejemplo el n_1, después tomaremos los elementos $n_1 + k$, $n_1 + 2k$, etc, hasta completar la muestra.

Como se ha visto en el **ejemplo 5.7**, los distintos tipos de muestreo pueden emplearse conjuntamente. Por ejemplo, en el análisis de diámetros de tuercas en cajas provenientes de dos fabricantes distintos (juntamos las ideas de los **ejemplos 5.8 y 5.9**).

Control de recepción: un campo donde el muestreo juega un papel fundamental es en el control de recepción que trata de comprobar que los productos cumplan las especificaciones de calidad.

El más empleado es el control de recepción por atributos, en el que se inspeccionan por muestreo las unidades de un lote. Se seleccionan artículos de cada lote y se toma una decisión con base a dicha muestra respecto a si se acepta o se rechaza el lote, según el número de unidades defectuosas que contenga.

Para resolver esta cuestión podemos emplear los llamados planes de muestreo. Éstos podemos clasificarlos en:

a) Planes de aceptación/rechazo: los más conocidos son:

- las normas japonesas JIS Z 9002
- las normas norteaméricanas Milatary Standard MIL- STD- 105D; UNE 66020

Éste último tiene en cuenta el tipo de inspección, así como el rigor de inspección.

Los muestreos pueden ser simples, dobles, múltiples y (en su caso extremo) secuencial (un muestreo es secuencial cuando después de cada observación se toma una de las siguientes decisiones: aceptar el lote, rechazarlo o seguir con el muestreo).

b) Planes de control rectificativo: los lotes rechazados se inspeccionan al 100 % sustituyendo los elementos defectuosos. Los más usados son los de Dodge-Romig.

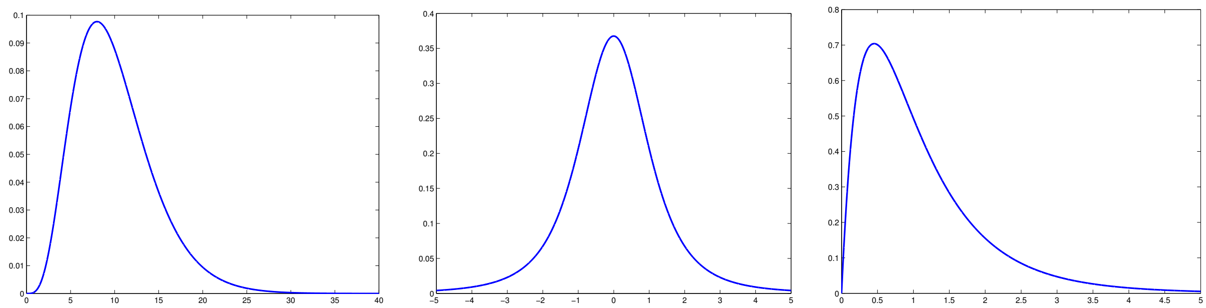
Las tablas de estos planes y una explicación más detallada sobre muestreo podéis encontrarlos por ejemplo en Johnson, Peña (vol. I) y sobre todo en cualquier libro sobre Control de Calidad. En ellos podréis también encontrar la relación entre el índice de capacidad (para cumplir con las especificaciones de calidad) y la frecuencia de inspección.

5.4. Teorema central del límite

En fotocopias

5.5. Distribuciones en el muestreo

En fotocopias.



(a) χ^2_{10} (b) t_3 (c) $F_{(4,20)}$

Figura 5.9: (a)χ^2 Chi-cuadrado; (b) t de Student; (c) F de Snedecor

Problemas del tema 5

Ejemplo La duración de las bombonas de butano de 40 kg de carga tiene aproximadamente una distribución normal. La probabilidad que una bombona dure más de 220 horas es de 0.1587 y la probabilidad que dure menos de 190 horas es de 0.3085. Calcula:

1. La media y la desviación típica de la distribución.
2. La probabilidad que dure entre 195 y 215 horas.
3. La duración máxima que puede garantizarse con un 80 % de confianza.

1. $0.1587 = P(X > 220) = P(\frac{X-\mu}{\sigma} > \frac{220-\mu}{\sigma}) = P(Z > \frac{220-\mu}{\sigma}) = 1$ - $P(Z \leq \frac{220-\mu}{\sigma})$, por tanto $P(Z \leq \frac{220-\mu}{\sigma}) = 0.8413$ y entonces mirando la tabla: $\frac{220-\mu}{\sigma} = 1$

   $0.3085 = P(X < 190) = P(\frac{X-\mu}{\sigma} < \frac{190-\mu}{\sigma}) = P(Z < \frac{190-\mu}{\sigma}) = 1$ - $P(Z \geq \frac{190-\mu}{\sigma})$, por tanto $P(Z \geq \frac{190-\mu}{\sigma}) = 0.6915$ y entonces mirando la tabla y teniendo en cuenta la simetría de la normal: $-\frac{190-\mu}{\sigma} = 0{,}5$

   Tenemos 2 ecuaciones y 2 incógnitas: 220 - $\mu = \sigma$, $-190 + \mu = 0.5\sigma$, se resuelve y se obtiene $\mu = 200$ y $\sigma = 20$

2. $P(195 < X < 215) = P(\frac{195-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{215-\mu}{\sigma}) = P(\frac{195-200}{20} < Z < \frac{215-200}{20}) = P(-0{,}25 < Z < 0{,}75) = P(Z < 0{,}75)$ - $P(Z < -0{,}25) = P(Z < 0{,}75)$ - $(1 - P(Z < 0{,}25))$ = (tabla) = 0,7734 - $(1 - 0{,}5987)$ = 0.3721

3. $0.8 = P(X \geq a) = P(\frac{X-\mu}{\sigma} \geq \frac{a-\mu}{\sigma}) = P(Z \geq \frac{a-200}{20}) = P(Z \leq -\frac{a-200}{20})$, consultando la tabla: $0{,}84 = -\frac{a-200}{20}$, y por tanto a = 183,2 horas

1. La duración de cierto motor pequeño sigue una distribución Normal con media 10 años y desviación típica 1.5 años.

   a) ¿Cuál es la probabilidad de que un motor dure menos de 7 años?
   b) El fabricante se compromete a cambiar gratis todos aquellos motores que fallen durante el tiempo de garantía. Si está dispuesto a reemplazar sólo el 3 % de los motores, ¿de qué duración debe ser la garantía que ofrezca?

   (Sol: 0.0228, 7.18 años)

2. Cierto tipo de batería de almacenamiento dura, en promedio 3.0 años, con una desviación estándar de 0.5 años. Suponiendo que las duraciones de la batería se distribuyen normalmente, encuentra la probabilidad de que una batería dada dure menos de 2.3 años. Calcula el percentil 5 de la distribución.

(Sol: 0.0808, 2.18 años)

3. Una empresa eléctrica fabrica focos que tienen una duración, antes de fundirse, que se distribuye normalmente con media igual a 800 horas y una desviación típica de 40 horas. Encuentra la probabilidad de que un foco se funda entre 778 y 834 horas. Calcula el percentil 10 de la distribución.

   (Sol: 0.5111, 748.8 horas)

4. En un proceso industrial el diámetro de un cojinete es una parte componente importante. El comprador establece que las especificaciones del diámetro sean 3.0 ± 0.01 cm. La implicación es que ninguna parte que caiga fuera de estas especificaciones se aceptará. Se sabe que en el proceso, el diámetro de un cojinete tiene una distribución Normal con media 3.0 y una desviación estándar 0.005. En promedio, ¿cuántos cojinetes se descartarán? Calcula el percentil 85 de la distribución.

   (Sol: 4.56 %, 3.0052)

5. Se utilizan medidores para rechazar todos los componentes donde cierta dimensión no está dentro de la especificación $1.50 \pm d$. Se sabe que esta medición se distribuye de forma Normal con media 1.50 y desviación estándar 0.2. Determina el valor d, tal que las especificaciones "cubran" 95 % de las mediciones.

   (Sol: 0.392)

6. Cierta máquina fabrica resistores eléctricos que tienen una resistencia media de 40 ohmios y una desviación estándar de 2 ohmios. Supongamos que la resistencia sigue una Normal y se puede medir con cualquier grado de precisión, ¿qué porcentaje de resistores tendrán una resistencia que exceda 43 ohmios? Calcula el percentil 95 de la distribución.

   (Sol: 6.68 %, 43.28)

7. Un fabricante produce pistones, cuyos diámetros se distribuyen según una Normal de media 5 cm y una desviación típica 0.001 cm. Para que un pistón sirva debe encontrase entre 4.998 y 5.002 cm. Si el diámetro del pistón es menor que 4.998 cm se rechaza; si es mayor que 5.002 cm se recicla. ¿Con qué probabilidad sirve? ¿Cuál es la probabilidad de rechazarlo?¿Y de reciclarlo?

   (Sol: 0.9544, 0.0228, 0.0228)

8. El tiempo necesario para armar cierta unidad es una variable aleatoria Normal con media 30 minutos y varianza 4 minutos. Determina el tiempo de armado de tal manera que la probabilidad de exceder éste sea 0.2. Además, determina la probabilidad que se tarde menos de 27 minutos, entre 27 y 32 minutos y que se tarde más de 32 minutos.

   (Sol: 31.68, 0.0668, 0.7745, 0.1587)

9. La demanda semanal de bombillas en una ferretería es una variable N(300,81). Suponiendo que se hacen pedidos semanales, calcula la cantidad de bombillas que necesitamos tener a principio de semana para poder satisfacer la demanda un 95 % de las semanas. ¿Cuál es la probabilidad que la demanda esté entre 280 y 310 bombillas?

   (Sol: $\lceil 314.76 \rceil = 315$ bombillas, 0.8533)

10. La dimensión principal de ciertas piezas tiene una distribución Normal(150, σ^2=0.4) y el intervalo de tolerancia es (149,150.4). Se pide:

    a) La proporción esperada de defectuosas resultantes de dicho proceso
    b) Se toman 15 piezas, calcula la probabilidad de que 9 sean aceptables (o sea, 6 defectuosas)
    c) Calcula el número esperado de piezas defectuosas en 50 piezas
    d) El percentil 40 de la distribución

    (Sol: 0.3214, 0.1683, 16,149.84)

11. En una gran fábrica, los tubos fluorescentes se mantienen encendidos día y noche. Las luces se funden con un promedio de 120 al mes (30 días). Suponiendo que se reemplazan al instante y que el número de roturas sigue una distribución Poisson. Calcula las probabilidades de que:

    a) Se rompan entre 100 y 120 luces (incluidas) en un mes
    b) No se rompa niguna luz un día

    (Sol: 0.4892, 0.0183)

12. Supongamos que el tiempo, en minutos, desde el momento en que llegas a la parada del bus hasta que llega el primer autobús se modeliza mediante una distribución exponencial con $\lambda = 0.37$. Encuentra las probabilidades de los siguientes sucesos:

    a) El siguiente bus llega entre 2 y 4 minutos después que tú llegas
    b) Esperas más de 2 minutos al primer bus
    c) El primer bus llega dentro de los primeros 90 segundos
    d) No pasa ningún bus en los primeros 5 minutos de estar en la parada
    e) Pasa al menos un bus en el primer minuto

    (Sol: 0.2495, 0.4771, 0.4259, 0.1572, 0.3093)

13. La gente que frecuenta un cierto karaoke tiene una probabilidad de 0.4 de levantarse y cantar. Cierto día hay 150 personas dentro del karaoke. ¿Cuál es la probabilidad que al menos 10 personas se levanten y canten? (supongamos que cada persona toma la decisión independentment de las otras).

    (Sol: 0.9988)

14. Un emisor envía una cierta señal. El receptor no lo recibe nítidamente, sino que un determinado ruido que puede modelizarse como una distribución uniforme en el intervalo [-1,1] se añade a la señal. ¿Cuál es la probabilidad que al recibir la señal, ésta tenga un ruido

    a) mayor que 0.5?

b) entre -0.25 y 0.75?
c) menor que 0.25?

(Sol: 0.25, 0.5, 0.625)

15. En una cierta fabricación mecánica el 96 % de las piezas resulta con longitudes admisibles (dentro de las tolerancias), un 3 % defectuosas cortas y 1 % defectuosas largas. Calcula la probabilidad de que:

 a) En un lote de 250 piezas sean admisibles 242 o más
 b) En un lote de 100 piezas sean cortas 4 o menos
 c) En un lote de 1000 piezas haya entre 6 y 12 inclusive largas

 (Sol: 0.3156, 0.815, 0.71)

16. Se ha comprobado que la duración de vida de ciertos elementos sigue una distribución exponencial con media 8 meses. Se pide:

 a) Calcula la probabilidad de que un elemento tenga una vida entre 3 y 12 meses
 b) El percentil 95 de la distribución

 (Sol: 0.46, 23.97)

17. El espesor del borde de un componente de una aeronave está distribuido de manera uniforme entre 0.95 y 1.05 mm.

 a) ¿Cuál es la proporción de bordes cuyo espesor es mayor que 1.02mm?
 b) ¿Qué espesor está excedido por el 90 % de los bordes?
 c) ¿Cuál sería la media y varianza del espesor del borde?

 (Sol: 0.3; 0.96; 1 y 0.00083)

18. Una compañía aérea observando que, en promedio, el 12 % de las plazas reservadas no se cubren decide aceptar reservas por un 10 % más de las plazas disponibles en aviones de 450 plazas. Calcula la proporción de vuelos en los que algún pasajero con reserva no tiene plaza.

 (Sol: 0.0197 $\approx$ 2 %)

19. Un aparato de medida da una lectura que puede considerarse distribuida según una N(μ,σ^2), donde μ es el valor real de la magnitud deseada y $\sigma^2 = 9$. Para mejorar la precisión se decide tomar la media $\overline{X}$ de n medidas.

 a) Calcula el mínimo n necesario para que la varianza de $\overline{X}$ sea inferior o igual a 0.1.
 b) Con el valor de n obtenido en el apartado anterior, calcula la probabilidad que $\overline{X}$ se aparte del valor real en más de ± 0.2 unidades.

 (Sol: 90, $P(|\overline{X} - \mu| > 0{,}2)$ = 0.5286)