Karla Marmolejo

Tema 1
Conceptos Básicos y Organización de Datos

VARIABLE	Conjunto de valores resultantes de medir una característica de interés sobre cada elemento individual de una población o muestra. Representación numérica de una característica que presenta más de una modalidad (valor) de un conjunto determinado. Si una característica tiene una única modalidad se llama **constante**.
MEDICION	Proceso por el cual se asignan números a objetos o sucesos según determinadas reglas
CARACTERISTICA	Cualquier propiedad de objetos o personas que deseemos estudiar
MODALIDAD	´
ESCALA DE MEDIDA	Conjunto de reglas o modelos desarrollados para la asignación de números a las variables.

Diferentes escalas de medida en función de la variable a medir

Escala Nominal	Asignación arbitraria de números o símbolos a cada uno de los valores de la variable; la única relación que tiene en cuenta es de igualdad y desigualdad —la pertenencia o no a una categoría determinada.		
	Los valores de la variable se denominan categorías y las variables son cualitativas o categóricas	Dicotómicas	Presenta sólo dos categorías (sexo)
		Politómicas	Más de dos categorías (estado civil)
Escala Ordinal	Se asignan números a objetos para indicar la extensión relativa en que se posee esa característica; los datos se pueden jerarquizar u ordenar, pero sin indicar la distancia que hay entre las posiciones—los números se asignan sólo para indicar el orden.		
	También reciben el nombre de variables cuasicuantitativas		

Escala de Intervalo	Ordenan los objetos según la magnitud del atributo que representan y proveen intervalos iguales entre las unidades de medida y la distancia entre los distintos valores consecutivos de la variable es la misma.	VARIABLES CUANTITATIVAS	Contínuas	Dados dos valores siempre se puede encontrar un tercer valor que esté incluido entre los dos primeros (peso)
	La caracteriza la existencia de una unidad de medición común y constante—se pueden sumar o restar			
	El origen es arbitrario y no refleja en ningún momento la ausencia de la magnitud que estamos midiendo			
Escala de Razón	Tiene todas las características de una medida de intervalo, pero sí se le puede asignar un punto de origen verdadero de valor cero, es un **valor absoluto** y no arbitrario		Discretas	Adopta valores aislados fijos—no se puede tomar ningún intermedio entre dos valores consecutivos (No. hijos)
	Los números asignados admiten las relaciones de igualdad-desigualdad, orden, suma, resta, multiplicación y división.			

Es muy importante la definición operativa de una variable (cómo se define y se registra) porque puede determinar su nivel medida. La mayoría de las variables psicológicas se considera que están medidas en una escala de intervalo. Una vez recabada la información a través del proceso o escala de medida y recogido los datos correspondientes, se dispone de un listado o base llamado **matriz de datos**. La generación de una base de datos supone la codificación previa de las observaciones, la introducción de los datos en algún programa informático, la depuración de los datos ya grabados y eventualmente la realización de transformaciones de variables que faciliten su posterior tratamiento estadístico.

Los **datos perdidos** son valores que no han sido registrados, habitualmente porque el participante no ha consignado ese dato. Existen procedimientos de imputación de datos, basados en los valores válidos de otros casos que se utilizan en ocasiones en variables cuantitativas.

Un **dato atípico** es un valor muy diferente del resto de valores de la misma variable—suelen ser ocasionados por errores al introducir datos o por valores extremos. Los datos atípicos distorsionan los resultados de los análisis, y por ello es importante identificarlos y tratarlos de manera adecuada, generalmente excluyéndolos.

Variable independiente	Cualquier suceso que sospechamos es la causa de otro—son las que manipula el investigador
Variable dependiente	Medida utilizada para estudiar la variable independiente--el resultado
Variable extraña	Aquéllas que pueden influir sobre la variable dependiente, pero no nos interesa estudiar sus efectos

Para referirnos a un valor cualquiera de la variable X se utiliza el subíndice i (Xi), siendo n el número de elementos que componen la muestra. i=1, 2, 3, 4… n

Karla Marmolejo

Tema 1
Conceptos Básicos y Organización de Datos

Estadística

Teórica	Se ocupa de los aspectos matemáticos formales y normativos
Aplicada	Constituye la aplicación a un campo concreto; también se le llama Análisis de Datos

Método Científico

Procedimiento estructurado que utiliza la ciencia para la ampliación del conocimiento. Consiste en dar razón sistemática, empírica y experimental de los fenómenos. Se convierte en la fuente de conocimiento más utilizada en el S. XVII. La Psicología lo utiliza para acercase a su objeto de estudio: **La Conducta**	
Sistemático	Tiene etapas definidas (sigue un sistema)
Replicable	Los datos obtenidos mediante su uso tienen que poder ser replicados o refutados por cualquier investigador interesado.

Pasos del Método Científico

1. Planteamiento del Problema	
2. Hipótesis—solución tentativa al problema	
3. Procedimiento para recogida de datos	
4. Análisis de datos	De estos se ocupa la Estadística
5. Discusión de los resultados en busca de conclusiones	
6. Elaboración de informe investigación	

Estadística

Rama de las matemáticas que se encarga del estudio de determinadas características en una población, recogiendo datos, agrupándolos, organizándolos en tablas, representándolos en gráficas y analizándolos para sacar conclusiones			
Descriptiva	Se organizan y resumen conjuntos de observaciones cuantificadas procedentes de una muestra o de la población total. Puede hacerse mediante tablas, gráficos o valores numéricos.	**Una variable**	Podemos recurrir a estadísticos que nos indicarán cuáles son los valores más habituales de esa variable—índices de tendencia central, estadísticos de variabilidad, estadísticos de asimetría, curtosis
		Dos variables	Se utilizan índices que nos indiquen hasta qué punto están ambas variables relacionadas entré sí así como procedimientos para predecir el valor de una en función de la otra—coeficientes de correlación y ecuaciones de regresión
Inferencial	Se realizan inferencias acerca de una población basándose en los datos obtenidos a partir de una muestra. Procedimientos—el cálculo de probabilidades	**Población**	el conjunto de todos los elementos que cumplen una determinada característica objeto de estudio; lo habitual es que se desee conocer un **parámetro**—propiedad descriptiva que se denota con letras griegas.
		Muestra	un subconjunto cualquiera de una población; se estudia la característica mediante **estadísticos** para estimar los parámetros de la población—propiedad descriptiva que se denota con legras latinas. Un estadístico se puede utilizar para estimar algún parámetro de la población.
			Las conclusiones obtenidas de una muestra sólo servirán para el total de una población si la muestra es **representativa**—se utilizan métodos de muestreo **probabilístico**
			También existen muestras **no probabilísticas**—las muestras de conveniencia o incidental (personas cercanas al investigador de más fácil acceso)

Karla Marmolejo

Tema 1
Conceptos Básicos y Organización de Datos

DESCRIPCION DE LAS VARIABLES

Distribución de frecuencias	Habitualmente el número de datos es elevado, por lo que se hace necesario organizar la información mediante una distribución de frecuencias —una tabla en la que se resume la información disponible de la variable		
	Una distribución de frecuencias es una representación de la relación entre un conjunto de medidas exhaustivas y mutuamente excluyentes y la frecuencia de cada una de ellas		
	Funciones	Ofrecer la información necesaria para realizar representaciones gráficas	
		Facilitar los datos para obtener los estadísticos muestrales	
Descripción de variable cualitativa	**Frecuencias absolutas (n_i)**	número de veces que se repite cada uno de los valores de una variable. La suma de todas las frecuencias absolutas representa el total de la muestra (n).	
	Frecuencias relativas o proporciones (p_i)	Se obtiene dividiendo la frecuencia absoluta n_i entre el número total de observaciones (n). $p_i = n_i / n$	
	Porcentajes (Pi)	La frecuencia relativa expresada en términos de porcentaje (P_i), multiplicar por 100. $P_i = p_i * 100$	
	Representación gráfica	mediante diagrama de barras o sectores (pie chart), son los más habituales.	
		Diagrama de barras conjunto	se utiliza cuando al menos una de las dos variables es cualitativa.
			Los datos se organizan en una tabla de entrada doble. Se sitúan los valores de una variable en las filas y los de la otra en las columnas. Una vez construida la tabla, se procede a dibujar los gráficos por columnas.
Descripción de variables ordinales o cuasicuantitativas	Misma distribución de frecuencias que la variable cualitativa, pero los valores deben ir situados de acuerdo al orden de las categorías, empezando por el valor más bajo o el más alto		
	Se añaden tres columnas más	**Frecuencia absoluta acumulada (n_a)**	número de veces que se repite cada valor o cualquiera de los valores inferiores.
		Frecuencia relativa acumulada (p_a)	cociente entre la frecuencia absoluta acumulada y el total de observaciones. $p_a = n_a / n$
		Porcentaje acumulado (P_a)	$P_a = p_a * 100$
	Representación gráfica	Diagrama de barras	Eje vertical—**ordenadas** o eje Y
		Diagrama de sectores (pie chart)	Eje horizontal—**abscisas** o eje X
	Cuando se obtienen una gran cantidad de filas con frecuencia absoluta de 0 ó 1, o si la modalidad es muy amplia, se recurre a la agrupación en **intervalos**—la formación de grupos de valores consecutivos de la variable. Primero se tiene que decidir el número de intervalos y la amplitud que tendrá la distribución de frecuencias, tratando de encontrar el equilibrio entre la precisión y la manejabilidad de los datos.		
	Los extremos de cada intervalo son los **límites aparentes**, para cada intervalo hay un límite inferior y un límite superior. Estos deben ser de la misma unidad que los valores de las variables (enteros o decimales). A un intervalo que no tiene límite inferior o superior se le llama **intervalo abierto**.		

Karla Marmolejo

Descripción de variables cuantitativas	Debido a la discontinuidad entre límites aparentes, es necesario calcular los **límites exactos o reales**. LIE se calcula **restando** al límite inferior aparente media unidad de medida; LSE se calcula **sumando** media unidad de medida al límite superior aparente.					
	Cuando los límites aparentes contienen decimales, los valores a sumar y restar dependerán del número de decimales—para enteros su mitad es 0.5, para 0.1 es 0.05, para 0.01 es 0.005 y así sucesivamente					
	El **punto medio** se calcula sumando los límites inferior y superior (ya sean los exactos o los aparentes) y dividiendo entre 2.					
	La **amplitud del intervalo** es la diferencia entre el LSE y el LIE.					
	Un gráfico es una forma rápida de visualizar un conjunto de datos o distribución de frecuencias. El sistema de coordenadas más habitual está formado por el **eje vertical (Y)** y el **eje horizontal (X)**. El **origen** es el punto donde se juntan ambos ejes.					
	Discreta (enteros—no hay intermedios)	Diagrama de barras—se puede utilizar también para variables nominales y ordinales. En el eje X van los valores y en el eje Y las frecuencias.				
		Diagrama de líneas o polígono de frecuencias—se construye situando un punto a una altura proporcional a la frecuencia en cada valor o en el punto medio de cada intervalo y luego se unen los puntos para formar una línea. También se puede usar en frecuencias acumuladas.				
		Diagrama de dispersión o nube de puntos—se utiliza en el caso de dos variables cuantitativas (discreta o continua). Se utiliza un eje XY—para cada par de datos se localiza la intersección entre ambas variables y se marca con un punto. Es una manera rápida de hallar relaciones lineales entre dos variables.				
	Continua (se puede encontrar un tercer valor entre los dos valores dados)	Agrupadas en intervalos	**Histograma**	Una extensión del diagrama de barras que dibuja los rectángulos unidos entre sí, indicando que existe una continuidad en los valores de las variables		
				Un gráfico de variable continua dividida en intervalos en los que se eleva un rectángulo con área proporcional a su frecuencia. En el eje horizontal se sitúan los límites exactos o punto medio.		
	Mediante índices estadísticos	**Tendencia central**	Distribución que se refiere al lugar donde se centra una distribución particular en la escala de valores			
		Variabilidad	Propiedad que se refiere al grado de concentración de los valores entre sí o con respecto a un valor central de la distribución de frecuencias		**Homogénea**	Poca variabilidad—los valores de la distribución están cercanos al promedio
					Heterogénea	Mucha variabilidad—los valores se dispersan mucho con respecto al promedio
		Forma de la distribución	**Asimetría o sesgo**	El grado en que los datos se reparten equilibradamente por encima y por debajo de la tendencia central	**Simétrica**	Al dividirla en dos partes iguales, las dos mitades se superponen
					Asimetría positiva	La mayor concentración de puntuaciones se produce en la parte baja de la escala—el apuntamiento hacia la izquierda
					Asimetría negativa	La mayor parte de las puntuaciones se sitúan en la parte alta de la escala—el apuntamiento hacia la derecha
			Curtosis	Grado de apuntamiento de los datos (lo "picudo" de la gráfica)	**Leptocúrtica**	Distribución de frecuencias muy apuntada
					Platicúrtica	Distribución de frecuencias muy aplastada
					Mesocúrtica	Grado de apuntamiento es intermedio

Karla Marmolejo

Tema 2
Indices de Tendencia Central y Posición

La Tendencia Central

Se puede resumir en un valor o puntuación que refleje esa tendencia central de la distribución y que represente al conjunto de observaciones		
Nos interesa calcular un valor central que actúe como resumen numérico para representar el conjunto de datos. Se denominan medidas, índices o estadísticos de tendencia central. Permiten representar toda distribución de frecuencias con un único valor y facilitan la comparación de diferentes conjuntos de puntuaciones de una variable.		
Indices de Tendencia Central	Media aritmetica	
	mediana	
	moda	
Medidas de Posición	Son útiles para informar sobre la posición relativa en la que se encuentra un sujeto con respecto al conjunto al que pertenece, a partir de su puntuación en la variable.	Percentiles
		Cuartiles
		Deciles

Karla Marmolejo

Tema 2
Indices de Tendencia Central y Posición

Media aritmética

También llamada promedio o media, el más conocido y usado		
Indica la tendencia general de una distribución de frecuencias de una variable y es el valor central alrededor del cual están la mayoría de las observaciones		
Denotada por X, se define como la suma de todos los valores observados de la variable divididos pro el número total de observaciones		
$X=X_1+X_2+X_3+\ldots+Xn = \sum Xi/n$		
Xi es el valor que toma la variable en el sujeto i		
n es el número total de observaciones		
Sólo se utiliza con variables cuantitativas de intervalo o de razón		
Para una distribución de frecuencias, hay que multiplicar cada frecuencia absoluta por el valor Xi		$X=(\sum niXi)/n$ n=número total de observaciones Xi=el valor i de la variable X o el punto medio del intervalo ni es la frecuencia absoluta del valor o intervalo.
Media a partir de una distribución de frecuencias relativas		$X=(\sum piXi)$ Xi=el valor i de la variable X o el punto medio del intervalo ni es la frecuencia absoluta del valor o intervalo.
En el caso de una distribución de frecuencias agrupadas en intervalos, se calcula igual, teniendo en cuenta que los varones de Xi serán los puntos medios de cada intervalo.		
Con las frecuencias absolutas acumuladas no se puede calcular la media, es necesario obtener primero las frecuencias absolutas		
Propiedades	En una distribución, la suma de las desviaciones de cada valor con respecto a la media es igual a cero. $\sum(Xi - X) = 0$	
	Si se disponen de un número mayor de observaciones en el que se repiten valores, agrupadas o no en intervalos, la expresión es $\sum ni(Xi - X) = 0$ (hay que tener en cuenta la frecuencia absoluta de cada valor ni, el número de veces que aparece cada puntuación o intervalo)	
	Si a cada puntuación Xi de la variable X le sumamos una constante a (elegida arbitrariamente), la media de las nuevas puntuaciones es igual a la media de X más la contante Yi = Xi + a, entonces Y = X +a	
	Si cada puntuación Xi de la variable X se multiplica por una constante b (elegida arbitrariamente), la media de las nuevas puntuaciones es igual a la media de X multiplicada por la constante. si Yi = b.X entonces Y = b.X	
En Psicología se utiliza con frecuencia este tipo de transformaciones de puntuaciones en la presentación de resultados para evitar valores negativos y decimales en las variables de interés.		
La media ponderada (de j muestras)	La media de todas las observaciones juntas de varios grupos en una variable $Xp = (n_1 \cdot X_1+n_2.X_2+n_3.X_3\ldots+nj+Xj)/n$	
	La media de cada grupo tiene un peso en la media total que está en función del tamaño de la muestra o número de casos de cada grupo—media ponderada; es necesario tener en cuenta el peso de la media.	
	También se utiliza para obtener la media global en una puntuación que se basa en distintas pruebas a las que se les ha otorgado pesos diferentes, en función de su importancia en la puntuación final. $Xp = (p_1.X_1+p_2.X_2)/(p_1+p_2)$	
Limitaciones	Cuando los datos están agrupados en intervalos, la media no se puede calcular si el intervalo máximo no tiene límite superior y/o si el intervalo mínimo no tiene límite inferior.	
	La media es sensible a la existencia de unas pocas observaciones con valore extremos en la distribución de frecuencias. Esta circunstancia se da en distribuciones marcadamente asimétricas, no es recomendable la utilización de la media—afecta su representatividad como valor central	

Karla Marmolejo

Tema 2
Indices de Tendencia Central y Posición

La Mediana

La mediana es mejor alternativa cuando la distribución es asimétrica—no se ve afectada por valores extremos ya que en su cálculo no intervienen todos los valores, sino únicamente los que ocupan las posiciones centrales.			
Se puede obtener en todo tipo de variables, **excepto en variables cualitativas o una variable con intervalo abierto y éste es el intervalo crítico**			
La mediana de una variable X, representada por Md, se define como el valor que divide la distribución de frecuencias de la variable en dos partes iguales, conteniendo cada una el 50% de las observaciones. La mediana es el valor que deja por debajo de sí el 50% de las observaciones.			
Para su cálculo con pocos casos:	Primero, se ordenan las n puntuaciones de menor a mayor		
	Segundo, se observa si *n* (número de observaciones) es par o impar	Si *n* es impar, el valor de la mediana es el de la observación que ocupa la posición central dentro de ese conjunto de observaciones ya ordenadas— Md = $(n+1)/2$	
		Si n es par, la mediana es la media aritmética de los dos valores centrales — Md = $(X_{(n/2)} + X_{(n/2)+1})/2$	
Para casos de frecuencias agrupadas en intervalos o no	Primero, ordenar los valores	Empezando por el valor más bajo de la variable hasta el más alto, porque la mediana se define como el número de veces que se repite cada valor o cualquiera de sus valores inferiores.	
	Segundo, determinar las frecuencias acumuladas	La frecuencia acumulada debe ir hacia el valor más alto, es decir, el valor n (el total de observaciones) debe coincidir con el intervalo de valor más alto.	
	Tercero, determinar el **intervalo crítico**, que es donde se encuentra la mediana	Aquél en el que la frecuencia acumulada *na* es igual o superior a n/2 o la proporción acumulada pa es igual o mayor a 0.5, es decir el 50% de *n*.	
	Cuarto, calcular la mediana	Md = Li + ((n/2 -nd)/nc) x I	
			Li = Límite inferior exacto del intervalo crítico (en una distribución no de intervalos, los límites críticos se obtienen restando 0.5 a cada valor de la variable)
			n = número de observaciones
			nd = frecuencia absoluta acumulada debajo del intervalo crítico
			nc = frecuencia absoluta del intervalo crítico
			I = amplitud del intervalo crítico (en una distribución de no intervalos es = a 1)
	El origen de la fórmula se basa en el método de interpolación—**se asume la distribución homogénea** de las puntuaciones dentro de cada intervalo.		

Karla Marmolejo

Tema 2
Indices de Tendencia Central y Posición

La Moda

Se puede obtener tanto en variables cualitativas como cuantitativas		
La moda es una distribución, representada por Mo, que se define como el valor o categoría de la variable con mayor frecuencia absoluta—la que más se repite		
Distribución unimodal	Un único valor con frecuencia absoluta máxima	
Distribución bimodal	Dos valores con la frecuencia más alta	
Distribución trimodal…	Tres valores con la frecuencia más alta	
Distribución amodal	Todos los valores tienen la misma frecuencia absoluta—no hay una moda	
Cálculo de la Mo para **variables cualitativas**	Determinar la categoría con máxima frecuencia absoluta	
Cálculo de la Mo para variables **cuantitativas no agrupadas en intervalos**		
Cálculo de la Mo para variables **cuantitativas agrupadas en intervalos**	Primero, localizar el intervalo modal	Intervalo con la máxima frecuencia absoluta
	Segundo, calcular la Mo	Mo = punto medio del intervalo (LES + LEI)/2
Principales características de la moda	Es un índice de cálculo sencillo y fácil interpretación	
	De los tres índices de tendencia central estudiados, la moda es el único que se puede aplicar a variables cualitativas	
	Cuando los datos están agrupados en intervalos y existen intervalos abiertos, la moda se puede calcular, **excepto si el intervalo modal coincide con el intervalo abierto**	

Karla Marmolejo

Cómo elegir el índice central que mejor resuma la tendencia central

La primera opción es la **media aritmética**	Pero se desaconseja cuando la distribución es asimétrica con unos pocos valores extremos que pueden distorsionar la representatividad de la media como tendencia central	
	Tampoco tiene sentido en el caso de variables nominales u ordinales—se pueden usar la mediana o la moda	
	También se desaconseja cuando los datos estén agrupados y existan intervalos abiertos en los extremos	
La siguiente opción es la **mediana**	Es más resistente a los valores extremos que generan asimetría en la distribución	
	Dos motivos por los que no se puede obtener la mediana	El nivel de medida de la variable es nominal
		Con datos agrupados en intervalos, la mediana se encuentra en el intervalo abierto
Finalmente la moda	Sólo se puede usar la moda con variables cualitativas	
	No se puede calcular cuando la distribución es amodal—todas las frecuencias absolutas son iguales	
	Tampoco se puede calcular si el intervalo abierto coincide con el intervalo modal	
Cuando la distribución de una variable cuantitativa es simétrica y unimodal—los valores de la media, mediana y moda son el mismo		

Karla Marmolejo

Tema 2
Indices de Tendencia Central y Posición

Indices estadísticos de Posición

Van dirigidos a un sujeto o dato en particular, para determinar qué posición ocupa este sujeto en la distribución con respecto a los demás participantes			
Proporcionan información relativa de un sujeto respecto a su grupo de referencia, dentro de la distribución de frecuencias de la variable			
Indican la situación de una puntuación con respecto a un grupo, utilizando al grupo como marco de referencia			
Primero, debemos dividir la distribución en un número de partes o secciones iguales entre sí en cuanto al número de observaciones			
La aplicación de estos índices requiere al menos de un nivel de medida ordinal en la variable de estudio			
Percentiles (o centiles)	Los 99 valores que dividen en 100 partes iguales la distribución de frecuencias de la variable.		
	El percentil 50, P_{50}, coincide con la mediana de una distribución		
	Denotado por Pk—valor que deja **por debajo de sí** un porcentaje k de observaciones (donde k = 1, 2, 3… 99 de la variable de interés). **No son porcentajes.**		
	En los percentiles se intenta localizar la posición del percentil con una regla de tres simple	$x = (n \times k)/100$ n= número de observaciones y k el percentil que se está calculando	
	Cálculo en datos agrupados en intervalos	Primero, ordenar los valores empezando por el valor más bajo de la variable hasta el más alto (de abajo hacia arriba). La frecuencia acumulada debe ir hacia el valor más alto, es decir, el valor n (el total de observaciones) debe coincidir con el intervalo de valor más alto.	
		Segundo, Determinar el número de casos por debajo del percentil k. $x = (n \times k)/100$	
		Tercero, localizar el intervalo en el que se encuentra el percentil k, denominado **intervalo crítico** —la frecuencia absoluta acumulada n_a es igual o superior a x (el k% de n).	
		Luego se aplica la fórmula $Pk = Li + ((n \times k/100 - n_d)/n_c) \times I$	
			Li = Límite inferior exacto del intervalo crítico (en una distribución no de intervalos, los límites críticos se obtienen restando 0.5 a cada valor de la variable)

Karla Marmolejo

	n = número de observaciones
	nd = frecuencia absoluta acumulada debajo del intervalo crítico
	nc = frecuencia absoluta del intervalo crítico
	k = percentil que estamos calculando
	I = amplitud del intervalo crítico (en una distribución de no intervalos es = a 1)
Cuando (nxk)/100 es exactamente igual a la frecuencia acumulada na del valor o intervalo, el percentil se corresponde directamente con el límite superior exacto del intervalo crítico.	
Cuando se tienen muy pocos datos, no es habitual calcular perceptibles porque tienen muy poca utilidad.	
Para determinar el percentil de un valor determinado de Xi, hay que despejar k. Si se obtiene un valor con decimales, se toma la cantidad entera más próxima (≥ 5 ↑, ≤ 5 ↓). Si el valor Xi es el LES de un intervalo y el LEI del siguiente, se puede elegir cualquiera de los dos intervalos como crítico.	k = (((Pk - Li) x nc) / I + nd) / n) x 100

Cuartiles	Los tres valores que dividen en cuatro partes iguales la distribución de frecuencias de la variable.	Primer cuartil $Q_1 = P_{25}$—deja por debajo el 25% de las observaciones
		Segundo cuartil $Q_2 = P_{50}$ = Md—deja por debajo el 50% de las observaciones.
		Tercer cuartil $Q_3 = P_{75}$—deja por debajo el 75% de las observaciones
	Se usan para construir índices de estudio de variabilidad de una distribución de frecuencias.	

Deciles	Los nueve valores que dividen en diez partes iguales la distribución de frecuencias de la variable.	$D_1 = P_{10,}$ —deja por debajo el 10% de las observaciones
		$D_2 = P_{20,}$ —deja por debajo el 20% de las observaciones
		Etc… $D_3 = P_{30}$, $D_4 = P_{40}$, $D_5 = P_{50}$, $D_6 = P_{60}$, $D_7 = P_{70}$, $D_8 = P_{80}$, $D_9 = P_{90}$

Karla Marmolejo

Tema 3
Medidas de Variabilidad y Forma

Medidas de Variabilidad

La **variabilidad** o **dispersión** hace referencia al grado de variación que hay en un conjunto de puntuaciones.		
Cuanto menor es la variabilidad en una distribución, más homogénea es la muestra. En el caso extremo y poco habitual de máxima homogeneidad, todos los valores de la variable serían iguale entre sí y la media, no habría variabilidad. Cuando existe cierta dispersión en los datos, la muestra es más o menos heterogénea y las puntuaciones difieren entre sí.		
Una distribución con poca variabilidad—las puntuaciones están muy próximas a la media		
Una distribución con mucha variabilidad—las puntuaciones están alejadas o muy alejadas a la media		
Dos tipos de Medidas o Indices	Aquellos que miden el grado en el que las puntuaciones se asemejan o diferencian entre sí	Amplitud total o rango (A_T)
		Amplitud intercuartil (A_{IQ})
	Aquellos en los que la dispersión se mide con respecto a alguna medida de tendencia central, como la media aritmética.	Varianza (S^2_x)
		desviación tipica (S_x)
Coeficiente de variación (CV)	Cuando se trata de comparar la dispersión de dos o más distribuciones, se basa en la relación entre la desviación típica y y la media, expresado en porcentaje	

Amplitud Total

También llamado recorrido de las observaciones o rango
Denotada como A_T, es la distancia que hay en la escala numérica entre los valles que representan la puntuación máxima y la puntuación mínima en un conjunto de puntuaciones.
A_T = Xmax - Xmin
En variables agrupadas en intervalos, la puntuación máxima es el LSE del intervalo máximo y la puntuación mínima es el LIE del intervalo mínimo.
No captura la poca o mucha dispersión que pueda existir entre los restantes valores, que son la gran mayoría de las puntuaciones.
Normalmente se recomienda incluir la amplitud total como información complementaria de otras medidas de dispersión más relevantes.

Amplitud Intercuartil

Se utiliza como índice de dispersión cuando la distribución es asimétrica. Se utiliza junto con la mediana como medida de tendencia central. También llamado **rango intercuartil**. Es la diferencia entre el tercer y primer cuartil.
$A_{IQ} = Q_3 - Q_1 = P_{75} - P_{25}$

Karla Marmolejo

Tema 3
Medidas de Variabilidad y Forma

Desviación Media

$DM = \sum n_i * \lvert X_i - X \rvert \div n$	$\lvert X_i - X \rvert$ implica utilizar sólo los valores absolutos; el valor absoluto es siempre positivo, si es negativo, se hace positivo.

Varianza

Conjunto de n puntuaciones en una variable Xi, se define como el promedio de los cuadrados de las desviaciones de las puntuaciones con respecto a la media.		
$S^2x = (\sum X^2_i / n) - X^2$		
Cuando los datos se presentan en tablas de distribución de frecuencias	$S2x = (\sum n_i X_i^2 / n) - X^2$	n = número total de observaciones Xi = valor i en la variable X o punto medio del intervalo ni = la frecuencia absoluta del valor o del intervalo i.
A partir de frecuencias relativas	$S2x = \sum p_i X_i^2 - X^2$	
Es un número positivo que se expresa en las unidades de la variable al cuadrado		
Cuasivarianza	$S^2_{n-1} = \sum (X_i - X)^2 / (n - 1)$	En la cuasivarianza se están midiendo la distancia entre los números, los espacios entre los valores.

Desviación Típica

La raíz cuadrada positiva de la varianza	
$Sx = \sqrt{S^2x}$	
Se suele utilizar más que la varianza porque se expresa en las mismas unidades que la variable objeto de estudio	
Tanto el cálculo de la varianza como el de la dispersión típica requieren el uso de todas las puntuaciones de la distribución	
Ambas miden la variabilidad de los datos con respecto a la media aritmética, por lo que sólo pueden aplicarse cuando sea apropiado utilizar la media aritmética.	
Ambas son siempre iguales o mayores a cero. Son iguales a cero únicamente cuando todas las puntuaciones son iguales entre sí. Sus valores son mayores a medida que aumenta la variabilidad de las puntuaciones.	
Si a las puntuaciones de la variable X les aplicamos una transformación lineal $Y_i = bX_i + a$; la varianza de las nuevas puntuaciones será $S^2y = b^2S^2x$ y la desviación típica será $Sy = b \cdot Sx$ (es decir, sólo les afecta cuando multiplicamos)	
Cuasidesviación típica	$S_{n-1} = \sqrt{S^2_{n-1}}$

Coeficiente de Variación

$CV = (Sx/X) \times 100$	Se expresa como porcentaje, definido para variables con $X > 0$; se recomienda acompañar su resultado por la media y la desviación típica.
	Entre mayor sea el CV, mayor es el grado de dispersión. Se utiliza para comparar dos variables diferentes.

Karla Marmolejo

Otro aspecto importante a considerar en la distribución de frecuencias es la forma que presente la distribución.

Asimetría

La asimetría de una distribución nos indica el grado en el que las puntuaciones se reparten por debajo y por encima de la medida de tendencia central.		
Dos índices que se utilizan con mayor frecuencia	Indice de asimetría de Pearson $Ap = (X - Mo) / Sx$	Se basa en la relación entre la media y la moda
		Es un índice adimensional (no tiene unidades de medida) que se aplica sólo a distribuciones unimodales.
		En distribuciones con asimetría positiva $Ap > 0$
		En distribuciones con asimetría negativa $Ap < 0$
	Indice de asimetría de Fisher $A_F = \sum (Xi - X)^3 / n \cdot S^3x$ $A_F = \sum ni(Xi - X)^3 / n \cdot S^3x$	Se basa en las distancias de las puntuaciones respecto a su media elevadas al cubo, por lo que su valor puede ser positivo, negativo o cero.
		Tiene en cuenta todas y cada una de las puntuaciones de la muestra, por lo que puede considerarse el mejor índice de asimetría.
		En distribuciones con asimetría positiva $A_F > 0$
		En distribuciones con asimetría negativa $A_F < 0$
		En distribuciones simétricas $A_F = 0$

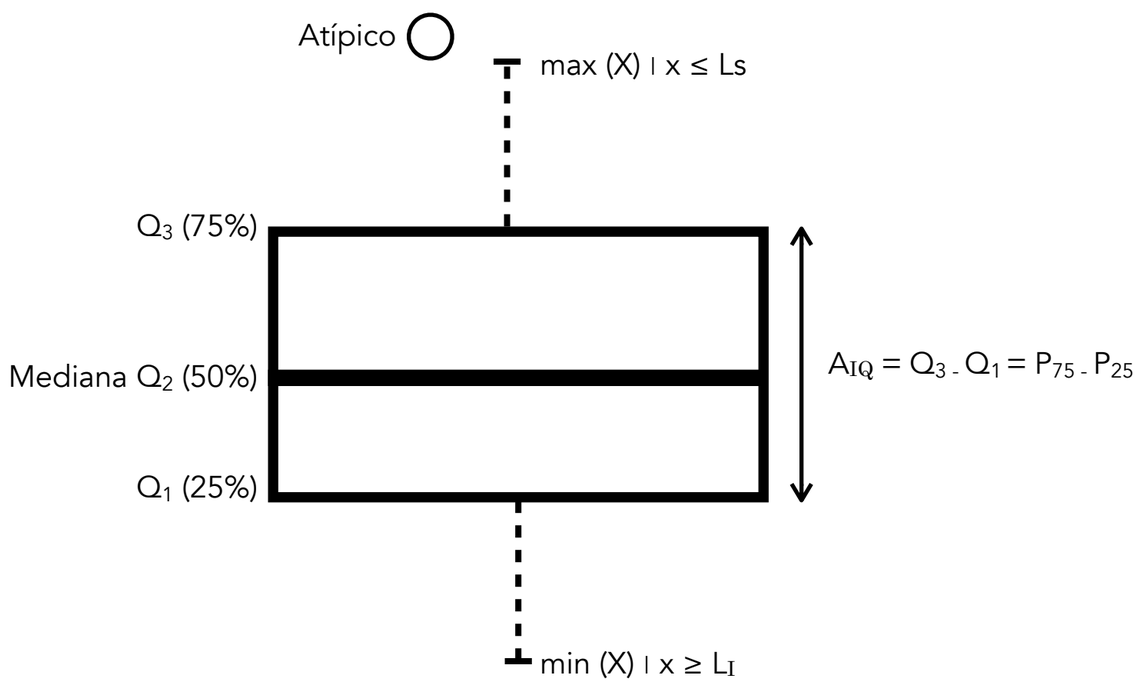
Curtosis

Llamado también **apuntamiento**, se refiere al grado de apuntamiento de los datos en la distribución de la frecuencia.		
Se basa en las distancias de cada puntuación con respecto a la media elevadas a la cuarta potencia.		
Tres formas diferentes	Leptocúrtica	Si la distribución es muy apuntada y mayor que la curva normal. $Cr > 0$
	Platicúrtica	Si la distribución es muy aplastada y menor que la curva normal. $Cr < 0$
	Mesocúrtica	Si muestra un grado de apuntamiento intermedio y tiene un grado de apuntamiento similar al de la curva normal. $Cr = 0$
Distribución de frecuencias	$Cr = (\sum (Xi - X)^4 / n \cdot S^4x) - 3$	
	$Cr = (\sum ni(Xi - X)^4 / n \cdot S^4x) - 3$	

Karla Marmolejo

Diagrama de Caja

También llamado gráfico de caja y bigotes	
Se trata de una presentación visual que resulta útil para estudiar la asimetría de una variable cuantitativa, así como para detectar si hay valores extremos o atípicos (outliers) en la distribución de frecuencias	
Se representa mediante una caja rectangular cuya altura corresponde con la amplitud o rango intercuartil. Dentro de la caja se dibuja una línea para indicar dónde se sitúa la mediana, que coincide con el segundo cuartil Q_2. La caja es atravesada por una línea vertical llamada bigote, en cuyos extremos se sitúan los valores mínimos y máximos de la variable (sin considerar los valores atípicos)	
Límites que determinan si un valor es atípico	Límite superior $L_s = Q_3 + A_{IQ} \times 1{,}5$
	Límite inferior $L_I = Q_1 - A_{IQ} \times 1{,}5$
Cuando existen casos extremos o atípicos, éstos aparecen como un círculo pequeño por encima o por debajo de los bigotes del diagrama de caja.	
Para estudiar la asimetría se va a tener en cuenta la longitud de los bigotes y el número de casos atípicos en ambas colas de la distribución	Si los bigotes tienen la misma longitud y el mismo número de casos atípicos en ambos lados, diremos que es aproximadamente simétrica
	Si los bigotes tienen la misma longitud pero hay más casos atípicos en un extremo en una cola de la distribución, entonces diremos que la distribución presenta asimetría.
	Si los bigotes presentan diferente longitud, estamos ante una distribución asimétrica



Karla Marmolejo

Puntuación diferencial	Las puntuaciones diferenciales aportan más información, nos indican si la puntuación coincide con la media de su grupo, es inferior o superior a ella.
	Su media es cero
	La varianza de las puntuaciones diferenciales es igual a la varianza de las puntuaciones directas
	Dos puntuaciones diferenciales idénticas pueden tener un significado muy diferente en función de la media y de la varianza de las distribuciones de las que proceden.
	$x_i = Xi - X$
Puntuación típica o tipificada	Nos permiten comparar dos sujetos distintos en dos pruebas o variables distintas
	Al proceso de obtener puntuaciones típicas se le llama tipificación—indica el número de desviaciones típicas que se aparta de la media una determinada puntuación
	Su media es cero
	Su varianza es igual a 1
	$z_x = x_i / Sx = (Xi - X) / Sx$
	Las puntuaciones típicas reflejan las relaciones entre las puntuaciones con independencia de la unidad de medida, por eso permiten hacer comparaciones entre distintos grupos y distintas variables.

Karla Marmolejo

Tema 4
Relación entre Variables I

Asociación entre Dos Variables Cualitativas

Tabla de Contingencia	Se dice que hay asociación entre dos variables si existe algún tipo de tendencia o patrón de emparejamiento entre los distintos valores de esas variables. La existencia de asociación entre dos variables indicaría que la distribución de los valores de una de las dos variables difiere en función de los valores de la otra	
	Una tabla de contingencia es una forma de ordenar los datos para estudiar la relación entre variables cualitativas con pocas categorías. Es una distribución de frecuencias clasificada de acuerdo a los valores que pueden tomar las dos variables. Por eso, se sitúan los valores de una de las variables en las filas (X)—los sujetos y los valores de la otra variable en las columnas (Y)—las variables.	
	Cada celdilla representa la frecuencia o número de elementos que reúne a la vez los valores de las dos variables que se cruzan en cada casilla	
	Frecuencias Conjuntas (n_{ij})	Se toma en consideración uno de los valores de las dos variables (donde las variables se "cruzan" o intersectan.
	Frecuencias Marginales	Los totales de cada valor de una única variable (las frecuencias marginales de Y son el total de cada columna, las frecuencias marginales de X son el total de cada fila).
	Es muy frecuente tener más de dos categorías en alguna de las variables. El formato general de una tabla de contingencia es el mismo, añadiendo filas y/o columnas, y calculándose las distintas frecuencias de la forma indicada	
	De manera habitual, las tablas de contingencia se presentan con las frecuencias absolutas, además de información de porcentajes en cada celdilla	
	Porcentaje Total (P_{IJ})	Es el número de casos década celdilla dividido por el total de casos n y multiplicado por 100
	Porcentaje Condicionado a X_i (P_I)	Porcentaje por fila—el número de casos de cada celdilla dividido por el total de casos por fila (frecuencia marginal Xi) y multiplicado por 100. El conjunto de estos valores se denomina **distribución condicional de filas** y sus porcentajes suman 100.
	Porcentaje Condicionado a Y_j (P_J)	Porcentaje por columna—el número de casos de cada celdilla dividido por el total de casos por columna y multiplicado por 100. El conjunto de estos valores se denomina **distribución condicional de columnas** y sus porcentajes suman 100.
	Al considerar los porcentajes para interpretar la relación entre las variables hay que tener en cuenta si la relación entre ellas es simétrica o asimétrica	**Asimétrica**—una de las dos variables se considera como **factor explicativo** de la distribución de la otra variable. Los porcentajes se calculan en el sentido de esta **variable explicativa**.
		Simétrica—no existe esta distinción, no cabe esperar que una de ellas influya en la otra.
	En una relación asimétrica los porcentajes se calculan en el sentido de la **variable explicativa**, por lo que la suma de los porcentajes de cada categoría de la variable explicativa referidos al total marginal de esa categoría será el 100%.	
	Cada uno de los tres tipos de porcentajes pone el énfasis en una distribución diferente y ofrece comparaciones distintas, según el sentido de la predicción. La utilización de porcentajes permite eliminar la influencia del tamaño de la muestra y del tamaño de los marginales, por lo que se pueden realizar comparaciones entre valores de las distribuciones condicionadas, y esta comparación indica la existencia de relación o no entre las variables, así como la naturaleza de la relación.	

Karla Marmolejo

	Distribución Marginal	Distribución de frecuencias unidimensional (marginal) que nos informan del número de observaciones para cada valor de una de las variables, prescindiendo de la información sobre los valores de las demás variables. Hay una distribución marginal de la variable X (contiene todas las frecuencias marginales de X) y una distribución marginal de la variable Y (contiene todas las frecuencias marginales de Y)
	Distribución Condicionada	Distribución que especifica las observaciones que hay de cada valor de una de las variables al imponer la condición de que la otra tome un valor determinado. Hay una **distribución de Y condicionada a un valor de Xi,** que se considera únicamente una fila de la tabla de contingencia, y una **distribución de X condicionada a un valor de Yj**, que únicamente tiene en cuenta una columna de la tabla.

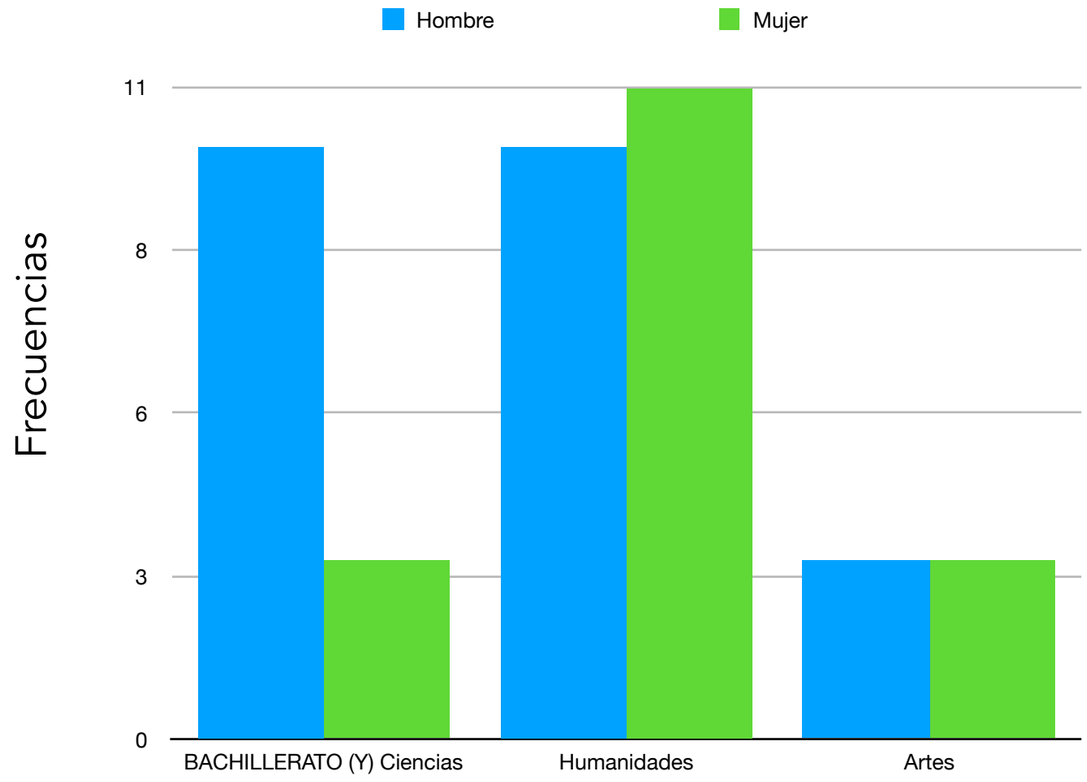
Karla Marmolejo

Diagrama de Barras Conjunto

Es apropiado cuando al menos una de las dos variables es cualitativa.	
Se construye sobre los datos de la tabla de contingencia, situando una de las dos variables en el eje horizontal y para identificar la otra variable se utilizan barras de distinto color o trama (especificado en la leyenda). Básicamente hay dos formas de representar un diagrama de barras conjunto:	
Diagrama de barras adosadas	En este diagrama se muestra con barras colocadas horizontalmente o verticalmente la frecuencia de cada casilla del interior de la tabla de contingencia. Es habitual mostrar el diagrama de barras en porcentajes del total o condicionales (por fila o por columna). Se suele situar la variable con más valores en aleje de abscisas, reservando la variable con menos valores para el color de las barras.
diagrama de barras apiladas	Este gráfico muestra una barra por cada valor que toma la variable Y, las cuales a su vez, se dividen en distintos colores que representa a cada valor de la variable X. Indica la frecuencia con la que aparece cada valor de X en cada valor de Y, comparando entre categorías, la aportación de cada valor al total. **Esta es la representación más adecuada para visualizar porcentajes condicionados.**

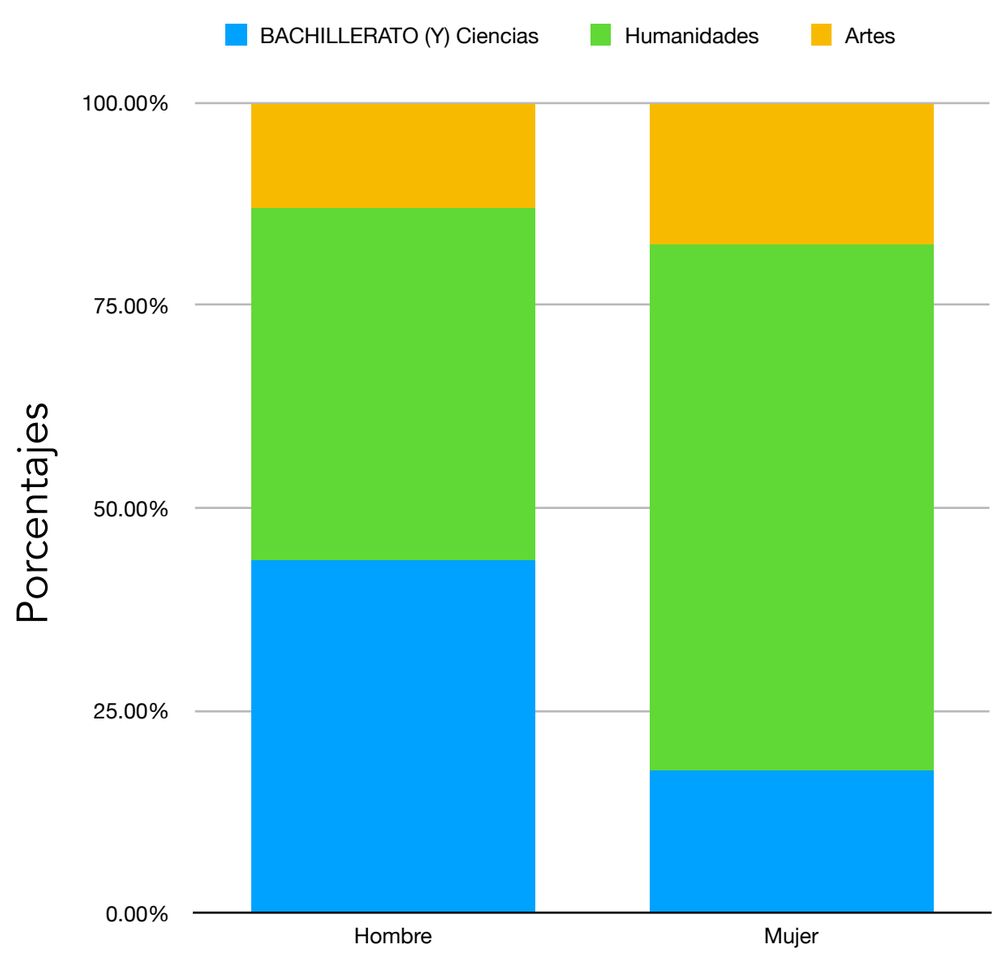
	BACHILLERATO (Y)			
	Ciencias	**Humanidades**	**Artes**	
Hombre	10	10	3	23
Mujer	3	11	3	17
	13	21	6	40

Barras Adosadas



	BACHILLERATO (Y)		
	Ciencias	**Humanidades**	**Artes**
Hombre	43.48%	43.48%	13.04%
Mujer	17.65%	64.70%	17.65%

Barras Apiladas



Karla Marmolejo

Medidas Globales de Asociación para Variables Cualitativas

Independencia χ^2

La existencia de una asociación entre dos variables indica que la distribución de las frecuencias de los valores de una de las dos variables difiere en función de los valores de la otra. Por el contrario, se habla de independencia entre variables cuando no existe tal patrón de relación entre los dos valores de las mismas.

Para saber si existe o no independencia entre dos variables se utiliza el estadístico χ^2, que se basa en la comparación de las frecuencias conjuntas. Así, se comparan las frecuencias empíricas (u observadas) con las frecuencias teóricas (o esperadas) suponiendo que no hubiera asociación—que fueran independientes.

$$\chi^2 = \sum\sum ((n_e - n_t)^2 / n_t)$$

(el sumatorio engloba toda la fracción)

n_e es la frecuencia empírica (observada)

n_t es la frecuencia teórica (esperada).

n_t = (Total fila x Total columna) / n La suma de la diferencia entre las teorías y empíricas siempre da 0

El índice χ^2 toma el valor 0 cuando dos variables son independientes, siendo mayor que 0 cuando existe asociación entre ellas, tanto mayor, más intensa sea esa relación. No tiene un límite máximo, lo cual supone una dificultad a nivel interpretativo.

Otro inconveniente de este índice es que al multiplicar las frecuencias de todas las casillas por una constante, el valor de χ^2 aumenta, a pesar de que las proporciones de todas las casillas sean las mismas antes y después de dicha multiplicación. Esto hace que su valor sólo pueda compararse para variables en tablas de contingencia del mismo tamaño (I x J) y con el mismo n.

Características

- Adopta valores entre 0 y $+\infty$
- Unicamente adopta el valor 0 si la frecuencia empírica de la celdilla es igual a la teórica que le corresponde en todas las celdillas de la tabla de contingencia
- El tamaño de la muestra, n, debe ser relativamente grande. El criterio que se utiliza habitualmente es que la frecuencia esperada mínima por casilla sea al menos de 5 en aproximadamente el 80% de las casillas, considerando además que la frecuencia mínima esperada en cada casilla sea de 1.
- Sirve para valorar la existencia o no de independencia, pero no resulta apropiada para medir la intensidad , pues el tamaño de la muestra y el número de categorías de las variables influyen sobre los valores de este estadístico.

Karla Marmolejo

Medidas Globales de Asociación para Variables Cualitativas

Coeficiente C de Contingencia

El coeficiente de Contingencia es una medida de asociación derivada de X^2 que es aplicable a tablas de contingencia de cualquier dimensionalidad (con independencia del número de filas y columnas).	$C = \sqrt{(X^2 / (X^2 + n))}$
El coeficiente de contingencia *C* puede asumir valores mayores o iguales a 0 y menores que 1. Cuanto mayor es el valor de *C*, mayor es la relación entre las dos variables, mientras que valores cercanos a 0 indican ausencia de relación entre las variables.	
Este coeficiente es especialmente útil cuando el número de filas y de columnas de la tabla de contingencia coinciden porque se puede precisar más su valor máximo, lo que permite una interpretación mejor con la siguiente fórmula. Sólo se puede calcular cuando el número de columnas y filas es el mismo.	$C = \sqrt{(k - 1) / k)}$ k = # filas = # columnas

Coeficiente V de Cramer

El coeficiente *V* de Cramer es una modificación de X^2 que alcanza un valor máximo de 1 en caso de máxima asociación o asociación perfecta y un valor mínimo de 0 en una situación de independencia perfecta.	$V = \sqrt{(X^2 / n (m - 1))}$ - m = valor más pequeño entre #filas y columnas
La experiencia muestra que con este estadístico es poco frecuente encontrar valores próximos a 1, de hecho pocas veces se alcanza un valor de 0,6. En términos empíricos, por tanto, se puede considerar al 0,6 prácticamente como un valor máximo habitual, por lo que un valor de 0,3 antes de considerarlo como bajo por su proximidad al 0, conviene interpretarlo más bien como un valor intermedio.	

Coeficiente φ

El coeficiente **φ** (phi) es una medida de asociación derivada de X^2 que se aplica a variables dicotómicas (por tanto con tablas de contingencia 2x2), excluyentes entre sí y binarias (a cada categoría le corresponde 1 ó 0)	$\varphi = \dfrac{(n_{11} \times n_{22} - n_{12} \times n_{21})}{\sqrt{((n_{11}+n_{12}) \times (n_{21} + n_{22}) \times (n_{11} + n_{21}) \times (n_{12} + n_{22})}}$
Con esta fórmula, **φ** puede adoptar valores entre -1 y 1	Un valor próximo a 0 indica ausencia de relación entre las variables y cuanto más se aproxime a 1 ó -1, más intensa será su relación. Si es positivo, los sujetos que adopten el valor "0" de la primera variable, también adoptarán "0" para la segunda, y los que adopen "1" para la primera también lo harán para la segunda. Si es negativo, es a la inversa—los sujetos que adopten el valor "0" en la primera variable, tienden a adoptar el valor "1" en la segunda y viceversa.

Karla Marmolejo

Habitualmente, cuando se estudia la relación entre una variable cualitativa y una variable ordinal se utilizan las mismas estrategias que en el estudio de dos variables cualitativas (χ^2, C, V y φ). En el caso del estudio de dos variables ordinales, la estrategia dependerá del número de valores distintos que puedan adoptar esas variables.	Si ambas variables adoptan un número reducido de valores, se suelen utilizar tablas de contingencia para su estudio, similar a las variables cualitativas. Cuando interesa estudiar la fuerza de la relación, teniendo encuentra el carácter ordinal, se utilizan como la *d* de Sommers o el coeficiente Gamma.
	Si alguna de las dos variables (o ambas) adoptan un número amplio de valores, las tablas de contingencia dejan de ser prácticas por el elevado número de filas y columnas. En estos casos se utiliza el coeficiente de correlación de Spearman (r_s) o el coeficiente *tau-b* de Kendall.

Karla Marmolejo

Relación entre Variables Ordinales

Cuando se estudia la relación entre una variable cualitativa y una variable ordinal, se utilizan las mismas estrategias que en el estudio de dos variables cualitativas. En el caso del estudio de dos variables ordinales, la estrategia dependerá del número de valores distintos que puedan adoptar esas variables.	Si ambas variables adoptan un número reducido de valores, se suelen utilizar tablas de contingencia para su estudio, de manera similar a lo visto en variables cualitativas. Cuando interesa estudiar la fuerza de la asociación, teniendo en cuenta el carácter ordinal de las variables, en lugar de los índices globales vistos, se utilizan otros desarrollados específicamente para este caso, como la *d* de Sommers, o el coeficiente Gamma.
	Si alguna de las dos variables (o ambas) adoptan un número amplio de valores, el estudio en tablas de contingencia deja de ser práctico, debido al elevado número de filas y columnas de las tablas. En estos casos se suele utilizar el coeficiente de correlación de Spearman o el coeficiente tau-b de Kendall.

Coeficiente de Correlación por rangos de Spearman

El coeficiente de correlación de Spearman se basa en los rangos de los datos en lugar de hacerlo en los valores reales. Resulta apropiado en el caso de variables ordinales o de variables cuantitativos que no tengan una distribución normal.

Para calcular este coeficiente, primero hay que ordenar todos los casos para cada una de las variables de interés y asignar un rango consecutivo a cada observación de cada una de las variables por separado. Esto es, primero hay que ordenar todos los valores de menor a mayor y luego asignar los rangos comenzando por el 1 al menor valor.	$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$	d_i = Rango (Xi) - Rango (Yi) n = nº. de sujetos
		Los valores del coeficiente de Spearman de -1 a +1. El signo del coeficiente indica la dirección de la relación y el valor absoluto del coeficiente de correlación indica la fuerza de la relación entre las variables.

Rangos empatados—cuando se producen puntuaciones que son iguales; en estos casos se asigna a las puntuaciones el rango promedio que ocuparían las observaciones empatadas. Cuando hay empates se recomienda utilizar el **coeficiente de correlación de Pearson** aplicado a los rangos.

Si la asociación entre ambas fuera perfecta, esperaríamos que el rango que corresponde a cada caso de la variable X fuera exactamente igual al rango de la variable Y, por lo tanto el coeficiente se calcula en base a las diferencias registradas en los rangos entre ambas variables, esperando que estas diferencias fueran 0.

Conforme mayores son las diferencias observadas en las ordenaciones de ambas variables, más se aleja la relación de ser perfecta	Si $r_s > 0$ —existe una relación directa entre las variables. Cuanto más se acerque a 1 el coeficiente de correlación, más fuerte será la relación existente entre las variables.
	Si $r_s < 0$ —existe una relación inversa entre las variables de forma que los valores de X se corresponderán con valores bajos de Y y viceversa. Cuanto más se acerque a -1 el coeficiente de correlación, más fuerte será la relación existente entre las variables
	Si $r_s \approx 0$ —apenas hay relación entre las variables, una variable poco o nada tiene qué ver con la otra.
	Sirve para valorar la existencia o no de independencia, pero no resulta apropiada para medir la intensidad , pues el tamaño de la muestra y el número de categorías de las variables influyen sobre los valores de este estadístico.

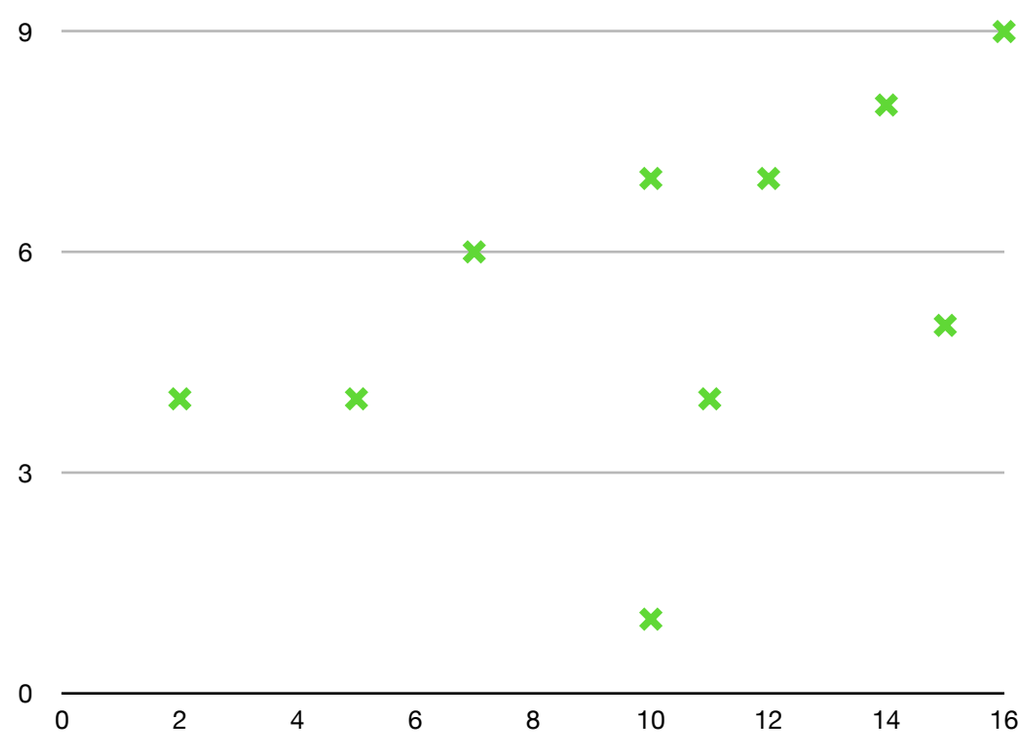
Karla Marmolejo

Tema 5
Relación entre Variables II

Representación Gráfica

Diagrama de Dispersión	El diagrama de dispersión, también denominado **nube de puntos,** se utiliza en el caso de dos variables cuantitativas, ofreciendo una primera aproximación de la relación que existe entre ambas variables.	
	Para utilizar el diagrama de dispersión se sitúa una de las variables en el eje de las abscisas (X) y la otra en el de las ordenadas (Y). Para cada par de datos, se localiza la intersección de ambas variables y se marca con un punto.	
	Relación Lineal Directa	Cuando los valores altos en Y tienden a emparejarse con valores altos en X, los valores intermedios de Y se emparejan con los valores intermedios de X y los valores bajos en Y tienden a emparejarse con los valores bajos de X.
	Relación Lineal Inversa	Cuando los valores altos en Y tienden a emparejarse con valores bajos en X, los valores intermedios de Y se emparejan con los valores intermedios de X y los valores bajos en Y tienden a emparejarse con los valores altos de X.
	Relación Lineal Nula	Cuando no hay un emparejamiento sistemático entre ellas en función de sus valores.
	Otra opción es que exista una relación entre las variables, pero no se trata de una relación lineal.	

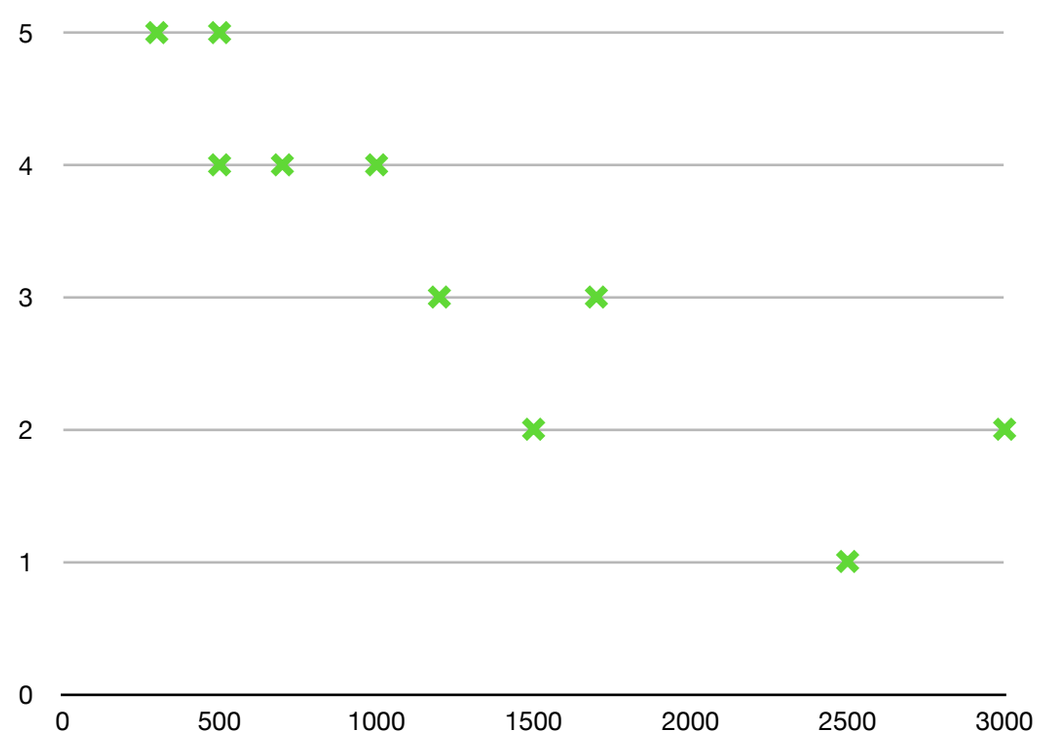
ID	Horas estudio	Calificación PAU
1	7	6
2	11	4
3	16	9
4	5	4
5	14	8
6	10	7
7	12	7
8	10	1
9	2	4
10	15	5



Relación Lineal Directa

Relación Lineal Inversa

Paciente	Calorías	Grado Anorexia
1	500	4
2	300	5
3	1000	4
4	700	4
5	2500	1
6	1500	2
7	500	5
8	1200	3
9	1700	3
10	3000	2



Karla Marmolejo

Covarianza

Es un índice que detecta la relación lineal entre X e Y. Hace referencia a la variación conjunta de dos variables. Presenta una grave limitación, al igual que χ^2, y es que se desconocen los valores mínimo y máximo que puede adoptar, lo que merma su capacidad para interpretar el grado de relación entre las variables.	
$S_{XY} = Cov(X,Y) = (\sum X_i Y_i / n) - \ddot{X} * \bar{Y}$	X_i = valor de la variable X en el caso i
	Y_i = valor de la variable Y en el caso i
	$\ddot{X}$ = media de la variable X
	$\bar{Y}$ = media de la variable Y
	n = número de casos de la muestra
Valor Positivo	La relación es directa
Valor Negativo	La relación es inversa
Valor en torno a 0	La relación es nula

Karla Marmolejo

Coeficiente de Correlación Lineal de Pearson

Es un índice que detecta la relación lineal entre X e Y, y lo hace superando los límites de interpretación de la covarianza, al tener establecido un valor máximo de 1 y un mínimo de -1. Sólo es apropiado para el estudio de las relaciones lineales entre variables..		
$r_{XY} = \frac{Cov(X,Y)}{S_X S_Y} = \frac{S_{XY}}{S_X S_Y}$		S_X = desviación típica de la variable X
		S_Y = desviación típica de la variable X
$r_{XY} = \frac{n\sum(XY) - \sum X \sum Y}{\sqrt{(n\sum X^2 - (\sum X)^2}\ \sqrt{(n\sum Y^2 - (\sum Y)^2}}$		S_{XY} = covarianza entre X e Y
Interpretación y Características	Cuanto mayor es el valor absoluto, la relación lineal entre las dos variables es más fuerte	
	Cuando el signo es positivo—a valores mayores de la variable X tienden a corresponder, en media, valores mayores de la variable Y—**relación lineal directa**	
	Cuando el signo es negativo—a valores mayores de la variable X tienden a corresponder, en media, valores menores de la variable Y—**relación lineal inversa**	
	El valor oscila siempre entre -1 y 1. Los valores 1 y -1 indican una correlación lineal perfecta.	
	Un coeficiente de correlación lineal de valor cercano a cero—no existe relación lineal, pero no excluye la posibilidad de que las variables tengan otras relaciones entre sí de carácter no lineal, lo cual indica una limitación importante y es que sólo detecta relaciones lineales entre dos variables.	
	Es un índice simétrico, por lo que $r_{XY} = r_{YX}$	
	El valor absoluto no se ve afectado por transformaciones lineales de las variables. $r_{XY} = r_{XY2}$ además $r_{YY2} = \pm 1$	
	La correlación entre dos variables, por alta que sea, no implica que X sea la cause de Y ni que Y sea la causa de X. Para poder hablar de causalidad se tiene que cumplir unos requisitos relativos al diseño de la investigación, que debe ser **experimental.**	
	Puede verse afectado por terceras variables.	
Casos Particulares	Relación entre variables ordinales	La formula de la correlación lineal de Spearman se deriva matemáticamente del coeficiente de correlación lineal de Pearson aplicado a rangos, por lo que su resultado es idéntico.
		El único caso en el que no coinciden es en el de empates en los rangos, en cuyo caso hay que utilizar el de Pearson entre los **RANGOS**.
		Para variables ordinales, X e Y son los RANGOS, no el valor de X e Y original, aplicados a la fórmula
	Relación entre variables dicotómicas	La formula del coeficiente **φ** se deriva del coeficiente de correlación lineal de Pearson, por lo que su resultado es igual.
		El cálculo de **φ** se basa en la tabla de contingencia, por lo que es bastante más rápido de calcular que r_{XY}
		Dado que para calcular **φ** se utiliza únicamente con puntuaciones de 0 y 1, los valores de la variable X pasarán a codificarse como 0 y los de Y pasarán a codificarse como 1.
		Se utiliza el **coeficiente de correlación biserial puntual**, el cual coincide con el de correlación lineal de Pearson.
		$r_{bp} = (\ddot{X}_p - \ddot{X}_q / S_x) * \sqrt{(p \cdot q)}$
		$\ddot{X}_p$ es la media de las puntuaciones de la variable cuantitativa X obtenidas por el grupo al que se le ha asignado un 1 en la variable dicotómica
		$\ddot{X}_q$ es la media de las puntuaciones de la variable cuantitativa X obtenidas por el grupo al que se le ha asignado un 0 en la variable dicotómica
		S_x = desviación típica de la variable **cuantitativa** X

Karla Marmolejo

	Relación de una variable dicotómica—sólo puede adoptar dos valores posibles, que se suelen representar por 0 y 1— y otra cuantitativa	p es la proporción de casos del grupo al que se le ha asignado un 1 en la variable dicotómica—**No. 1s / n**
		q es la proporción de casos del grupo al que se le ha asignado un 0 en la variable dicotómica—**No. 0s / n**
		El rango en valor absoluto que se puede obtener con el coeficiente de correlación biserial puntual está entre 0 y 1. Es 0 cuando las medidas de la variable cuantitativa para cada categoría de la variable cualitativa son iguales. Entre más cercano a 1, más alta será la relación. El signo dependerá del orden de las medias en el numerador.
		Si la variable en lugar de ser dicotómica es **dicotomizada** (una variable se ha dicotomizado si hay un continuo latente entre las dos categorías pero se han establecido dos únicos valores; i.e. el peso se puede considerar en gramos o dicotomizar en alto y bajo), se debe utilizar el **coeficiente de correlación biserial.**

Karla Marmolejo

Coeficientes de correlación apropiados en función del tipo de variable

		Variable Y			
		Cualitativa dicotómica	**Cualitativa politómica**	**Ordinal**	**Cuantitativa**
Variable X	**Cualitativa dicotómica**	**Estadístico χ^2**	**Estadístico χ^2**	**Estadístico χ^2**	**Correlación Biserial Puntual**
		Coeficiente C de Contingencia	**Coeficiente C de Contingencia**	**Coeficiente C de Contingencia**	Correlación Biserial
		Coeficiente V de Cramer	**Coeficiente V de Cramer**	**Coeficiente V de Cramer**	
		Coeficiente φ			
		Coeficiente Q de Yule			
		Coeficiente Kappa			
	Cualitativa politómica		**Estadístico χ^2**	**Estadístico χ^2**	
			Coeficiente C de Contingencia	**Coeficiente C de Contingencia**	
			Coeficiente V de Cramer	**Coeficiente V de Cramer**	
	Ordinal			**Coeficiente de correlación de Spearman**	**Coeficiente de correlación de Spearman**
				Coeficiente de correlación de Kendall	Coeficiente de correlación de Kendall
				Coeficiente de correlación de Goodman y Kruskal	Coeficiente de correlación de Goodman y Kruskal
	Cuantitativa				**Covarianza**
					Coeficiente de correlación de Pearson

Karla Marmolejo

Sir Francias Galton en sus estudios encontró un fenómeno que definió como una **regresión a la media de la distribución** (padres altos tenían hijos más altos que la media pero más bajos que ellos y padres bajos tenían hijos más bajos que la media pero más altos que ellos). Buscó una ecuación matemática que sirviera para estimar los valores que adoptarían en una variable sujetos para los que se conoce sus puntuaciones en otra variable y la relación entre ambas. Esta ecuación resultaría ser la ecuación de una recta.	
Si dos variables X e Y se relacionan linealmente, entonces la representación gráfica de su distribución conjunta se aproximará visualmente bastante a una línea recta, y por consiguiente, podemos escribir una variable en función de la otra con esta ecuación.	
Correlación y regresión son dos conceptos muy cercanos. La diferencia fundamental entre ambos estiba en su objetivo	En la regresión, el interés se centra en predecir los valores de una variable (Y) a partir de los valores conocidos de la otra variable (X), además de que en la regresión la relación entre las variables es asimétrica o direccional.
	En la correlación, se busca la variación conjunta de las dos variables, y la relación entre las variables es simétrica
Tres fases	1. La identificación del modelo de regresión, que supone obtener los coeficientes de regresión que le caracterizan.
	2. La valoración del modelo, que supone el estudio de la capacidad predictiva del mismo
	3. La aplicación del modelo para predecir variables

$y_i' = a + bX_i$	La regresión se suele utilizar en situaciones en las que se dispone de la medida de dos variables X e Y en una muestra de participantes y, después, para otros sujetos de esa misma población, se predice cuáles serán los valores de Y, desconocidos en ese momento, en función de los valores de X, que sí son conocidos.
	La recta de regresión, desde el unto de vista estadístico, tiene una característica fundamental—se trata de la recta que ajusta a la nube de puntos del diagrama de dispersión con menos error.
	a es una constante que se denomina **origen**, porque corresponde al valor que adopta la variable Y cuando la variable X vale cero. Por tanto, indica el origen, el punto en el que la recta corta al eje de ordenadas.
	b es una constante que se denomina **pendiente**, porque de ella depende la inclinación de la recta. Indica en qué medida cambian los valores de Y por cada incremento de una unidad en los valores de X
	Y es la variable cuyo valor se desea conocer y va a ser pronosticado a partir del valor de la variable X. Se suele denominar **variable pronosticada** o **criterio** y denotarla como Y'
	X es la variable cuyo valor se conoce y va a ser utilizado para pronosticar el valor del criterio. Se suele denominar **variable predicadora** o simplemente **predictor.**

Karla Marmolejo

Modelo de Predicción o de Regresión Lineal Simple

Cálculo de los coeficientes de regresión a y b	$b = (n \sum(XY) - \sum X \sum Y) / (n \sum X^2 - (\sum X)^2)$	
	$b = r_{xy} \cdot (S_Y/S_X)$	
	$a = \bar{Y} - b\ddot{X}$	
Dado que se trata de una predicción, hay cierto nivel de error.	$y_i' = a + bX_i + E_i$	
	$E_i = Y_i - y_i'$	
E_i es una medida del error individual cometido para cada una de las observaciones, un error lo más pequeño posible para todos los sujetos. Para hacer esto, la regresión lineal se vale del denominado **criterio de mínimos cuadrados**, que es un procedimiento que proporciona valores tales que la suma de los errores al cuadrado (SCE) para los n participantes sea mínimo. Se establece obteniendo los valores *a* y *b* que minimizan la siguiente expresión:	$SCE = \sum(Y_i - y_i')^2 = \sum [Y_i - (a + bX_i)]^2$	
Valoración del Modelo ¿Hasta qué punto el modelo de regresión es un buen modelo	Cuanta mayor relación haya entre las variables, mejor será el pronóstico realizado. Antes de realizar cálculos numéricos, lo ideal para valorar el ajuste es representar los datos mediante un diagrama de dispersión, para tener una primera aproximación de la posible relación entre las dos variables. En ocasiones a simple vista se puede apreciar que utilizar el modelo lineal para predecir los datos de Y puede ser una mala opción.	
	Para examinar la utilidad predictiva de un modelo de regresión, además de la aproximación gráfica, se pueden utilizar dos índices (ayudan a determinar qué tan preciso es el modelo):	
	Varianza error	Es la varianza de los errores cometidos al pronosticar la variable Y a partir de la variable X_i, definiendo estos errores (o residuos) como la diferencia entre la puntuación que realmente obtendría el sujeto en esa variable (Y) y la puntuación que se le ha pronosticado con el modelo de regresión a partir de su valor en la variable X (Y'). Conocida también como **error cuartico medio.**
		$S_E^2 = S^2_{y \cdot x} = \sum E_i^2 / n - \bar{E}^2$
		Cuanto menor sea el valor de la varianza error, más similares serán las puntuaciones pronosticadas por el modelo y las puntuaciones que realmente obtendrían los sujetos en el criterio—**mejores serán las predicciones realizadas por el modelo de regresión**.

Karla Marmolejo

para predecir la variable criterio?	**Coeficiente de Determinación**	En regresión lineal simple, éste es igual al coeficiente de correlación de Pearson elevado al cuadrado— r^2xy
		Indica la proporción de varianza de la variable pronosticada o criterio (Y) que es explicada por el modelo lineal, por la variable predicadora X. El porcentaje de Y que es explicado por la variable X.
		No depende de las unidades en que se expresen los datos y toma valores entre 0 y 1. Cuanto mayor sea su valor, más similares serán las puntuaciones pronosticadas por el modelo y las puntuaciones que realmente obtendrían los sujetos en el criterio—mejores serán las predicciones realizadas por el modelo de regresión. Si es igual a 0, significa que la variable predicadora tiene nula capacidad predictiva de la variable Y. Si llegara a ser 1, la variable predicadora explicaría toda la variación de Y y las predicciones no tendrían error.
		Al multiplicarse por 100, indica el porcentaje de la variabilidad con que se puede predecir el criterio explicado por la variable predictora. A mayor porcentaje, más exacto el modelo.
Características del Modelo	La pendiente de la recta de regresión siempre será del mismo signo que el coeficiente de correlación lineal de Pearson, por lo que informará sobre el tipo de relación lineal entre las variables—directa o inversa. Dado que las desviaciones típicas siempre son positivas, *b* adopta el signo del coeficiente de correlación lineal de Pearson. $b = rxy \cdot (SY/SX)$	
	La media de los errores de predicción o residuos es 0. $\bar{E} = 0$.	
	La media de las puntuaciones pronosticadas coincide con la media de las verdaderas puntuaciones en Y. $\bar{Y}' = \bar{Y}$	
	La varianza de las puntuaciones en Y, es igual a la suma de la varianza de los pronósticos (hechos mediante la recta de regresión) más la varianza de los errores. $S_Y^2 = S_{Y'}^2 + S_{y \cdot x}^2$	
	El coeficiente de determinación es igual al cociente entre la varianza de las puntuaciones pronosticadas y la varianza de las puntuaciones en Y. De ahí que sea un indicador de la proporción de varianza del criterio que queda explicada con el modelo de regresión lineal. $r_{xy}^2 = S_{Y'}^2/S_Y^2$	
	El complementario del coeficiente de determinación es igual al cociente entre la varianza de los errores y la varianza de las puntuaciones en Y, e indica la proporción de la varianza del criterio que NO queda explicada por el modelo de regresión lineal. $1- r_{xy}^2 = S_{y \cdot x}^2 / S_Y^2$	

Karla Marmolejo

Regresión Múltiple

Si se utiliza más de una variable predictora, la capacidad predictiva del modelo puede mejorar. Los modelos de regresión lineales múltiples suelen ser más realistas que los simples, ya que es raro encontrar criterios que se puedan predecir a partir de una única variable predictora. En la actualidad este tipo de análisis se suele realizar mediante software estadístico, aquí no se expondrán las fórmulas para el cálculo de los coeficientes de regresión múltiple.
El coeficiente de determinación varía y se denota como $R^2_{Y \cdot X_1X_2}$
$R^2_{Y \cdot X_1X_2} = (r^2_{YX_1} + r^2_{YX_2} - 2 \cdot r_{YX_1} \cdot r_{YX_2} \cdot r_{X_1X_2}) / (1 - r^2_{X_1X_2})$
Para faciliar la utilización de esta fórmula, se presentan los datos en una matriz de correlaciones—una tabla con el mismo número de filas y columnas que de variables, en la que en cada casilla aparece la correlación entre las variables correspondientes a la fila y la columna
El valor predictivo de este modelo de regresión múltiple es muy superior al de ambos modelos de regresión simple. Teniendo en cuenta ambas predictoras simultáneamente se explica la variabilidad de la variable Y a partir de las variables X_1 y X_2, (en porcentaje si se multiplica por 100)

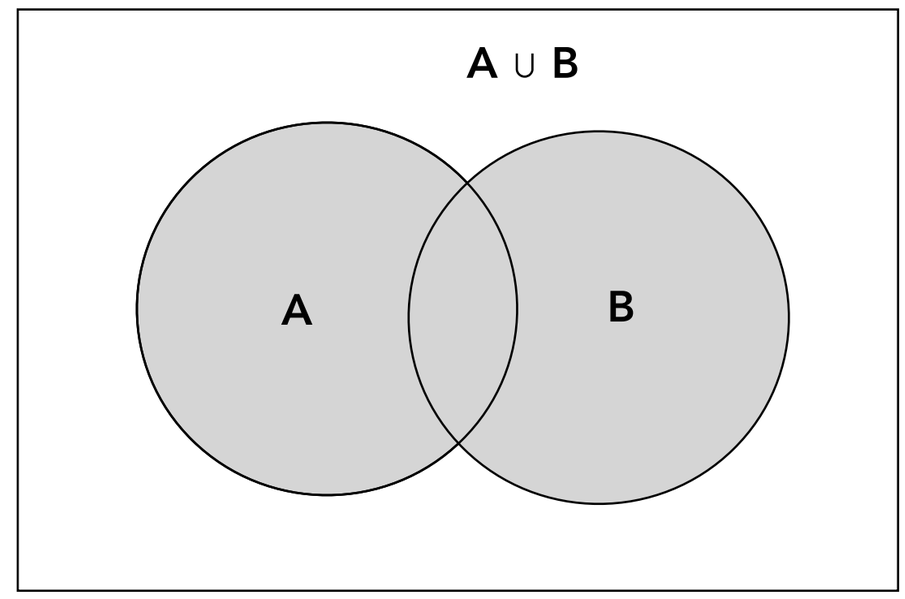
Karla Marmolejo

Tema 6
Nociones básicas de probabilidad

Probabilidad	Teoría matemática que permite desarrollar modelos matemáticos adaptados al estudio de este tipo de situaciones, mediante la asignación de probabilidades (certidumbre) de dichas situaciones. Su objeto es el estudio de variables aleatorias, que son valores que dependen básicamente del azar o de la posibilidad de que puedan o no ocurrir.	
Estadistica	Otra rama de las matemáticas cuyo objeto de estudio son los datos, entendidos como valores o atributos que tienen los objetos de estudio de interés (personas, por ejemplo).	
	Inferencial	Conjunto de métodos y técnicas que permiten inducir, a partir de la información empírica proporcionada por una muestra, cuál es el comportamiento de una determinada población, con un riesgo de error medible en términos de probabilidad.
Probabilidad y Estadistica	Son disciplinas que se refieren al estudio de un mismo tipo de situaciones con incertidumbre. La Probabilidad aporta los modelos matemáticos (las distribuciones) para el estudio de la incertidumbre y la Estadística adapta estos modelos a los datos reales (datos con incertidumbre).	
Experimento Aleatorio	Proceso que se puede repetir indefinidamente en las mismas condiciones y cuyo resultado no se puede predecir con certeza.	
	Tres caracteristicas	Todos los resultados posibles son conocidos con anterioridad a su realización➜tiene un conjunto definido de resultados posibles
		No se puede predecir con certeza el resultado concreto del experimento, pudiéndose obtener cualquiera de los resultados posibles en función del azar.
		El experimento puede repetirse teóricamente un número infinito de veces en idénticas condiciones.
Espacio Muestral	El conjunto de todos los resultados posibles de un experimento aleatorio, también llamado **espacio de resultados**. Se representa por la letra *E*.	
Sucesos	Los resultados de un experimento aleatorio o subconjuntos del espacio muestras (se representan por letras mayúsculas)	
	Elemental, simple o punto muestral	Cada uno de los resultados posibles del experimento aleatorio➜consta de un solo elemento del espacio muestral *E*.
	Compuesto	Consta de dos o más sucesos elementales (o resultados).
	Seguro	El que ocurre siempre➜consta de todos los sucesos elementales (posibles resultados) del espacio muestral y se identifica con el espacio muestal total *E*.
	Posible	Aquel que puede contener algún elemento del espacio muestral *E*➜tiene posibilidad de que ocurra.
	Imposible	Es el que no ocurre nunca, se identifica con ∅.

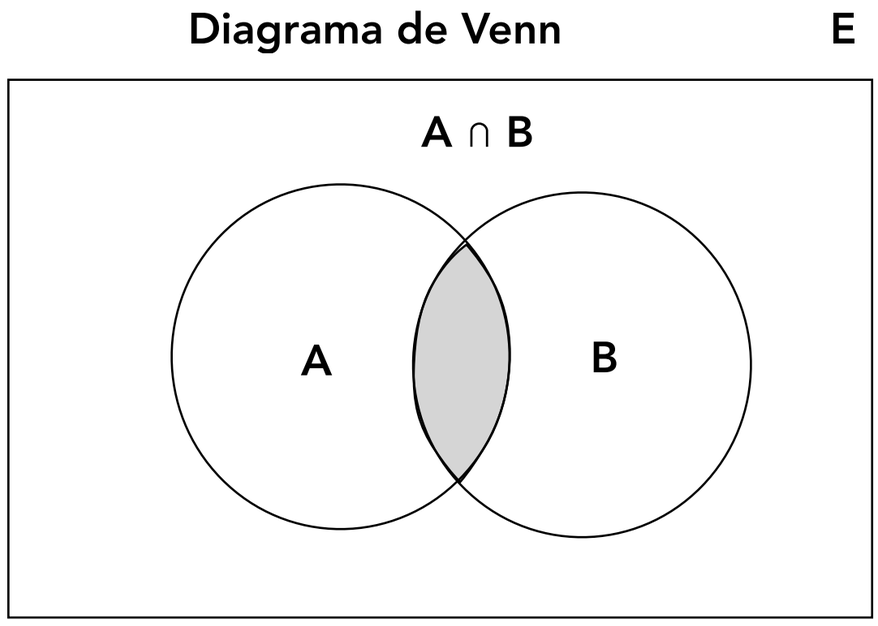
Karla Marmolejo

Diagrama de Venn



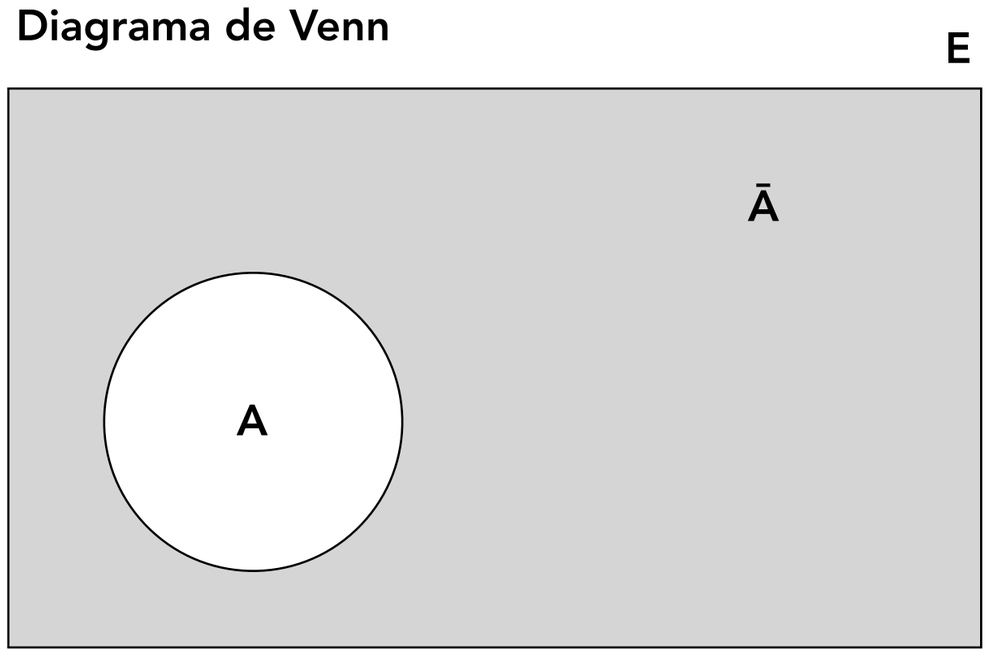
Unión—los sucesos elementales que pertenecen a A, a B o ambos a la vez.

Diagrama de Venn



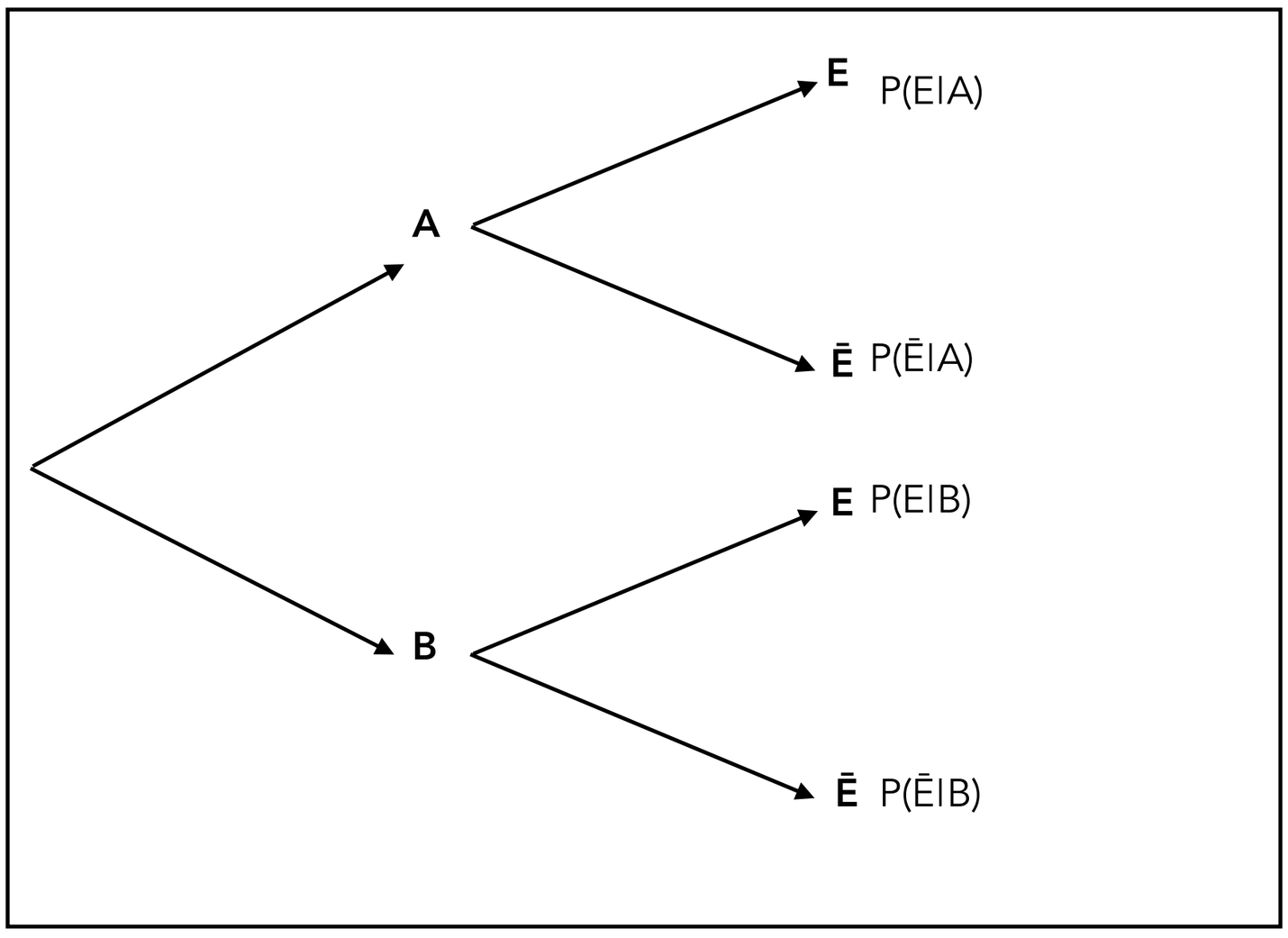
Intersección—los sucesos elementales que pertenecen a A y B simultáneamente.
Cuando la intersección de dos sucesos no contiene ningún elemento se dice que son sucesos incompatibles o excluyentes.

Diagrama de Venn



Complementario—Subconjunto de E formado por los sucesos elementales que no pertenecen a A.

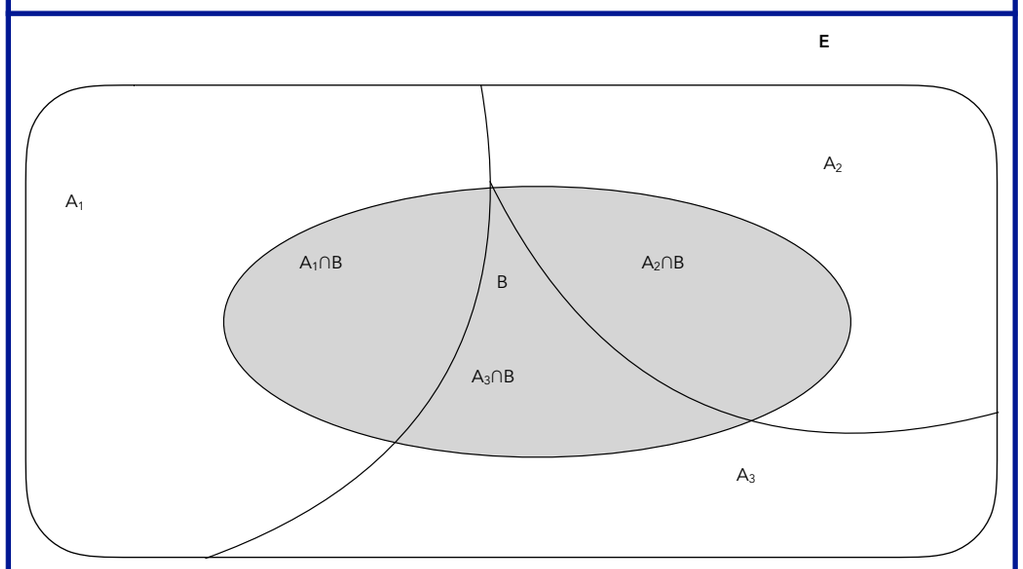
Diagrama de árbol



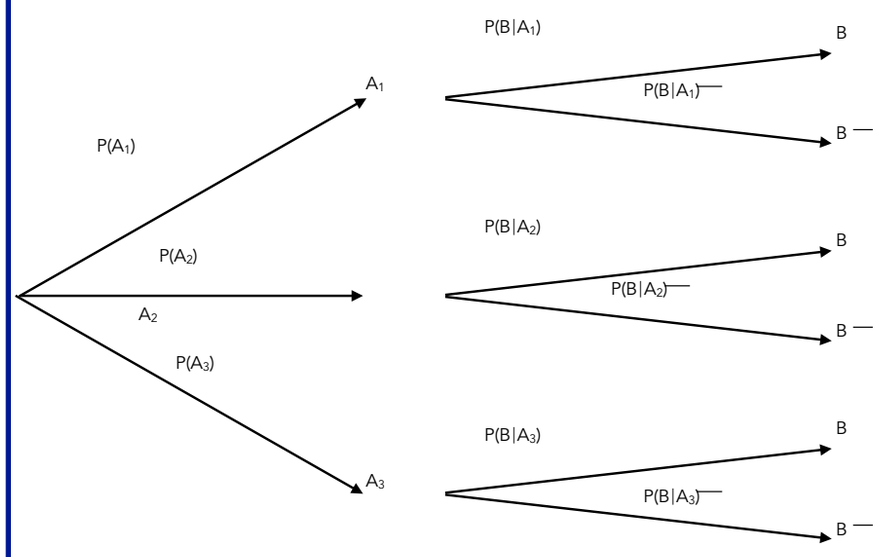
Karla Marmolejo

En teoría de la probabilidad se toman todos los posibles resultados de un experimento aleatorio como elementos del espacio muestral E.		
Probabilidad de un suceso	Una medida numérica que cuantifica la posibilidad de que dicho suceso ocurra. Los valores de probabilidad se encuentran comprendidos entre 0 y 1, en función de su cuantía de probabilidad de ocurrencia. Sucesos muy probables están cerca del 1 y los menos probables están próximos al 0. 1—sucesos seguros y 0—sucesos imposibles.	
Definición clásica o a priori y/o Regla de Laplace	Indica que la probabilidad de un suceso A es igual al cociente entre el número de casos favorables de que ocurra eses suceso y el número de casos posibles en el supuesto de que todos los casos tengan la misma oportunidad de ocurrir (sean equiprobables).	
	$P(A) = n_A / n$ P(A) = probabilidad de un suceso n_A = número de casos favorables n = número de casos posibles.	
	Problema—asume que todos los sucesos son equiprobables (tienen la misma probabilidad de ocurrencia)	
Definición estadística o a posteriori	Se basa en la estabilidad de las frecuencias relativas cuando el número de repeticiones de un suceso aleatorio es muy elevado y tiende al infinito.	
	Podemos definir P(A) como el límite al que tiende la frecuencia relativa de aparición de un suceso A cuando el número de ensayos n o repeticiones tiende a infinito	
	$P(A) = \lim_{n \to \infty} n_A / n$	
	Se llama a posteriori porque las probabilidades se atribuyen a cada suceso después de un gran número de repeticiones del experimento aleatorio.	
	La frecuencia relativa tiende a estabilizarse cuando el número de repeticiones del experimento es muy elevado. A este fenómeno de estabilización de las frecuencias se le conoce como **Ley del azar o ley de regularidad estadística.**	
Axiomática	Dado un espacio muestral E, se denomina probabilidad del suceso Ai, definido en el espacio muestral E y designado por P(Ai), a un número real asignado al suceso Ai, al que cumple las siguientes propiedades	$0 \leq P(Ai) \leq 1$
		$P(E) = 1$
		Si A_1, A_2.. Ak son sucesos incompatibles dos a dos entonces $P(A_1 \cup A_2 \cup Ak) = P(A_1)+P(A_2)\cdots+P(Ak)$
	Ventajas	Recoge las definiciones de probabilidad anteriores, ya que cumplen con la axiomática propuesta
		Permite el desarrollo matemático de la teoría de la probabilidad.
	Los axiomas son las condiciones mínimas que se deben cumplir par que la función definida sobre los sucesos determine consistentemente sus valores de probabilidad.	
	Teorema de la Suma	Establece la probabilidad de que ocurra el suceso A o el suceso B es igual a la probabilidad de que ocurra A más la probabilidad de que ocurra B, menos la probabilidad de que ocurran A y B (la intersección de ambos sucesos).
		$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Karla Marmolejo

Teoremas	Probabilidad condicionada	Los sucesos A y B son **dependientes** cuando la probabilidad de A depende o está condicionada al suceso B, lo que se denota como P(A | B)
		Para dos sucesos cualquiera A y B, la **probabilidad de A condicionada a B—P(A | B)** es igual a la probabilidad de la intersección dividida por la probabilidad de B.
		P(A | B) = P (A∩B) / P(B); siempre que P(B) ≠ 0
		P(B | A) = P (B∩A) / P(A); siempre que P(A) ≠ 0
		La forma de calcular la independencia o no de los dos sucesos puede realizarse aplicando la probabilidad condicionada. Nótese que si A y B son independientes: P(A | B) = P (A) y P(B | A) = P (B).
	Teorema del Producto	Se aplica a situaciones en las que se quiere calcular la probabilidad de que aparezcan dos sucesos de forma simultánea➜sirve para calcular la probabilidad de intersección entre dos sucesos.
		La probabilidad de la ocurrencia de A y B es igual a la probabilidad de ocurrencia de A por la probabilidad de ocurrencia de B, dado que A ha ocurrido previamente (condición previa).
		P (A∩B) = / P(A) • P(B | A)
		Cuando los sucesos son independientes, P(B | A) = P(B), por tanto P (A∩B) = / P(A) • P(B)
		Extracción con reposición—se mantiene siempre el mismo número de "papeletas" dado que las extraídas se devuelven.
		Extracción sin reposición—no se devuelven las "papeletas", por lo que las probabilidades de obtener una papeleta concreta en la segunda extracción dependerá de lo obtenido en la primera. Ver ejemplo 6.9 del libro, pag. 271
	Teorema de la Probabilidad Total	En un espacio muestral E, se dice que k sucesos A_1, A_2, …, Ak forman una partición del espacio muestral si se cumplen simultáneamente las siguientes condiciones:
		Ai∩Aj = ∅ para cualquier par de sucesos Ai y Aj, es decir, son incompatibles y su intersección es el conjunto vacío.
		$A_1 \cup A_2 \cup \cdots \cup Ak = E$ lo que implica que la unión de todos los sucesos es igual al espacio muestral (son exhaustivos).
		

Karla Marmolejo

		$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$
		$P(B) = P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_1) + P(A_3) \cdot P(B \mid A_1)$
		$\mathbf{P(B) = \sum P(Ai) \cdot P(B} \mid Ai)$
		
	Teorema de Bayes	En sucesos dependientes, la probabilidad condicionada tiene en cuenta información de un suceso para conocer la probabilidad de otro.
		Permite calcular cómo se modifican las probabilidades de determinados sucesos cuando se conoce alguna información adicional.
		Nos permite conocer las probabilidades de los sucesos A dado el suceso B.
		$P(Ai \mid B) = P(Ai) \times P(B \mid Ai) / P(B)$
		$P(A_i)$ son las probabilidades de cada suceso Ai,
		$P(B \mid A_i)$es la probabilidad del suceso B condicionada a los sucesos Ai
		$\sum(Ai)$ x $\mathbf{P(B \mid Ai)} = P(B)$
	Tanto el Teorema de Bayes como el de Probabilidad Total tienen como punto de partida la partición del espacio muestra en dos o más sucesos excluyentes.	

Karla Marmolejo

Aplicaciones de la Probabilidad Condicionada en Psicología de la Salud

En el ámbito de la Psicología de la Salud, y más en concreto, en el ámbito de la Epidemiología, se emplea con bastante frecuencia la probabilidad condicionada.

Prevalencia	Proporción de casos existentes de una enfermedad en un momento determinado. Nos indica la probabilidad de personas que tienen una característica o enfermedad en relación a la población.
Incidencia	Proporción de casos nuevos de una enfermedad en una población durante un período determinado. Representa la probabilidad de personas nuevas que pueden tener una característica o enfermedad en un período corto.

Exite una relación entre las anteriores➜si los casos nuevos (incidentes) no se resuelven se vuelven crónicos (prevalentes). Además, una disminución en la incidencia (casos nuevos) repercute en una menor prevalencia (casos existentes) y al revés.

Factores de riesgo

Probabilidad de que aumente un problema o enfermedad al estar expuesto a un riesgo. Este tipo de análisis parte de la base de que sujetos expuestos a un factor (X+) tienen más posibilidades o riesgo de sufrir una enfermedad o tener un problema psicológico (E+) en comparación con el grupo no expuesto (E-) a dicho factor (X-).

Valoración de la calidad de las pruebas de diagnóstico

Los datos se presentan en una tabla de doble entrada como la de abajo

	Resultado de la prueba	
	+	**-**
Trastorno (T)	Verdaderos positivos	Falsos negativos
No Trastorno (NT)	Falsos positivos	Verdaderos negativos

Sensibilidad P (+ | T) ➜probabilidad de discriminar a los verdaderos positivos—la probabilidad de que los que **SI** tienen el trastorno T den **positivo** en la prueba.

P (+ | T) = $n_{+\cap T}$ (verdaderos positivos) / n_T (Total con trastorno)

Especificidad P (- | NT) ➜probabilidad de detectar los verdaderos negativos—la probabilidad de que los que **NO** tienen el trastorno den **negativo** en la prueba.

P (- | NT) = $n_{-\cap NT}$ (verdaderos negativos) / n_{NT} (Total sin trastorno)

Una buena prueba diagnóstica es la que presenta alta sensibilidad y alta especificidad.

Sin embargo, las pruebas nunca son exactas y siempre se trabaja con un margen de error en la evaluación. En consecuencia hay dos valores predictivos:

Valor predictivo positivo P (T | +) ➜probabilidad de que todos los que den positivo tengan el trastorno—los falsos positivos disminuyen este valor.

P (T | +) = $n_{T\cap +}$ (verdaderos positivos) / n_+ (Total positivos)

Valor predictivo negativo P (NT | -) ➜probabilidad de que todos los que den negativo no tengan el trastorno—los falsos negativos disminuyen este valor.

P (NT | -) = $n_{NT\cap -}$ (verdaderos negativos) / n_- (Total negativos)

Karla Marmolejo

Tema 7
Variables aleatorias y modelos discretos de probabilidad

Variable Aleatoria	Intuitivamente➜cualquier característica medible que toma diferentes valores con probabilidades determinadas.	
	Es una función que asigna un número real, y sólo uno, a cada resultado posible de un espacio muestra E de un experimento aleatorio. Define las características o condiciones de los sucesos.	
	Cuando a cada suceso elemental o resultado posible del espacio muestran le asignamos un valor numérico se obtiene una variable aleatoria denominada X. Las variables aleatorias toman valores numéricos y se pueden definir diferentes variables sobre los resultados de un mismo experimento.	
	El azar interviene en el resultado que obtenemos al realizar el experimento aleatorio y no en la variable o función.	
	Las variables aleatorias se representan por letras mayúsculas del alfabeto latino, y se utilizan las letras minúsculas con subíndice para referirnos a los valores concretos que toman estas variables aleatorias.	
	Tipos de variables	**Discretas➜**adopta valores enteros; fijados dos valores consecutivos, no puede tomar ninguno intermedio
		Contínua➜cuando puede tomar infinitos valores o un conjunto de valores no numerable. Ejemplos: tiempo de reacción, estatura, cociente intelectual.

Karla Marmolejo

Función de probabilidad de una Variable Aleatoria

Distribución	En el caso de una variable aleatoria discreta X, vendrá dada por los valores que la variable puede tomar y su correspondiente probabilidad.
Función de probabilidad	Función que se asocia a cada valor de la variable la probabilidad de que ésta adopte ese valor
	$f(x) = P(X=x)$
	Se puede ilustrar en una tabla. La primera columna recoge el espacio muestral (todos los posibles resultados), siendo cada fila un suceso. En la segunda columna se muestran los valores que puede tomar la variable X anteriormente definida para cada suceso, y la tercera columna sus correspondientes probabilidades.
	La función de probabilidad de una variable aleatoria discreta puede representarse mediante un diagrama de barras donde en aleje de abscisas se recogen los valores que toma la variable y en el eje de ordenadas las correspondiente probabilidades.

X	0	1	2	3
f(x)	0.125	0.375	0.375	0.125

x = valores de X
f(x) = probabilidad de

Propiedades fundamentales	Para cualquier valor de x, f(x) siempre toma valores positivos o nulos. $f(x) \geq 0$
	La suma de todas las probabilidades correspondientes a cada valor de x es igual a uno. Formalmente $\sum f(x) = f(X_1) + f(X_2) + \ldots + f(Xn) = 1$

Karla Marmolejo

Función de distribución de una V.A. discreta

Nos indica cuál es la probabilidad de que la variable aleatoria tome un valor menor o igual que un valor concreto x.		
Función de Distribución	Se representa por F(x), función que asocia a cada valor de la variable de probabilidad de que ésta adopte ese valor o cualquier otro inferior.	
	$F(x) = P(X \leq x)$	
	$F(Xk) = P(X \leq x) = f(X_1) + f(X_2) + \ldots + f(Xk)$	
	Para diferenciar el concepto de función de probabilidad y función de distribución, hay que tener siempre presente que en el primero se le asigna a la probabilidad de un valor concreto, mientras que la función de distribución es **acumulada➜**se le asigna un valor concreto y todos los anteriores.	
	Propiedades fundamentales	Todos los valores que toma la función de distribución de la probabilidad son positivos o nulos. $F(x) \geq 0$
		F(x) es nula o vale 0 para todo valor inferior al menor valor de la variable aleatoria x_1. $F(x) = 0$ si $x < x_1$
		F(x) es igual a uno para todo valor igual o superior al mayor valor de la variable aleatoria. $F(x) = 1$ si $x \geq Xn$
		La función F(x) es no decreciente ya que es una acumulación o suma de probabilidades que son siempre positivas o nulas.
		La probabilidad P de que la variable aleatoria X tome valores x mayores que x_1 y menores o iguales que x_2 ($x_1 < x < x_2$) es la diferencia entre los valores de la función de distribución correspondientes a su valor superior menos su valor inferior. $P(x_1 < x < x_2) = F(x_2) - F(x_1)$

Karla Marmolejo

Media y Varianza de una V.A. Discreta

Media	Designada por la letra μ (mi) se calcula el sumatorio de los productos de cada uno de los valores que toma la variable por su correspondiente probabilidad.
	$\mu = E(X) = \sum x \cdot f(x)$
	La media de una variable X también se denomina **esperanza matemática o valor esperado** de x. Referido a una variable aleatoria representa el promedio teórico que tomaría la variable aleatoria si se repitiese el experimento aleatorio infinitas veces.

Varianza	Designada por la letra σ^2 (sigma cuadrada) o V(X), se calcula el sumatorio del producto de cada uno de los valores que toma la variable menos su media elevados al cuadrado multiplicados por su correspondiente valor de la función de probabilidad	
	$\sigma^2 = V(X) = \sum[(x-\mu)^2 \cdot f(x)]$	
	Fórmula alternativa	$\sigma^2 = V(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$
		$E(X^2) = \sum x^2 f(x)$
		$[E(X)]^2$ es la media de la variable elevada al cuadrado, μ^2
	Puede definirse también como la esperanza de los cuadrados de X, $E(X^2)$, menos el cuadrado de la esperanza de X, $[E(X)]^2$.	

Desviación típica	Es la raíz cuadrada de la varianza.
	$\sigma = \sqrt{\sigma^2} = \sqrt{\sum[(x-\mu)^2 \cdot f(x)]}$

Karla Marmolejo

Distribuciones Discretas de Probabilidad

En función d elas condiciones de partida del experimento y de las características de la variable aleatoria, se pueden ajustar estas distribuciones a alguna distribución (modelo teórico de probabilidad) ya conocida.
Existen diversas distribuciones teóricas para variables discretas, bien conocidas, por utilizarse frecuentemente como modelo, o por su interés como instrumento estadístico.
Por lo general, en Psicología y Ciencias de la Salud se trabaja con variables aleatorias discretas que sólo pueden tomar dos valores (dicotómicas) y que habitualmente se representan por 1 y 0.

Distribución de Bernoulli	La realización de un experimento aleatorio con sólo dos resultados posibles se trata de un experimento o ensayo denominado Bernoulli. El acierto o fallo a una pregunta con dos alternativas respondida al azar. A una de ellas se le denomina "éxito o acierto" (1) y a la otra "fracaso o error" (0), sin que los términos tengan connotaciones ni positivas ni negativas. **Se utiliza cuando hay un único ensayo.**	
	La **variable aleatoria discreta que sigue el modelo de Bernoulli** se define como una variable aleatoria dicotómica X, con dos posibles valores mutuamente exclusivos➜1 (éxito) con probabilidad p y 0 (fracaso) con probabilidad q, tal que la suma de ambas probabilidades sea igual a uno.	
		$f(1) = P(X=1) = p$
		$f(0) = P(X=0) = q$
		$p+q = 1$, por lo que $q = 1-p$
	Una variable aleatoria X que sigue el modelo de Bernoulli con parámetro p, se denota abreviadamente como $X \rightarrow Ber(p)$	
	Características fundamentales	Función de probabilidad $f(x) = P(X=x) = p^x q^{1-x}$
		Función de distribución $F(x) = P(X \leq x) = \sum p^x q^{1-x}$
		Media $\mu = p$
		Varianza $\sigma^2 = p(1-p) = pq$
		Desviación típica $\sigma = \sqrt{\sigma^2} = \sqrt{pq}$
		Donde x puede adoptar el valor de 1 (éxito) ó 0 (fracaso)
		p = probabilidad de éxito en el único ensayo del experimento
		q = la probabilidad del fracaso (1-p) en el único ensayo del experimento.

Karla Marmolejo

Distribuciones Discretas de Probabilidad

Distribución binomial	Una generalización de la distribución de Bernoulli en la que el experimento se repite más de una vez, consiste en repetir n veces y de forma independiente un ensayo Bernoulli en el que la probabilidad de éxito p se mantiene constante en cada un de los n ensayos.	
	Se representa por B(n,p) donde B indica binomial, n es el número de ensayos o veces que se repite el experimento y p la probabilidad de éxito. La distribución de Bernoulli sería un caso particular de la binomial con un parámetro n igual a 1.	
		$f(1) = P(X=1) = p$
		$f(0) = P(X=0) = q$
		$p+q = 1$, por lo que $q = 1-p$
	Una variable aleatoria X que sigue el modelo de Bernoulli con parámetro p, se denota abreviadamente como $X \rightarrow Ber(p)$	
	Características fundamentales	Función de probabilidad $f(x) = P(X=x) = \binom{n}{x} p^x q^{1-x}$
		Función de distribución $F(x) = P(X \leq x) = \sum \binom{n}{x} p^x q^{1-x}$
		Media $\mu = np$
		Varianza $\sigma^2 = np(1-p) = npq$
		Desviación típica $\sigma = \sqrt{\sigma^2} = \sqrt{npq}$
		x = número de aciertos
		n = número de ensayos
		p = probabilidad de éxito de cada uno de los ensayos
		q = la probabilidad del fracaso (1-p) de cada uno de los ensayos.
		$\binom{n}{x}$

Karla Marmolejo

Tema 8
Modelos continuos de probabilidad

<table>
<tr><td rowspan="5">Variable aleatoria continua</td><td colspan="3">Puede adoptar infinitos valores o un conjunto de valores no numerarles. Dado que estas variables pueden tomar infinitos valores y que, entro de cada intervalo de valores existen a su vez infinitos valores posibles, la probabilidad de que tome un valor determinado es nula o igual a cero➜las probabilidades se van a asignar a un determinado intervalo de la variable.</td></tr>
<tr><td rowspan="3">Función de Densidad</td><td colspan="2">Función de densidad de probabilidad de una variable aleatoria continua f(x)</td></tr>
<tr><td rowspan="2">Cumple las dos condiciones siguientes</td><td>$f(x) \geq 0$ Los valores de f(x) nunca son negativos</td></tr>
<tr><td>$\int_{\infty}^{\infty} f(x)dx = 1$ la integral en variables continuas es el análogo al sumatorio de las variables discretas➜el área total bajo la curva, que es igual a 1.</td></tr>
<tr><td colspan="2">Con la función de densidad de probabilidad de X podemos calcular la probabilidad de que X se encuentre en un determinado intervalo (a,b) mediante el cálculo integral. $P(a \leq X \leq b) = \int_a^b f(x)dx$ donde la integral está definida por el intervalo (a,b)</td></tr>
<tr><td></td><td colspan="2">Función de distribución acumulada o función de distribución de probabilidad de una variable aleatoria continua F(x)➜se asocia a cada valor de la variable X la probabilidad de obtener valores menores o iguales que un valor dado (lo mismo que decir menor, ya que la probabilidad de ser igual al valor dado es 0)</td></tr>
</table>

Karla Marmolejo

	Función de Distribución	$F(x) = P(X \leq x) = \int_{\infty}^{x} f(x)dx$	
		Mismas propiedades que en las variables discretas	$F(x) \geq 0$ Los valores de $F(x)$ nunca son negativos
			$F(-\infty) = 0$ y $F(+\infty) = 1$ por lo que $0 \leq F(x) \leq 1$ Está acotada entre 0 y 1.
			$\forall a \leq b$, $P(a \leq X \leq b) = F(b) — F(a)$ La probabilidad de que X se encuentra een el intervalo (a,b) es la diferencia entre la función de distribución para X = b y la función de distribución para X = a
	Media o valor esperado	$\mu = E(X) = \int_{\infty}^{\infty} x \cdot f(x)dx$	
	Varianza	$\sigma^2 = V(X) = \int_{\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx$	

Karla Marmolejo

La Distribución Normal

También llamada Campana de Gauss o curva normal, fue definida por De Moivre en un intento de encontrar las probabilidades acumuladas en una distribución binomial cuando n (número de ensayos) es grande.

Características y Propiedades	**Función de densidad de probabilidad**	$f(x) = (1 / (\sigma \sqrt{2\pi}))\, e^{-1/2(X-\mu/\sigma)^2}$ para $-\infty < X < \infty$
		μ = media o valor esperado de la distribución
		σ = desviación típica de la distribución
		π = 3,1416
		e = 2,718 (base de los logaritmos neperianos)
		Si una variable X tiene una función de densidad que se ajusta a la fórmula anterior➜se distribuye normalmente y lo expresamos por $X \rightarrow N(\mu, \sigma)$, indicando que tiene una distribución normal (N) con parámetros μ y σ.
		Su forma de campana es más apuntada cuanto menor es su desviación típica.
		Su figura nos indica que la puntuación de la mayoría de los individuos, en una variable que sigue esta distribución, se encuentra en torno a la media, y a medida que nos alejamos de esa puntuación, por su lado izquierdo y derecho, va disminuyendo la frecuencia.
		Una de sus propiedades fundamentales➜si le aplicamos una transformación lineal $Y = bX + a$, la nueva variable Y también se distribuirá normalmente pero con media $\mu_Y = b\,\mu_X + a$ y desviación típica $\sigma_Y = \lvert b \rvert \cdot \sigma_X$

Karla Marmolejo

Características y Propiedades		Propiedades Fundamentales	Es **simétrica** en torno a su media, que coincide con la mediana y su moda.
			La curva normal tiene **dos puntos de inflexión**➜donde la curva pasa a ser de cóncava a convexa. Estos puntos están situados a una distancia de una desviación típica de la media.
			Es **asintótica** en el eje de abscisas➜se extiende desde -∞ hasta +∞ sin nunca llegar a tocar el eje X.
	Distribución normal tipificada o normal estandarizada N(0,1)	$z = (X—\mu) / \sigma$	
		Se utilizan las tablas donde, la primera columna es la letra z, que corresponde a la puntuación típica, y la primera fila corresponde al segundo decimal de la puntuación z; los valores interiores representan probabilidades. Las puntuación típicas negativas están por debajo de la media y las positivas por encima de la media. Si la tabla no recoge el valor exacto de z, se puede utilizar el valor más próximo.	
		Cálculo de la probabilidad para valores menores o iguales a una determinada puntuación típica	Se utiliza la tabla directamente.
		Cálculo de la probabilidad para valores mayores que una determinada puntuación	Se busca en la tabla la probabilidad que esa puntuación deja por debajo y se resta de 1.

Karla Marmolejo

		Cálculo de la probabilidad entre dos puntuaciones determinadas	Se restan las probabilidades que dejan debajo de sí las dos puntuaciones típicas.
		Si se dispone de los datos originales de un grupo de sujetos en una determinada variable X y ésta se distribuye normalmente, se puede utilizar las tablas de la distribución normal estándar, primero transformando las puntuaciones directas en puntuaciones típicas.	
Para aproximar la distribución binomial a la normal se establecerá un intervalo entre 0,5 unidades a la izquierda y a la derecha de la puntuación. $P(X = x) = P[(x-0,5) \leq x \leq (x+0,5)]$			
La aproximación de la binomial a la normal se define como		$P(X=x) = P[((x-0,5)-np)/\sqrt{npq} \leq z \leq ((x+0,5)-np)/\sqrt{npq})]$	
Sumar y restar el valor 0,5 se llama corrección por continuidad, permitiendo así utilizar las puntuaciones discretas como si fuesen continuas. Para ello se interpreta cada puntuación X como si fuesen los puntos medios de los intervalos.			

Karla Marmolejo

Distribución χ^2 de Pearson	χ^2 se va a utilizar para hacer referencia a una distribución normal continua. Sean $X_1, X_2, \ldots, Xn$ un conjunto de n variables aleatorias independientes con una distribución N (0,1), entonces una nueva variable aleatoria $X = X_1^2+X_2^2+\ldots+X^2n$ sigue una distribución X^2n (se lee "ji cuadrado" con n grados de libertad) y se representa como $X \rightarrow X^2n$. Los parámetros correspondientes a la distribución χ^2 son:	
	Media o valor esperado $\mu = n$	Varianza $\sigma^2 = 2n$
	Los grados de libertad n indican que cada una de las n variables aleatorias puede tomar cualquier valor de sus posibles valores, sean cuales sean los valores tomados por las n-1 restantes.	
	Propiedades	Nunca adopta valores menores de 0
		Es asimétrica positiva pero a medida que aumentan sus grados de libertad se va aproximando a la distribución normal.
		Para n>30 la podemos aproximar a una distribución $N(n, \sqrt{2n})$
		Para usar la tabla, la primera fila recoge las probabilidades o proporciones (la proporción que ese valor deja por debajo de sí o el percentil al que corresponde) y la primera columna los grados de libertad correspondientes.
		Para hallar $P(X\geq X) = 1- P(X\leq X)$
Distribución t de Student	Sean X e Y dos variables aleatorias independientes, donde X sigue una distribución N(0,1) e Y una distribución χ^2. Entonces, la variable aleatoria $T = X/\sqrt{Y/n}$ sigue una distribución t con n grados de libertad y se expresa por $T \rightarrow t_n$	
	Media o valor esperado $\mu = 0$	Varianza $\sigma^2 = n/n-2$
	Una distribución t se puede definir como el cociente entre una variable N(0,1) y la raíz cuadrada de una variable χ^2_n, dividida por sus grados de libertad.	
	Propiedades	Es simétrica, con . Su forma es muy parecida a la N(0,1), aunque menos apuntada.
		Puede tomar cualquier valor entre $-\infty$ y $+\infty$
		A medida que aumentan los grados de libertad, la distribución se aproxima más y más a una distribución normal.
		La curva es asistólica al eje de abscisas.
		Fundamentalmente esta distribución se utiliza en estadística inferencia.
		La tabla presenta los valores positivos para esta distribución; en la primera columna se presentan los grados de libertad y en la primer fila las distintas probabilidades o proporciones de valores menores o iguales que un valor positivo dado (el precentil). Como es una distribución simétrica, podemos hallar las probabilidades asociadas a valores negativos a partir de los valores positivos. Los valores positivos se encuentran a la derecha de la media y los negativos a la izquierda.
	Si X_1 y X_2 son variables aleatorias independientes, con distribución χ^2 con n_1 y n_2 grados de libertad respectivamente, entonces una nueva variable F definida por $F = X_1/n_1 \div X_2/n_2$ sigue una distribución F con n_1 y n_2 grados de libertad (Fn_1,n_2). Siendo n_1 los grados de libertad del numerador y n_2 los grados de libertad del denominador.	
	Media o valor esperado $\mu = n_2/n_2-2$ para $n_2>2$	Varianza $\sigma^2 = 2n_2^2 (n_1+n_2-2)/n_1(n_2-4)(n_2-2)^2$ para $n_2>4$

Karla Marmolejo

Distribución F de Fisher-Snedecor	Propiedades	Es asimétrica positiva, por lo que nunca toma valores menores que 0.
		Una importante propiedad de esta distribución es la llamada propiedad recíproca, por lo que si X es una variable con distribución F con n_1 y n_2 grados de libertad, entonces la variable Y=1/X es también una distribución F con n_2 y n_1 grados de libertad.
		${}_pFn_1,n_2 = 1/(1-{}_pFn_2,n_1)$
		donde p y 1-p son las probabilidades acumuladas asociadas al valor de la variable. Esta propiedad es útil para obtener algunos perceptibles o probabilidades que no aparecen en la tabla. La tabla recoge solamente la probabilidad de que X sea menor o igual que 0,9000; 0,950; 0,975; 0,990 y 0,995.

Karla Marmolejo

Tema 9
Muestreo y distribución muestral de un estadístico

Inferencia Estadística	Disciplina que abarca la técnicas y métodos que permiten deducir las propiedades desconocidas de la población a partir de los datos obtenidos en la muestra. Los campos que comprende son **estimación de parámetros** y **contraste de hipótesis**. La piedra angular de ambos es el concepto de **distribución muestral de un estadístico**, que establece la relación entre las características de la población y el comportamiento de los estadísticos de las muestras que las representan.			
ESTADISTICA DESCRIPTIVA Recogida, organización y análisis de datos	→	PROBABILIDAD Permite y legitima el salto de las características (conocidas) de la muestra hasta las características desconocidas de la población	→	ESTADISTICA INFERENCIAL Permite obtener información acerca de la población a partir de la aportada por la muestra mediante la combinación de los modelos de probabilidad y de los estadísticos
En la mayoría de los casos es imposible trabajar con el conjunto total **población**. Por lo tanto, tendremos que buscar una **muestra**. Lo ideal es elegir muestras de tal forma y modo que representen y sean fiel reflejo de las características relevantes a la investigación en la población de trabajo.				
	Una colección, finita o infinita, de elementos que comparten ciertas características comunes, la cual definida por esas características que tiene en común dichos elementos que la componen.			
	Es independiente de la cantidad de elementos que la compongan, puede estar compuesta por un solo elemento.			

Karla Marmolejo

Población	Un elemento puede pertenecer a más de una población si cumple los criterios necesarios que definen cada una de ellas.	
	Un elemento pertenece a una población sólo durante el tiempo que se cumplen las propiedades que la definen.	
	FINITAS	Formadas por un número finito de elementos
	INFINITAS	Formadas por un número infinito de elementos.
	En general, las poblaciones son muy grandes y esto hace que sea prácticamente inviable trabajar con ellas. Por esta razón, lo habitual es trabajar con muestras.	
Parámetros	Los índices que representan los valores que resumen las características de las poblaciones. Son constantes, ya que se calculan con todos los elementos de la población	
Censo	El estudio de todos los elementos que componen la población. Ejemplo➜censo poblacional	
Muestra	Una parte de la población.	**REPRESENTANTIVA➜**cuando reúne las mismas características que la población. La representatividad de la muestra es la que garantiza que los resultados del estudio realizado puedan ser generalizados a toda la población.
	Tiene que permitir hacernos una idea general de cómo es la población. Debe ser una "*imagen reducida*" de las variables que pueden influir en el estudio para permitir la **generalización➜**desde la parte al todo, supone siempre un cierto error que se puede cuantificar y controlar a través de la Estadística mediante **error máximo (E_{max})** y **nivel de confianza ($1 - \alpha$).**	
Estadísticos	Los índices que representan los valores que resumen las características de las muestras, son variables aleatorias cuyos valores varían en función de los elementos que compongan la muestra.	

Karla Marmolejo

Muestreo	Conjunto de procedimientos y reglas que permiten extraer muestras de poblaciones garantizando su representatividad.	
PASOS	a) Definir los casos sobre los cuales se habrán de recolectar los datos.	
	b) Delimitar la población mediante una característica que defina, de forma exhaustiva y excluyente, a los individuos que la componen.	
	c) Elegir el método de selección de la muestra.	
	d) Calcular el tamaño de la muestra	
	e) Aplicar el procedimiento de selección	
	f) Obtener la muestra	
TIPOS	PROBABILISTICO	Se conoce la probabilidad que tiene cada elemento de la población de ser elegido para formar parte de la muestra y se conoce el ➜listado de elementos que componen la población. Su ventaja más importante es garantizar la representatividad y permitir hacer inferencias.
	NO PROBABILISTICO	No se conoce la probabilidad que tiene cada elemento de ser elegido. No garantiza la representatividad➜las conclusiones que se pueden extraer quedan circunscritas a la situación en la que se realizó el trabajo sin posibilidad de generalizar más allá de ese contexto. En general, se selecciona a los sujetos siguiendo determinados criterios procurando que la muestra resultante sea lo más parecida posible a la población.

Karla Marmolejo

MUESTREO PROBABILISTICO

Aleatorio Simple	Tomar de una población de tamaño N una muestra de tamaño n, mediante un procedimiento que garantice que todos los elementos de la población tienen la misma probabilidad de ser elegidos.
	Se asigna un número a cada elemento de la población
	A través de algún medio mecánico o informático se eligen al azar tantos sujetos como sea necesario para completar el tamaño de la muestra requerido. Todas las muestras posibles son equiprobables.
	Pero la probabilidad de cada una de ellas y la probabilidad de pertenencia de los elementos será distinta en función de la forma en que se genere la muestra:
	CON REPOSICION→tras elegir un elemento, éste se reincorpora a la población de forma que pueda ser elegido en la siguiente extracción.
	SIN REPOSICION→una vez seleccionado un elemento de la población no se reintegra, de esta forma la población va perdiendo tamaño y cambia con cada extracción. Aunque todos los elementos tienen la misma probabilidad de ser elegidos, ésta va cambiando según se realizan las extracciones→el resultado de una extracción no es independiente del resultado obtenido en las demás.
	La gran mayoría de los procedimientos de la Estadística Inferencial exigen el principio de independencia en la obtención de las muestras, lo cual no se cumple en el muestreo aleatorio sin reposición. Este problema se resuelve considerando, cuando el tamaño de la población N es grande con respecto al tamaño de la muestra n, que las probabilidades calculadas con ambos muestreos (con y sin reposición) son prácticamente iguales. Este tipo de muestreo tiene poca o nula utilidad práctica cuando la población que se está manejando es muy grande.

Karla Marmolejo

Aleatorio Sistemático	Cuando los elementos de la población están ordenados o pueden ordenarse, se puede utilizar el muestreo sistemático. En lugar de extraer n números aleatorios sólo se extrae uno (i). En número i del que se parte es un número elegido al azar, y los elementos elegidos para componer la muestra son los que ocupan los lugares i, i + k, i + 2k, i + 3k, …., i + (n-1)k; donde **k = N/n**
	En este tipo de muestreo no todos los elementos tienen la misma probabilidad de ser extraídos y las extracciones no son independientes.
	El riesgo de este tipo de muestreo está en aquellos casos en que se dan periodicidades en la población, ya que al seleccionar una periodicidad constante, los elementos seleccionados para la muestra pueden no ser representativos de la población.
Aleatorio Estratificado	Se utiliza cuando la población no es homogénea, debido a que existen grupos o estratos heterogéneos entre sí con gran homogeneidad dentro del estrato.
	Se pretende asegurar que todos los estratos estén representados adecuadamente en la muestra. Cada estrato funciona de forma independiente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o sistemático para elegir los elementos concretos que formarán parte de la muestra. El procedimiento de composición de la muestra se denomina afijación
	Afijación simple→a cada estrato le corresponde igual número de elementos muestrales.
	Afijación proporcional→la distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.
	La muestra total se forma por la suma de las muestras de cada estrato. Cada submuestra es independiente del resto. Permite aplicar técnicas de selección diferentes dentro de cada estrato y obtener estimaciones separadas de cada una de ellas.

Karla Marmolejo

Aleatorio por Conglomerados	La unidad muestral es un grupo de elementos de la población que conforman una unidad más amplia, a la que se llama **conglomerado**. Ejemplos➜áreas sanitarias, departamentos universitarios, una caja de determinado producto. **Los grupos YA están formados**.
	El procedimiento consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido) y trabajar con todos los elementos pertenecientes a los conglomerados elegidos.
	Si el número de elementos del conglomerado es muy amplio seleccionamos algunos de ellos al azar, en este caso decimos que es un muestreo por conglomerados bietápico o polietápico (más de dos etapas). Para aplicarlo a cada etapa se van seleccionando conglomerados de menor tamaño hasta que en la última etapa se trabaja con los n elementos que componen esos conglomerados
	Normalmente los tipos de muestreos que se combinan en polietápico son el conglomerado y el estratificado.

Karla Marmolejo

MUESTREO NO PROBABILISTICO

Por Cuotas o Muestreo Accidental	Se asienta sobre la base de un buen conocimiento de los estratos de la población y de los individuos más adecuados para los fines de la investigación.
	Tiene semejanzas son el muestreo aleatorio estratificado, pero sin carácter de aleatoriedad.
	Se fijan unas "**cuotas**", que consisten en un número de individuos que reúnen unas determinadas condiciones. Se eligen los primeros que se encuentren que cumplan esas características.
Opinático o Intencional	Un esfuerzo deliberado de obtener muestras "representativas" mediante la inclusión en la puesta de grupos supuestamente típicos.
	Un ejemplo muy conocido es el Estado de Ohio en EEUU, que se utiliza frecuentemente para la predicción de las elecciones a presidente de los EEUU. Es un estado cambiante (swing state), no tiene una tendencia del voto, unas veces votan republicanos y otras demócratas y desde principios del siglo XX sólo en dos ocasiones falló la predicción. Tiene una gran diversidad en su población en la que hay representantes de todos los tipos de votantes; es un Estado muy representativo por la gran mayoría de los diferentes tipos de votantes presentes en él.
Casual o Incidental	El investigador selecciona directa e intencionadamente los individuos de la población.
	El caso más frecuente es seleccionar individuos a los que tiene fácil acceso.
De Bola de Nieve	Las unidades muestrales van incorporándose paulatinamente a la muestra, a partir de las referencias de los sujetos que ya han participado en la investigación.
	Se localizan algunos individuos, los cuales conducen a otros, y estos a su vez a otros y así hasta conseguir la muestra suficiente.
	Se emplea frecuentemente en estudios con poblaciones "marginales", difíciles de identificar y de localizar.

Karla Marmolejo

Distribución Muestral de un Estadístico

Distribución Muestral de un Estadístico	Si en lugar de trabajar con toda la población se hace con una muestra, la distribución así obtenida es la **distribución muestral**. Es un concepto central, tanto de la estimación como del contraste de hipótesis.
	En una población cualquiera es posible extraer más de una muestra diferente del mismo tamaño. El valor concreto de un estadístico dependerá de los valores concretos que tomen cada uno de los elementos de la muestra extraída.
	El estadístico obtenido ya no será una constante, sino una variable, ya que su valor concreto dependerá de la muestra en la que se haya calculado.
	La distribución de probabilidad de todos los posibles valores del estadístico en las diferentes muestras➜**distribución muestral del estadístico**
	Dada una población de tamaño N de la que se van a extraer varias muestras de tamaño n, y para cada muestra se calcula un estadístico (por ejemplo la media $\ddot{x}$) de una variable aleatoria X cualquiera. Puesto que la media $\ddot{x}$ toma diferentes valores, dependiendo de cada muestra, el conjunto de las distintas medias forman a su vez una variable aleatoria que tendrá su propia distribución de probabilidad con sus características➜forma, media y varianza➞los parámetros que la definen. Estos parámetros se representan con letras griegas con su subíndice, que nos indica a qué estadístico nos estamos refiriendo.
	Es la distribución de probabilidad teórica de los valores de un estadístico cuando estos se calculan sobre las k muestras (siendo k muy grande, teóricamente infinito) de tamaño n, extraídas de la población y obtenidas mediante muestreo aleatorio simple.
Si de cualquier población con media μ y desviación típica σ, se toman todas las posibles muestras aleatorias (muestreo aleatorio simple) **con reposición**, cada una de tamaño n, la distribución muestral del estadístico media tiene como parámetros:	$\mu_{\ddot{x}} = \mu$, es decir, la media de las medias es igual a la media poblacional.
	Desviación típica o **error típico de la media** $\sigma_{\ddot{x}} = \sigma / \sqrt{n}$
	El número de muestras posibles es N^n
En el caso de muestras **finitas y muestreo sin reemplazamiento**, la desviación típica deberá multiplicarse por **$\sqrt{(N-n)/(N-1)}$**, donde N es el tamaño de la población y n el tamaño de la muestra.	$\mu_{\ddot{x}} = \mu$
	$\sigma_{\ddot{x}} = (\sigma / \sqrt{n})(\sqrt{(N-n)/(N-1)})$
A medida que aumenta el número de muestras, los valores de los parámetros de la distribución muestral de la media ($\mu_{\ddot{x}}$ y $\sigma_{\ddot{x}}$) se aproximan más a la media poblacional (μ) y al error típico de la media antes definido. $\sigma / \sqrt{n}$	

Karla Marmolejo

Distribución Muestral de la Media

La distribución muestral del estadístico media se puede caracterizar mediante su valor esperado o media $\mu_{\ddot{x}}$, su error típico $\sigma_{\ddot{x}}$, y la forma de la distribución. Hay que tener en cuanta la distribución de la variable de partida X en la población y si la varianza poblacional es conocida. A mayor variabilidad de la distribución de la media, mayor el error típico.	
Distribución normal de la variable X con varianza poblacional conocida	Si se extrae un número muy amplio de muestras de tamaño n, mediante muestreo aleatorio simple, la distribución de sus media tiende a una distribución normal definida como $N(\mu, \sigma/\sqrt{n})$ a medida que n tiende al infinito.
	$\mu_{\ddot{x}} = \mu$ esperanza matemática
	$\sigma_{\ddot{x}} = \sigma / \sqrt{n}$ varianza teórica
	$\sigma^2_{\ddot{x}} = \sigma^2 / \sqrt{n}$ desviación típica o error típico. A medida que aumenta el tamaño de la muestra, el error típico de la media disminuye. El error típico es inversamente proporcional al tamaño de la muestra (n),➜se puede afirmar que a medida que aumenta el tamaño muestral la distribución muestral de las medias se hace más homogénea (menor variabilidad).
	Normal con media mu y error típico sigma sobre raíz cuadrada de n➜ $\ddot{X} \rightarrow N(\mu, \sigma/\sqrt{n})$
	Podemos tipificar la variable $\ddot{X}$, calcular su puntuación $Z_{\ddot{x}}$ y la distribución sigue siendo normal.
	$Z_{\ddot{x}} = (\ddot{X} - \mu) / (\sigma / \sqrt{n})$
Distribución normal de la variable X con varianza poblacional desconocida	Es el caso más común. Se deberá estimar el valor del error típico de la media mediante la cuasidesviación típica muestral dada por:
	$\sigma_{\ddot{x}} = S_{n-1} / \sqrt{n}$ S_{n-1} = cuasidesviación típica
	$\mathbf{S_{n-1} = nS^2 / n-1}$
	La distribución muestral de la media ya no se ajusta a la normal, sino que sigue la distribución t de Student.
	$T = (\ddot{X} - \mu) / (S_{n-1} / \sqrt{n})$ (donde T son los valores internos de la tabla)
	Sigue una distribución t de Student con n-1 gl, se extraen muestras mediante muestreo aleatorio simple de tamaño n; la distribución muestral de las medias forma una variable aleatoria que se distribuye $t(\mu, S_{n-1}/\sqrt{n})$ con n-1 grados de libertad.
	Lo que ocurre generalmente es que las variables no se ajustan al modelo de la normal o simplemente se desconoce cuál es su varianza poblacional.

Karla Marmolejo

La variable X no se distribuye normalmente

Teorema Central del límite→ permite calcular las probabilidades asociadas a los valores de las medias sin necesidad de conocer la forma de la distribución de las variables, siempre que las muestras tengan tamaño suficiente **(n ≥ 30)**. Si una población tiene una media μ y una varianza σ^2 finitas, la distribución de las medias menstruales de tamaño n extraídas de manera aleatoria e independiente se aproxima a la forma de una **distribución normal con desviación típica y media** conforme el tamaño de n se va incrementando $N(\mu, \sigma/\sqrt{n})$, independientemente de las distribuciones que presenten $X_1, X_2, \ldots, X_3$

	población	muestra	Distribución muestral de la media
Media	$\mu = \frac{\Sigma X}{N}$	$X = \frac{\Sigma X}{n}$	$\mu_{\bar{x}} = \mu$
Varianza	$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$	$S^2_{n-1} = \frac{\Sigma(X - \bar{X})^2}{n - 1}$ Cuasivarianza	$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$
Desviación típica	$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$	$S^2_{n-1} = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}}$ Cuasidesviación típica	$\Sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$ Error típico de la media

Estimador

Un estadístico que se utiliza para estimar un parámetro. La media de la muestra es un estimador de la media poblacional; y el valor del estimador en una muestra se denomina **estimación** o **estimación puntual.**

La media muestral X es un **estimador insesgado** de la media poblacional μ. El error típico de la media es el indicador de la precisión de la estimación de la media; cuanto menor es el error típico, mayor es la precisión. Dependiendo de la desviación típica de la población y del tamaño de la muestra.

Karla Marmolejo

Distribución Muestral del Estadístico Proporción

Definimos π como la proporción de aciertos en la población.		
P = proporción de éxitos en las n extracciones, constante en todas las muestras, es decir, $P_1 = P_2 = \ldots = Pn = \pi$		
Para muestras pequeñas n<30	La distribución muestral de X (número de éxitos en n ensayos) es X → B (nπ, √nπ(1-π). donde $\mu_x = E(X) = n\pi$ y $\sigma_X = \sqrt{n\pi(1-\pi)}$. **DISTRIBUCION LINEAL DE EXITOS➜hay que multiplicar por n**	
	P es una transformación lineal de X (P = X/n), donde P es la proporción de aciertos en n ensayos, su distribución lineal es P → B (π, √π(1-π)/n). **DISTRIBUCION LINEAL DE LA PROPORCION**	
	Donde E(P) = π y $\sigma_P = \sqrt{(\pi(1-\pi)/n)}$. **DISTRIBUCION LINEAL DE LA PROPORCION**	
	Se utiliza la tabla de la binomial para determinar las probabilidades para diferentes tamaños muestrales y valores de π.	
	Supuestos	La variable aleatoria X es una variable Bernoulli (sólo dos valores éxito o fracaso)
		Se conoce π proporción en la población
		Las n observaciones son independientes
		La distribución muestral de la variable P es una distribución Binomial definida por B (π, √π(1-π)/n)
Para muestras suficientemente grandes n≥30	A medida que n crece, por el Teorema del Límite Central, se sabe que la distribución de las promociones se aproxima a la distribución normal con parámetros μ = E(P) = π y $\sigma^2_P = \pi(1-\pi)/n$	
	Aplicando la tipificación	Z = (P - π) / √(π(1-π)/n)
	Supuestos	La variable aleatoria X es una variable Bernoulli (sólo dos valores éxito o fracaso)
		Se conoce π proporción en la población (o el valor P en la muestra)
		Las n observaciones son independientes
		La muestra es suficientemente grande n≥30; si n<30, entonces deberá cumplirse que nπ≥5 y n(1-π)≥5
		Entonces los valores de las proporciones se aproximan a una normal definida por P → N (π, √π(1-π)/n))

Karla Marmolejo

Distribución Muestral del Estadístico Varianza

La distribución muestral del estadístico varianza S^2x sigue una distribución χ^2_{n-1}		
Parámetros $\mu_{S2x} = E(S^2) = (n-1)/n \; x \cdot \sigma^2$ y error típico o desviación típica $\sigma_{S2x} = \sigma^2 \sqrt{(2(n-1)/n)}$		
La distribución muestral del estadístico cuasivarianza (S^2_{n-1}) sigue una distribución χ^2_{n-1}		
Parámetros $\mu_{S2n-1} = E(S^2_{n-1}) = \sigma^2$ y error típico o desviación típica $\sigma_{S2n-1} = \sigma^2 \sqrt{(2/(n-1))}$		
VARIANZA MUESTRAL	$n\, S^2x / \sigma^2 = \chi^2_{n-1} \rightarrow S^2x = \sigma^2 \chi^2_{n-1} / n$	
	$E(S^2x) = \sigma^2 (n-1) /n$	
	$V(S^2x) = \sigma^2 \sqrt{(2(n-1))} /n$	
CUASIVARIANZA MUESTRAL	$(n-1)\, S^2_{n-1} / \sigma^2 = \chi^2_{n-1} \rightarrow S^2_{n-1} = \sigma^2 \chi^2_{n-1} / n-1$ **$S^2_{n-1} = nS^2/n-1$ (se obtiene por despeje de multiplicando por n/n Tema 3)**	
	$E(S^2_{n-1}) = \sigma^2$	
	$V(S^2_{n-1}) = \sigma^2 \sqrt{(2/(n-1))}$	
	La cuasivarianza es la mejor estimación de la varianza poblacional σ^2	
Para calcular las probabilidades asociadas a las varianzas, mediante la tabla de χ^2 se utilizan directamente las expresiones		
Se entiende que un test no mejora su capacidad de discriminación cuando la varianza poblacional se mantiene, por lo que si se hacen modificaciones en el test y se calcula la probabilidad de obtener varianzas iguales o mayores que la varianza de la muestra (S^2x), si esa probabilidad es pequeña, será indicador de que las modificaciones incrementaron la discriminación.		
$\chi^2 = n\, S^2x / \sigma^2$		$\chi^2 = (n-1)\, S^2_{n-1} / \sigma^2$
que se distribuyen χ^2_{n-1} (n-1, $\sqrt{(2(n-1))}$), **siempre y cuando X siga la distribución normal N (μ, σ) en la población.**		
CUANDO LA MUESTRA ES SUFICIENTEMENTE GRANDE n >= 100	La distribución de las variables nS^2x / σ^2 y $(n-1)\, S^2_{n-1} / \sigma^2$ se aproxima a la normal con media μ = n-1 y desviación típica $\sigma = \sqrt{(2(n-1))}$	
	Se tipifican los valores	$Z = (nS^2x/\sigma^2) - (n-1) / (\sqrt{(2(n-1))})$ VARIANZA
		$Z = (((n-1)S^2_{n-1})/\sigma^2) - (n-1)) / (\sqrt{(2(n-1))})$ CUASIVARIANZA
	Ambas con distribución aproximada N(0,1) lo que facilita la búsqueda de las probabilidades asociadas en la tabla de la Normal	

Karla Marmolejo

Tema 10
Estimación de parámetros y cálculo del tamaño muestral

El muestreo y la distribución muestral de un estadístico son la base de la ***Inferencia Estadística***➜la que permite derivar, a partir de los resultados de la muestra, los resultados que con una cierta probabilidad se pueden determinar para la población.

Se aplican estos modelos cuando se diseña una investigación y el investigador se plantea preguntas teóricas. Una vez planteadas las preguntas y evaluadas en relación con los conocimientos adquiridos a través del estudio y análisis de la bibliografía, se traducen a términos estadísticos y se comprueba si la situación planteada se parece a algún modelo; si es así, obtendremos una respuesta estadística a la pregunta.

Esta forma de trabajo exige garantizar que la realidad representada en el modelo teórico se ajusta al modelo matemático elegido para dar la respuesta estadística➜la comparación entre la pregunta estadística (la expresión matemática del modelo teórico) y los modelos de probabilidad y/o estadísticos debe hacerse cumpliendo los supuestos referentes a:

➤La métrica de las variables o nivel de medida

➤La forma de la distribución de la variable

Inferencia Estadística	Conjunto de métodos y técnicas basadas en los modelos Estadísticos y de Probabilidad➜permiten inducir a partir de la muestra, el comportamiento de las variables en la población➜emplea un razonamiento que va de lo particular a lo general (de muestra a población).	
	Dos ramas	**Estadística paramétrica**➜la distribución de las variables en la población es conocida, la muestra se selecciona por muestreo aleatorio simple y los datos están medidos al menos en escala de intervalos.

Karla Marmolejo

		Estadística no paramétrica➜la distribución de las variables no se ajusta a ninguna distribución conocida o los datos están medidos en una escala inferior a la escala de intervalo.
Estadística paramétrica	**Emplea dos procedimientos**	**Estimación de parámetros**➜asignar un valor numérico o determinar un intervalo de valores numéricos al parámetro o parámetros que deseamos conocer. Permite hacer conjeturas como ➤si el valor más próximo o el intervalo de valores entre los que se encuentre el valor de media de la población μ con un cierto grado de confianza ➤si el valor más próximo o el intervalo de valores entre los que se encontrará el valor de la diferencia a nivel poblacional con un cierto grado de confianza. ➜Se parte de los datos muestrales para responder a una pregunta sobre la población
		Contraste de hipótesis➜tiene como objetivo comprobar si un determinado supuesto, referido a un parámetro o parámetros poblaciones, es compatible con la evidencia empírica que no proporciona la muestra. Responde a preguntas como ➤es el valor de la media poblacional un valor determinado según el grado de confianza considerado suficiente ➤la diferencia entre los valores numéricos de la media es relevante o simplemente se debe al azar. ➜Se hace una afirmación sobre la población que luego se contrasta con la realidad d los datos obtenidos en la muestra.

Karla Marmolejo

ESTIMACION DE PARAMETROS—CUATRO TIPOS

Estimación Puntual	Asignamos un único valor al parámetro desconocido a partir del resultado obtenido en una muestra.
	La estimación puntual consiste en dar un valor numérico único al parámetro desconocido➜utilizar el valor del estadístico para estimar el parámetro
	No se puede establecer ni la fiabilidad de la estimación ni el error que se comete➜lo único que podemos afirmar es que el error cometido en la estimación se hará menor a medida que aumente el tamaño de la muestra
	No siempre es el más aconsejable y útil
Estimación por Intervalos	Un rango de posibles valores dentro del cual estimamos se encuentra el verdadero valor del parámetro con un determinado grado de confianza.
Estimación Bayesiana	Los parámetros se presentan como variables aleatorias con una cierta distribución a priori. Las observaciones aportan información que transforman las probabilidades a priori en probabilidades a posteriori.
Estimación Bootstrap	Se basan en el remuestreo y en las técnicas de simulación, por lo que se requieren el uso de ordenadores y consiste en extraer de una misma muestra varias (muchas) muestras y estudiar el conjunto de muestras así obtenidas
Para que un estadístico pueda considerarse un buen estimador de un parámetro deberá cumplir con las siguientes propiedades➜**carencia de sesgo, eficiencia, consistencia y suficiencia.**	

Karla Marmolejo

PROPIEDADES DE LOS ESTIMADORES

Carencia de Sesgo	**Estimador insesgado**➜aquel en el que se cumple que la media de la distribución muestral (esperanza matemática de la distribución) coincide con el parámetro estimado.	
	Sea θ el parámetro a estimar y θ el valor del estimador. Se define como $E(\theta) = \theta$	
	LA MEDIA	Se sabe que el estadístico media sigue una distribución normal o una t de Student. En cualquiera de las dos distribuciones $E(\ddot{X}) = \mu$ ➢la media de la muestra es un estimador insesgado de la media poblacional.
	LA PROPORCION	$E(P) = \pi$ ➢P es un estimador insesgado de π.
	LA VARIANZA Y CUASIVARIANZA	La varianza S^2x no es exactamente su correspondiente valor poblacional σ^2 ➢ S^2x es un estimador sesgado, cuyo sesgo es el factor (n-1)/n
		La cuasivarianza si es un estimador insesgado ya que $E(S^2_{n-1}) = \sigma^2$ ➢si queremos realizar una estimación puntual de σ^2, es preferible utilizar la cuasivarianza en lugar de la varianza.
	Dos estimadores θ_1 y θ_2 del mismo θ (parámetro), diremos que θ_1 es más eficiente que θ_2 si su varianza (la de la distribución muestral) es menor ➢ $\sigma^2_{\theta 1} < \sigma^2_{\theta 2}$	
	Entre dos estimadores insesgados será preferible seleccionar el que presente una menor varianza (menor error típico de la distribución muestral del estadístico). El error típico refleja el mayor o menor alejamiento de los posibles valores del estadístico a su esperanza matemática (media de la distribución muestral). Un estimador es tanto mejor cuanto su distribución muestral esté más concentrada➢tenga una varianza más pequeña.	

Karla Marmolejo

Eficiencia	Dado que $\sigma^2_{S2x} = \sigma^2\sqrt{(2(n-1))} / n$ y $\sigma^2_{S2n-1} = \sigma^2\sqrt{(2 / n-1)}$ ➢el error típico de la cuasivarianza es mayor que el de la varianza, por lo que la varianza es un estimador más eficiente que la cuasivarianza.
	La eficiencia del estimador siempre es relativa, ninguno puede ser perfectamente eficiente➜el error típico acompaña a cualquier distribución muestral.
	Eficiencia relativa ER = $\sigma^2_{\theta 1} / \sigma^2_{\theta 2}$ Si el cociente es igual a 1, ambos estimadores son igualmente eficientes Si ER > 1 el estimador del dominador es más eficiente Si ER < 1 el estimador del numerado es más eficiente La media es más eficiente que la mediana➜preferimos la media a pesar de que ambos estimadores son igualmente insesgados.
Consistencia	Indica que, a medida que el tamaño muestra se hace grande (que tiende al infinito), el valor del estadístico se aproxima al valor del parámetro
	Un estimador es consistente cuando la probabilidad de que su valor se aproxime al del parámetro es mayor conforme aumenta el tamaño de la muestra
	$\lim_{n\to\infty} P(\lvert\theta - \theta\rvert < \delta) = 1$ Si n tiende al infinito, la probabilidad de que $\lvert\theta - \theta\rvert$ sea menor que cualquier valor δ, por pequeño que sea éste, tiende a 1.

Karla Marmolejo

Suficiencia	Se refiere a la capacidad del estimador de utilizar toda la información existente en la muestra en relación al parámetro➜que el estimador emplee todos los valores de los datos para estimar el parámetro.
	El estimador suficiente de μ es $\ddot{X}$, al igual que la varianza, la cuasivarianza insertada y la proporción son estimadores suficientes de σ^2 y π respectivamente.
Propiedades de los estimadores media, proporción y cuasivarianza	La **media** muestral se considera buen estimador de la media poblacional ($\ddot{X}$ = μ). Cumple con las propiedades de ser **insesgado, consistente y suficiente**
	La **proporción** muestral (P) se considera buen estimador de la proporción poblacional (P = π). Cumple con las propiedades de ser **insesgado, consistente y suficiente**
	La **varianza** muestral (S^2x) es **consistente y suficiente**, pero **NO insesgado➜**la **cuasivarianza** se considera buen estimador de la varianza poblacional ($\sigma^2 = S^2_{n-1}$) cumple con las propiedades de ser **insesgada, consistente y suficiente,** aunque su eficiencia es menor en relación con la varianza muestral.

Karla Marmolejo

Métodos de obtención de estimadores

Existen varios métodos para la obtención de estimadores que garanticen las propiedades de carencia de sesgo, eficiencia, consistencia y suficiencia➜en Psicología y Ciencias de las Salud se utilizan habitualmente dos:	
MÉTODO DE MÍNIMOS CUADRADOS	Obtener aquel estimador que minimice las distancias (al cuadrado) entre el valor estimado y el parámetro y los resultados muestras observados
	Que $\sum(Xi - \theta)^2$ sea mínimo, donde i = 1, 2, 3, …., n
	No siempre es el mejor, pero resulta muy útil para estimar los parámetros de la regresión.
MÉTODO DE MÁXIMA VEROSIMILITUD	Obtener como estimador aquel valor del estadístico que hace lo más verosímil posible la muestra obtenida➜el que maximice la probabilidad de obtener el resultado particular observado en la muestra.

Karla Marmolejo

Estimación por Intervalos

La estimación por intervalos consiste en obtener una medida del error (diferencia entre el estimador y el parámetro) que se comete al realizar la estimación con una determinada probabilidad➜atribuir al parámetro un rango de valores posibles dentro del cual estará incluido el parámetro con una determinada probabilidad.	
Intervalo de confianza	Un rango de posibles valores del parámetro, sus límites se les llama **límites del intervalo de confianza**
Nivel de confianza n.c.	Corresponde con la probabilidad asociada al intervalo que contiene todos los posibles valores que puede tomar el parámetro θ. Se le llama de confianza y no de probabilidad➜el intervalo de confianza contendrá al verdadero valor del parámetro o no
Su principal ventaja➜se puede valorar la seguridad con la que se realizan las estimaciones mediante el nivel de confianza➞se expresa en términos de probabilidad.	
Características del intervalo de confianza	**Relación entre amplitud del intervalo y nivel de confianza➜**cuanto mayor es el intervalo de valores, mayor es la probabilidad de que se encuentre dentro de él el verdadero valor de θ y la estimación es menos precisa➞la **precisión de la estimación** se relaciona de forma inversa con el **nivel de confianza➢**a mayor confianza en la estimación, mayor será el intervalo (el rango de posibles valores del parámetro) y la estimación es menos precisa➢**a mayor confianza menor precisión**
	Fijación del nivel de confianza➜El investigador es quien decide y fija el valor del nivel de confianza. Por convenio en general se adoptan los niveles de confianza de 1 - α = 0,95 ó 1 - α = 0,99
	Nivel de riesgo o significación α➜el opuesto al nivel de confianza se llama **nivel de riesgo, margen de error o nivel de significación** y se representa por α➞indica la probabilidad de que el valor del parámetro NO se encuentra dentro de los límites definidos ➢se divide en dos partes iguales siendo el área correspondiente de cada una α/2
	Entre en **nivel de confianza (1- α)** y el **nivel de riesgo o significación α** existe una relación inversa➜cuanto mayor es el nivel de confianza menor es margen de error.

Karla Marmolejo

Cálculo del Intervalo de Confianza

Para calcular el intervalo de confianza de un parámetro es necesario conocer la distribución muestral del estadístico correspondiente y los parámetros que la definen➜la esperanza matemática y el error típico	
Para el parámetro μ con σ^2 conocida	**X sigue una distribución normal en la población y se conoce la varianza poblacional**
	$P(\ddot{X} - \lvert Z_{\alpha/2}\rvert\, \sigma/\sqrt{n} \le \mu \le \ddot{X} + \lvert Z_{\alpha/2}\rvert\, \sigma/\sqrt{n}) = 1-\alpha$
	Los límites del intervalo son: $L_{inf} = \ddot{X} - \lvert Z_{\alpha/2}\rvert\, \sigma/\sqrt{n}$ $L_{sup} = \ddot{X} + \lvert Z_{\alpha/2}\rvert\, \sigma/\sqrt{n}$
	Error máximo de estimación➜a semi-amplitud (la mitad de la amplitud) del intervalo de confianza➜indica que el investigador asume con un nivel de confianza del (1- α)% que la diferencia máxima entre el valor estimado a partir de la muestra y el valor real del parámetro es igual a $E_{max} = \lvert Z_{\alpha/2}\rvert\, \sigma/\sqrt{n}$
	Cuando el tamaño muestral sea grande (n>30) y queramos estimar la media poblacional, podemos utilizar el intervalo antes definido
Para el parámetro μ con σ^2 desconocida	**X sigue una distribución normal en la población pero se desconoce la varianza poblacional**
	La distribución muestral de la media $\ddot{X}$ sigue la distribución *t* de Student definida por: $\ddot{X} \rightarrow t(\mu;\, S_{n-1}/\sqrt{n})$ con n-1 grados de libertad $T = (\ddot{X} - \mu) / S_{n-1}/\sqrt{n}$ El intervalo de confianza para T sería $P(\ddot{X} - \lvert t_{n-1;\alpha/2}\rvert\, S_{n-1}/\sqrt{n} \le \mu \le \ddot{X} + \lvert t_{n-1;\alpha/2}\rvert\, S_{n-1}/\sqrt{n}) = 1-\alpha$
	Los límites del intervalo son: $L_{inf} = \ddot{X} - \lvert t_{n-1;\alpha/2}\rvert\, S_{n-1}/\sqrt{n}$ $L_{sup} = \ddot{X} + \lvert t_{n-1;\alpha/2}\rvert\, S_{n-1}/\sqrt{n}$
	X sigue una distribución desconocida y el tamaño muestral sea n≥30, Teorema del Límite Central $L_{inf} = \ddot{X} - \lvert Z_{\alpha/2}\rvert\, S_{n-1}/\sqrt{n}$ $L_{sup} = \ddot{X} + \lvert Z_{\alpha/2}\rvert\, S_{n-1}/\sqrt{n}$
Intervalo de confianza para el parámetro π aproximación a la normal	La distribución muestral del estadístico P cuando se cumple que la muestra es grande (n≥30; ó nP≥5 y n(1-P)≥5) es N(0,1) aproximadamente, por lo que $Z = (P - \pi) / \sqrt{\pi(1-\pi)/n}$. A partir de la tipificación de P $P(P - \lvert Z_{\alpha/2}\rvert \sqrt{P(1-P)/n} \le \pi \le P + \lvert Z_{\alpha/2}\rvert \sqrt{P(1-P)/n}) = 1-\alpha$ $L_{inf} = P - \lvert Z_{\alpha/2}\rvert \sqrt{P(1-P)/n}$ $L_{sup} = P + \lvert Z_{\alpha/2}\rvert \sqrt{P(1-P)/n}$

Karla Marmolejo

Intervalo de confianza para el parámetro σ^2	La varianza S^2x es un estimador sesgado de σ^2, siendo la cuasivarianza el estimador insesgado. Se utiliza la cuasivarianza muestral S^2_{n-1} como estimador de σ^2 $P\left([(n-1)\, S^2_{n-1}] / [\chi^2_{n-1,\alpha/2}] \geq \sigma^2 \geq [(n-1)\, S^2_{n-1}] / [\chi^2_{n-1,1-\alpha/2}] \right) = 1-\alpha$ $L_{inf} = [(n-1)\, S^2_{n-1}] / [\chi^2_{n-1,1-\alpha/2}]$ $L_{sup} = [(n-1)\, S^2_{n-1}] / [\chi^2_{n-1,\alpha/2}]$
	Cuando las muestras son grandes (n>100) la distribución muestral de la varianza insesgada se puede aproximar a la normal $N(\sigma^2; S^2_{n-1}\sqrt{2/n})$. Por lo tanto, **cuando n>100** $P\left(S^2_{n-1} - \lvert Z_{\alpha/2}\rvert\, S^2_{n-1}\sqrt{2/n} \leq \sigma^2 \leq S^2_{n-1} + \lvert Z_{\alpha/2}\rvert\, S^2_{n-1}\sqrt{2/n}\right) = 1-\alpha$ $L_{inf} = S^2_{n-1} - \lvert Z_{\alpha/2}\rvert\, S^2_{n-1}\sqrt{2/n}$ $L_{sup} = S^2_{n-1} + \lvert Z_{\alpha/2}\rvert\, S^2_{n-1}\sqrt{2/n}$

Karla Marmolejo

Significado del Nivel de Confianza

La probabilidad de que el parámetro desconocido se encuentre entre los límites del intervalo. Sin embargo, el concepto de probabilidad sólo es aplicable a variables y los límites del intervalo son valores constantes.
Para interpretar correctamente el nivel de confianza asociado al intervalo (por ejemplo 0,95)➜de cada 100 intervalos que construyamos, se puede esperar que 95 capten el valor del parámetro (intervalos correctos) y 5 no lo capten (intervalos incorrectos). Una proporción de 1- α contendrá al parámetro poblacional y una proporción de α no los contendrá.
Se habla de probabilidad cuando se hace alusión a la variable media, por eso al referirnos al intervalo hablaremos de confianza y no de probabilidad.

Generalización del la Construcción de Intervalos

Para **estimadores con distribución muestral conocida**, los pasos son:
1. Determinar el parámetro que queremos estimar y el estadístico (estimador) cumpliendo con las propiedades de un buen estimador
2. Conocer la distribución muestral del estadístico (estimador) y los parámetros que la definen (media y error típico)
3. Fijar el nivel de significación de α o el nivel de confianza 1- α ➜α lo fija el investigador; suele ser por convenio α=0,05 ó α=0,01
4. Determinar el **error máximo de estimación (E_{max})**➜producto del error típico de la distribución muestral del estadístico que estima al parámetro por el valor del estadístico (Z, T, F, etc) correspondiente al nivel de significación prefijado.
IC = [Estimador ± Error máximo de estimación] = [$\theta \pm E_{max}$]

Factores que afectan al intervalo de confianza

Nivel de confianza➜con 1- α = 0,95 tendremos intervalos menos amplios (o más precisos) que con la elección de 1- α = 0,99
Error de confianza➜medida de la variabilidad de la distribución muestral del estadístico➜depende del tamaño muestral n y de la homogeneidad de la muestra, afectando ambos factores al intervalo de confianza.
Tamaño muestral➜inversamente proporcional al error típico➜a mayor tamaño muestral n menor es el error típico➢ menor amplitud del intervalo y mayor precisión
Homogeneidad de la muestra➜si las muestras son homogéneas➜la varianza es pequeña➢ la desviación típica (poblacional o muestral), o sea el error típico será pequeño➢la amplitud del intervalo es menor y por tanto la precisión será mayor.

Karla Marmolejo

Cálculo del Tamaño Muestral

El tamaño muestral tiene que ser suficiente para garantizar la precisión deseada en la estimación de los parámetros. Debe ser un número entero por lo que se redondea siempre al inmediato superior. Gira en torno a los conceptos de error típico y de error máximo de estimación

La **Estadística Paramétrica** tiene dos procedimientos y para ambos se parte de datos muestrales.	**Estimación de parámetros**	Se responde a una pregunta sobre la población
		Los factores que influyen en la determinación del tamaño muestral El parámetro que se va a estimar El error máximo E_{max} que el investigador está dispuesto a admitir El nivel de confianza 1- α con el que se trabaja La precisión que se desea La variabilidad que presenta la población
	Contraste de hipótesis	Se hace una afirmación sobre la población que luego se comprueba
		Los factores son: El error tipo I (α) y error tipo II (β) y la potencia estadística Magnitud de la diferencia (o tamaño del efecto) Direccionalidad de la hipótesis Variabilidad de la población respecto a la variable en estudio

Tamaño muestral para el parámetro media

Conocida la varianza poblacional

$E_{max} = |Z_{\alpha/2}|\ \sigma/\sqrt{n}$

$n = (\ |Z_{\alpha/2}|\ \sigma/E_{max})^2 = \sigma^2 Z^2_{\alpha/2}/E^2_{max}$

Para poblaciones finitas y muestreo sin reposición debe multiplicarse por el factor de corrección $\sqrt{[(N-n)/(N-1)]}$

$n = (\sigma^2 Z^2_{\alpha/2} N) / (E^2_{max}(N-1) + \sigma^2 Z^2_{\alpha/2})$

Desconocida la varianza poblacional

La distribución muestral del estadístico media se ajusta a una distribución t de Student con n-1 grados de libertad

$E_{max} = |t_{n-1;\alpha/2}|\ S_{n-1}/\sqrt{n}$

$n = t^2_{n-1;\alpha/2}\ S^2_{n-1} / E^2_{max}$

Para poblaciones finitas y muestreo sin reposición debe multiplicarse por el factor de corrección $\sqrt{[(N-n)/(N-1)]}$

$n = (t^2_{n-1;\alpha/2}\ S^2_{n-1}\ N) / (E^2_{max}(N-1) + t^2_{n-1;\alpha/2}\ S^2_{n-1})$

Como no se conoce n, dos soluciones posibles:

Procedimiento iterativo

Iteración 0➜ g.l.=26; $t_{26;0,025}$ = -2,06; n = $-2{,}06^2 \times 23{,}48 / 0{,}75^2 = 177{,}14 \approx 178$

Iteración 1➜ g.l.=177; $t_{177;0,025}$ = -1,98 (aprox); n = $-1{,}98^2 \times 23{,}48 / 0{,}75^2 = 163{,}65 \approx 164$

Iteración 2➜ g.l.=163; $t_{163;0,025}$ = -1,98 (aprox); n = $-1{,}98^2 \times 23{,}48 / 0{,}75^2 = 163{,}65 \approx 164$

Aproximación a la normal

$n = Z^2_{\alpha/2}\ S^2_{n-1}/E^2_{max}$

Cómo podemos saber cuál es la varianza insesgada en nuestra muestra?

El procedimiento más cómodo y efectivo es obtener un valor aproximado de la desviación típica insesgada (S_{n-1}). A partir de estudios previos, o de la realización de un estudio piloto, partiendo de este valor de S_{n-1} se calcula el tamaño muestral n

Karla Marmolejo

Factores que afectan al intervalo de confianza

Cuando n>30

$E_{max} = Z_{\alpha/2} \sqrt{[P(1-P) / n]}$

$n = [Z^2_{\alpha/2} P(1-P)] / E^2_{max}$

Lo que se suele hacer en la práctica es suponer que la varianza de la distribución muestral es máxima (p = q = 1-p = 0,5), con lo que la muestra será casi con toda seguridad superior a lo que estrictamente necesario, pero no habrá que hacer suposiciones arriesgadas sobre el valor de p

Si la población es finita o el muestreo es sin reposición

$n = [Z^2_{\alpha/2} P(1-P) N] / [E^2_{max} (N-1) + Z^2_{\alpha/2} P(1-P)]$

Si no se conoce el valor de P se supone varianza máxima, es decir p = 0,5