

UNIVERSIDAD NACIONAL DE SAN LUIS

FACULTAD DE PSICOLOGÍA

METODOLOGÍA DE LA INVESTIGACIÓN I

Licenciatura en Psicología

Licenciatura en Psicomotricidad

Prof. Responsable: Dr. HORACIO DANIEL GARCIA

Trabajos prácticos: Lic. DANIEL PITONI, Lic. MAXIMILIANO SAPINO y Mg. ELIANA ZÁRATE



Unidad 3: Recolección de datos y organización de la información

Estadística: concepto y definición. Estadística descriptiva e inferencial. Recolección de datos: población y muestra. Organización de datos: matriz de datos. Introducción a los softwares estadísticos. Tabla de frecuencias, intervalos de clase; frecuencias absoluta, relativa y porcentual. Representación gráfica de una distribución de datos.

Autor: HORACIO DANIEL GARCIA

Año 2023



Estadística: concepto y definición

Seguramente, si nos ponemos a pensar qué es la estadística, vendrá a nuestra mente un conjunto de ideas relacionadas con números apilados y gráficos de diversos aspectos, que hacen referencia a una disciplina actual, reciente. A decir verdad, la estadística trabaja en gran medida con datos numéricos y representa sus resultados muchas veces a través de tablas y gráficos, pero, sin embargo, no podemos decir que se trate de una disciplina reciente.

Desde el origen de nuestra civilización, el ser humano se ha interesado por tratar de representar con números los datos que eran valiosos para su época; por ejemplo: cantidad de bolsas de maíz cosechadas en una determinada superficie, mortandad comparativa de las ovejas según la estación del año, cantidad de enfermos de tuberculosis, etcétera.... Por lo tanto, podemos decir que la estadística ha sido una disciplina auxiliar de todas las ciencias y actividades. Así, actualmente la estadística asiste a la economía con la finalidad de medir la evolución de precios y de la pobreza; a la sociología le permite estudiar los perfiles y la dinámica de los grupos sociales; al marketing le brinda información para establecer las estrategias más eficaces para difundir un determinado producto; en la salud la estadística brinda datos que permiten estipular el grado de prevalencia de las enfermedades, la edad de ocurrencia y los grupos más vulnerables de la población, etc. Como podrás ver, constituye una poderosa herramienta siempre vinculada con la generación del conocimiento.

La estadística que conocemos hoy en día, debe gran parte de su estatus a los trabajos matemáticos de aquellos hombres que desarrollaron la teoría de las probabilidades, con la cual la estadística fue adhiriéndose a las ciencias formales y haciendo contribuciones realmente importantes.

Dado que es una disciplina muy amplia, existen diferentes definiciones de la misma según el enfoque en el que se plantee. Para nuestro curso, podemos decir que:

....es una disciplina que tiene el objetivo de reunir, procesar e interpretar información cuantitativa concerniente a individuos, grupos techos etcétera, con la finalidad de describir, realizar previsiones para el futuro o tomar decisiones más efectivas.

Importancia del estudio de la Estadística

Cómo te puedes imaginar, en la investigación, la estadística cumple un rol fundamental. Permite describir fenómenos, indagar relaciones entre variables, encontrar diferencias entre grupos y, a veces, utilizando una muestra de participantes, lograr inferencias válidas para una población de individuos (siempre que la muestra tenga características similares a la población). Ahora bien, ¿por qué un psicólogo, un psicomotricista, u otros profesionales tienen que aprender estos conceptos?

Salazar & Castillo (2018) entienden que la estadística puede ser pensada al menos como:

- 1. Herramienta de trabajo en las prácticas profesionales:** La estadística aporta sus métodos para sintetizar, representar y establecer conclusiones sobre el comportamiento de datos que son obtenidos desde la misma práctica profesional. Por ejemplo, cuando estés trabajando con tus pacientes, los datos que obtengas acerca de cómo repercuten tus intervenciones podrán ayudarte a mejorar tu desempeño y tu efectividad en la ayuda que brindarás a ellos.
- 2. En la solución de problemas de los procesos investigativos:** Aquí la estadística es fundamental para resolver las preguntas: ¿Cómo mejorar mi investigación?, ¿Cómo puedo verificar si existe alguna relación entre las variables en estudio?, ¿Cómo puedo saber si hay diferencias en una determinada característica en dos grupos distintos?, ¿Cómo puedo determinar si existe alguna relación causal?, etc.
- 3. En la investigación teórica:** Acá la estadística configura una excelente herramienta que ayuda a la generación y enriquecimiento de teorías. Los datos que se analizan sobre hechos reales pueden permitir predecir el

comportamiento bajo circunstancias determinadas, refutar (rechazar) aquellas presunciones que no son válidas o confirmar aquellas que tengan suficiente respaldo empírico.

La Estadística es fundamental para muchas ramas de la ciencia, desde la medicina a la economía. En este sentido, para la psicología y la psicomotricidad resulta esencial para interpretar los datos que se obtienen de la investigación científica y de las prácticas profesionales que realizamos.

Entonces, más puntualmente... ¿Para qué sirve la Estadística?

Para Gorgas-García, Cardiel-López. & Zamorano-Calvo (2011), la estadística permite entre otras cosas:

- Analizar muestras: eligiendo un pequeño grupo, extraído de una población, se pueden hacer deducciones de ella a partir de lo observado en la muestra, sin necesidad de gastar tantos recursos ni tiempo como demandaría conocer las características de todas las personas que componen la población (sondeos de opinión, control de calidad, etc.).
- Descripción de datos: toda la información obtenida se puede resumir y representar de una manera clara y concreta que favorezca su interpretación (por ejemplo, usando gráficos y/o tablas).
- Contrastar hipótesis: por medio de diversos procedimientos estadísticos podemos determinar la validez de una hipótesis y, de esta manera, garantizamos que las conclusiones a las que arribamos en nuestras investigaciones sean lo suficientemente sólidas y válidas.
- Medir relaciones entre variables: por ejemplo, podemos interesarnos en averiguar si existe relación entre un rasgo de personalidad y la forma de abordar las situaciones traumáticas, en este caso, la estadística provee procedimientos que no solo corroboran tal posibilidad, sino que además nos brindan información acerca de cuál es la tasa de error que podemos tener en esa afirmación.
- Predecir: estudiando las posibles combinaciones de relaciones entre las variables, se puede determinar los valores desconocidos de una variable puntual.

Estadística descriptiva e inferencial

Verás, hay dos tipos de estadísticas que difieren tanto en sus propósitos, en sus requerimientos, así como en sus alcances.

Se puede decir que el objetivo central de LA ESTADÍSTICA DESCRIPTIVA es detallar pormenorizadamente el comportamiento o las características del objeto de estudio. Para esto, recoge información de determinadas particularidades de una población y luego la representa en forma conveniente, útil y comprensible; de tal forma que se puede describir las distintas facetas que componen un mismo suceso.

Es importante subrayar que *la estadística descriptiva aporta conclusiones valederas únicamente para el conjunto analizado*. Por ejemplo, si un investigador estudia la sintomatología asociada a la depresión en un grupo determinado, su análisis y posibles explicaciones quedarán circunscriptas a ese grupo en particular, no pudiéndolas generalizar a otros grupos ni a la población. Con la estadística descriptiva podemos brindar información de niveles promedios en diversas variables (Ansiedad, Depresión, Rasgos de personalidad, Habilidades sociales, Tono muscular, Postura corporal etc. Etc.), y de porcentajes en variables como Género, Nivel socioeconómico, Escolaridad, etc., además de una serie de análisis complementarios que veremos en esta unidad y la unidad cuatro.

La realización de gráficas (visualización de los datos en diagramas) también forma parte de la Estadística Descriptiva, dado que proporciona una manera visual directa de organizar la información.

La finalidad de la Estadística Descriptiva no es, entonces, extraer conclusiones generales sobre el fenómeno que ha producido los datos bajo estudio, sino solamente su descripción (de ahí el nombre) (Gorgas-García, Cardiel-López, & Zamorano-Calvo, 2011).

En cambio, LA ESTADÍSTICA INFERENCIAL se ocupa de procurar generalizar la información (muchas veces obtenida por los métodos de la estadística descriptiva) intentando hacer inferencias (estimaciones) acerca de las características de la población, partiendo de datos de una muestra (un grupo de personas). Es decir, investiga a una población, valiéndose de los datos y resultados que se obtienen de una muestra. Es importante que sepas que para que esto sea posible, las muestras deben tener ciertas características que verás un poco más adelante en esta unidad. Además, este tipo de estadística se utiliza para analizar asociaciones entre variables, para establecer si los niveles medios de una variable difieren en dos o más grupos y para determinar la eficacia de un tratamiento o de una técnica, entre otras tantas posibilidades.

Este poder de generalización de datos, de predicción de resultados, ha favorecido que la estadística inferencial tenga un desarrollo creciente y, asimismo, una acentuada evolución de diversos métodos y estrategias estadísticas (Salazar. & Castillo, 2018).

Población y muestra

Decimos que la población es el conjunto total de datos provenientes de personas, objetos, ideas o acontecimientos, que le interesa al investigador y que pretende estudiar. En estadística, la población tiene un concepto un tanto más amplio que el demográfico. Para los investigadores, la población incluye, no a las personas de manera indefinida, sino a todos los elementos que han sido escogidos para su estudio porque tienen una característica de interés. Por ejemplo; cuando nos proponemos estudiar las características de la población de Argentina que presenta dificultades psicomotoras, nos estamos refiriendo a la totalidad de personas que poseen una particularidad bien definida.

Como podrás ver en el ejemplo estamos haciendo referencia a un conjunto de elementos (totalidad de personas) que poseen características comunes (dificultades psicomotoras) y sobre los cuales el investigador hará extensivas las conclusiones de su trabajo. Igualmente, lo que entendemos como población, guarda relación con la delimitación del problema y el objeto de estudio (ya que esto queda determinado por el interés del investigador); siguiendo con el ejemplo que propusimos más arriba, no será considerado como población al conjunto de personas con deserción académica, ya éstas tendrán características distintas a las fijadas por el investigador.

La muestra, en cambio, es una porción de la población. Siguiendo con el ejemplo, podrás apreciar que hacer un estudio que evalúe la totalidad de las personas que tienen problemas psicomotores en Argentina puede traer muchos inconvenientes (dificultad de localizar la totalidad de las personas, problemas de costos, demoras en el relevamiento de los datos, etcétera), por tal motivo surge la necesidad de adecuar la recolección de datos a una forma que permita, tomando un menor número de individuos, arribar a las mismas conclusiones que si se evaluara a toda la población de interés.

Ahora bien, para que podamos considerar que un subconjunto de la población pueda ser una muestra, ésta tiene que reflejar adecuadamente las características de la población de la cual fue extraída. A esto se le llama representatividad de la muestra y es un aspecto muy importante para poder generalizar los resultados. Decimos que una muestra es representativa cuando contiene características similares de la población y presenta márgenes de error pequeños, los que deben ser calculados por el investigador.

En función a esto, podemos hacer una primera categorización entre muestras probabilísticas y no probabilísticas:

- No probabilísticas: la característica principal de este tipo de muestra es que son fácilmente accesible para el investigador porque no representan mayores dificultades en la selección de los participantes que la componen, pero por lo general son mucho menos representativas que las probabilísticas, ya que tienen un riesgo mayor de tener un sesgo.
- Probabilísticas: este tipo de muestra implica una cuidadosa selección de los participantes en base a las probabilidades de que la misma posea características similares a la población de la cual es extraída.

Más adelante veremos en mayor detalle y profundidad estas estrategias de muestreo, explicaremos las conveniencias y las limitaciones de los distintos tipos de cada una de ellas.

Hay algo importante que debes saber.

*Cuando nos referimos a una característica global de la población hablamos de **Parámetro** (en general, un parámetro no es conocido. Por ejemplo, la edad promedio de una población de habitantes de una región). En cambio, cuando lo que nos interesa saber es una característica de la muestra hablamos de **Estadístico** (es un valor conocido, que varía de una muestra a otra y se puede usar para obtener conclusiones acerca de la población. Por ejemplo, el promedio de edad de los habitantes seleccionados en una muestra es un estadístico y se puede utilizar para estimar la edad promedio de la población de la que se obtuvo la muestra).*

Recolección de datos

Bien, sabemos que en una investigación recopilamos un conjunto de información que luego analizaremos para tratar de arribar a determinadas conclusiones. Cada información que obtenemos, a través de distintos instrumentos de medida, pasa a configurar un dato. Un dato estadístico es cada uno de los valores que se han obtenido de una o de múltiples variables al realizar un estudio.

Un dato puede ser el sexo de una persona, la edad, una característica psicológica, un atributo físico... en fin, cualquier tipo de atributo que configure una variable que pueda ser registrada o medida.

Seguramente estarás pensando que en una investigación habrá mucha cantidad de datos; en efecto, verás... si a 10 personas le preguntamos de que género son tendremos 10 datos (uno por cada persona), pero si además indagamos otra variable como escolaridad tendremos entonces 20 datos (10 acerca del género y 10 del nivel de formación que poseen). Usualmente, una investigación cuantitativa en el campo de la psicología o de la psicomotricidad suele estar conformada por numerosos participantes y muchas variables, por lo que la cantidad de datos es enorme.

Pensemos por un momento en el siguiente escenario: Estudiaremos en una muestra de 200 participantes las siguientes variables: Sexo, Edad, Escolaridad, Lugar de residencia, Cantidad de horas de trabajo, Depresión, Ansiedad y Apoyo social percibido (8 variables)... ¿Cuántos datos tendremos?... bueno, la cuenta es fácil: $200 \text{ (participantes)} \times 8 \text{ (variables)} = 1600 \text{ datos}$. Como verás es un número importante de datos que debemos organizar adecuadamente ya que, de lo contrario, no podremos obtener ninguna información concreta.

Organización de datos: matriz de datos

Como hemos visto, un dato es la información que tenemos acerca de una característica de un individuo en particular. Sin embargo, será sobre un conjunto grande de datos que el investigador podrá arribar a conclusiones acerca de las hipótesis o los objetivos que planteó. Es por este motivo que debemos organizar los datos a través de lo que se denomina una matriz (de datos).

¿Qué es la matriz de datos?

La matriz de datos es un conjunto de información (datos) que ha sido ordenado en una estructura de filas y columnas. En las filas se ubicarán los casos o individuos (por ejemplo, participante 1, participante 2, participante 3, participante 4, etc.), en las columnas las variables, por ejemplo, Edad, Sexo, Experiencia laboral, Autofocalización, Experiencia emocional, Evitación, Búsqueda de apoyo, Religión, etc. y, finalmente, los datos se ubicarán en la intersección de cada fila con cada columna.

VARIABLES

sujeeto	Edad	Sexo	Expe.laboral	Autofocalizacion	Exper.emocional	Evitacion	Eusq.Apoyo	Religion
1	32	Mujeres	0-5 años	8	7	4	19	24
2	33	Mujeres	6-10 años	8	4	7	17	23
3	29	Mujeres	6-10 años	13	5	10	6	15
4	30	Mujeres	0-5 años	5	6	12	18	16
5	29	Mujeres	0-5 años	9	6	10	18	24
6	43	Hombres	+ de 20 años	2	3	4	7	11
7	44	Mujeres	11-20 años	4	4	7	5	17
8	25	Hombres	0-5 años	14	10	9	18	2
9	42	Mujeres	6-10 años	8	4	13	24	13
10	27	Hombres	0-5 años	12	8	11	21	9
11	53	Mujeres	+ de 20 años	12	7	15	12	24
12	45	Hombres	11-20 años	9	8	13	23	18
13	49	Mujeres	+ de 20 años	4	5	13	11	20
14	25	Mujeres	0-5 años	9	5	18	16	14
15	32	Hombres	6-10 años	6	5	16	23	13
16	40	Mujeres	11-20 años	6	3	8	5	0
17	25	Hombres	0-5 años	5	12	11	7	8

SUJETOS (circulo rojo a la izquierda) **DATOS** (circulo rojo a la derecha)

Figura 1: Matriz de datos extraída del programa estadístico SPSS

Los datos pueden ser numéricos, o también características que se pueden expresar por palabras, esto depende de la naturaleza de la variable y su nivel de medición (tema que hemos visto con anterioridad). Los datos numéricos pueden ser: la altura, el peso o bien el resultado de un determinado test. Otras variables, como por ejemplo sexo, no son naturalmente numéricas, pero se las suele codificar mediante números para procesarlas a través de los programas estadísticos (por ejemplo, Hombre = 0; Mujer = 1).

Ahora bien, como podrás ver, una vez ordenados los datos tenemos mejores condiciones para analizar la información... es decir, están mejor organizados, aunque aún no podemos tener una impresión acabada de cuáles son las características del conjunto de esos datos, hasta tanto no los procesemos. Cuando lo hagamos podremos determinar qué cantidad de mujeres y hombres componen el análisis, cuál es el valor promedio de las variables cuantitativas, etcétera.

Para obtener esa información existen diversos métodos estadísticos. En nuestro curso veremos más adelante algunos de ellos, pero antes deseamos que te familiarices con el uso de los programas estadísticos que tanta ayuda nos ofrecen al realizar automáticamente análisis matemáticos extremadamente complejos.

Introducción a los softwares estadísticos

Sin duda uno de los hechos que más ha acercado a la gente a la estadística es el uso de la informática, ya que, con anterioridad, procesar la información que provenía del trabajo de campo llevaba un considerable de tiempo aplicado a las matemáticas, hasta tanto llegar a alguna conclusión. Era un esfuerzo considerable que se traducía en una menor producción de conocimiento.

Hoy en día el proceso es mucho más simple, el usuario no tiene por qué conocer las fórmulas, aunque si tiene que saber interpretar información y procesarla de acuerdo a las condiciones de sus variables y a los requerimientos de los estadísticos aplicados. Si la persona tiene desconocimiento o un plan de trabajo deficiente, el Software no resolverá sus problemas, más bien puede complicarlo aún más... el Software debe ser usado con seriedad, con conocimiento de los principios que rigen la estadística, eligiendo la técnica adecuada en base a los objetivos que se quieran alcanzar.

Un software estadístico es un programa de computación que está particularmente orientado a resolver las formulaciones matemáticas de los análisis estadísticos. Existe una gran cantidad de ellos, algunos te los comentaremos a continuación.

EXCEL: Si bien no es un programa estadístico en el sentido estricto, posee muchas capacidades para ser utilizado. La ventaja del mismo es que la mayoría de los ordenadores tienen una versión de éste o de algún otro sistema similar con licencia libre.

SPSS: Podemos decir que es el software estadístico más conocido a nivel mundial para ciencias sociales y humanas. Presenta una gran flexibilidad para transformar los datos y para analizarlos a través de un gran número de pruebas estadísticas. Tiene la

posibilidad de realizar gráficos, aunque éstos son un tanto básicos y deslucidos. La desventaja que presenta es que es una versión paga, sin embargo, hay una versión demo en [spss.com/es/](https://www.spss.com/es/)

MINITAB: es otro de los programas más usados. Permite calcular la mayoría de los estadísticos habituales: análisis exploratorio de datos, gráficos estadísticos, estadística no paramétrica, regresión y sus variantes, análisis multivariado de datos, etc. Similar al SPSS, es una versión paga en la cual se puede encontrar versiones de prueba por un tiempo limitado

STATGRAPHICS: es un programa de fácil manejo y, por lo tanto, una buena herramienta para la enseñanza de la estadística y para la investigación. En [statgraphics.net](https://www.statgraphics.net) se puede descargar una versión demo del programa (en español de duración limitada)

INFOSTAT: Es un programa desarrollado por docentes de estadística de la Universidad Nacional de Córdoba -Facultad de Ciencias Agropecuarias. Tiene una muy buena facilidad de transportar datos a programas de Windows y gran potencia como herramienta de cálculo estadístico.

PSPP ya que es una aplicación libre y gratuita para el análisis de datos. Proporciona análisis útiles como: frecuencias, tablas cruzadas, comparación de media; regresión lineal, fiabilidad, reordenamiento de datos, pruebas no paramétricas, análisis factoriales, entre otras características. Lo puedes descargar en <https://www.gnu.org/software/pspp/get.html>

Finalmente te presentamos el software con el cual trabajaremos en este curso: JAMOVI

Jamovi, es un programa estadístico de distribución gratuita, se puede descargar en <https://www.jamovi.org/download.html>. Además, para quienes tengan un ordenador antiguo, sin muchas prestaciones, existe la posibilidad de trabajar en la nube <https://cloud.jamovi.org/>

Jamovi se desarrolló en base a un proyecto que buscaba desarrollar una plataforma estadística de uso intuitivo y que además sea abierta, gratuita y que pueda proporcionar los últimos avances en metodología estadística. La filosofía de base que tienen sus desarrolladores es que el software científico debe ser "impulsado por la comunidad", donde cualquiera puede desarrollar y publicar análisis, y ponerlos a disposición de una amplia audiencia.

Este programa funciona como una hoja de cálculo avanzada que tiene la capacidad para realizar complejos cálculos estadísticos de una forma simple y eficaz, lo que favorece su uso para aquellas personas que se están iniciando. Sin embargo, su capacidad para incorporar diversos módulos, lo hace realmente atractivo para investigadores más avanzados.

Instalación

El proceso de instalación en un equipo, es simple. A continuación, te mostramos los pasos que deberías seguir:

Primero deberás ingresar a la web oficial de Jamovi (<https://www.jamovi.org/download.html>), luego tienes que elegir la plataforma que usa tu dispositivo (Windows, Mac o Linux), en ese momento descargarás el archivo instalador y una vez descargado podrás instalarlo definitivamente en tu computadora.

Visualización

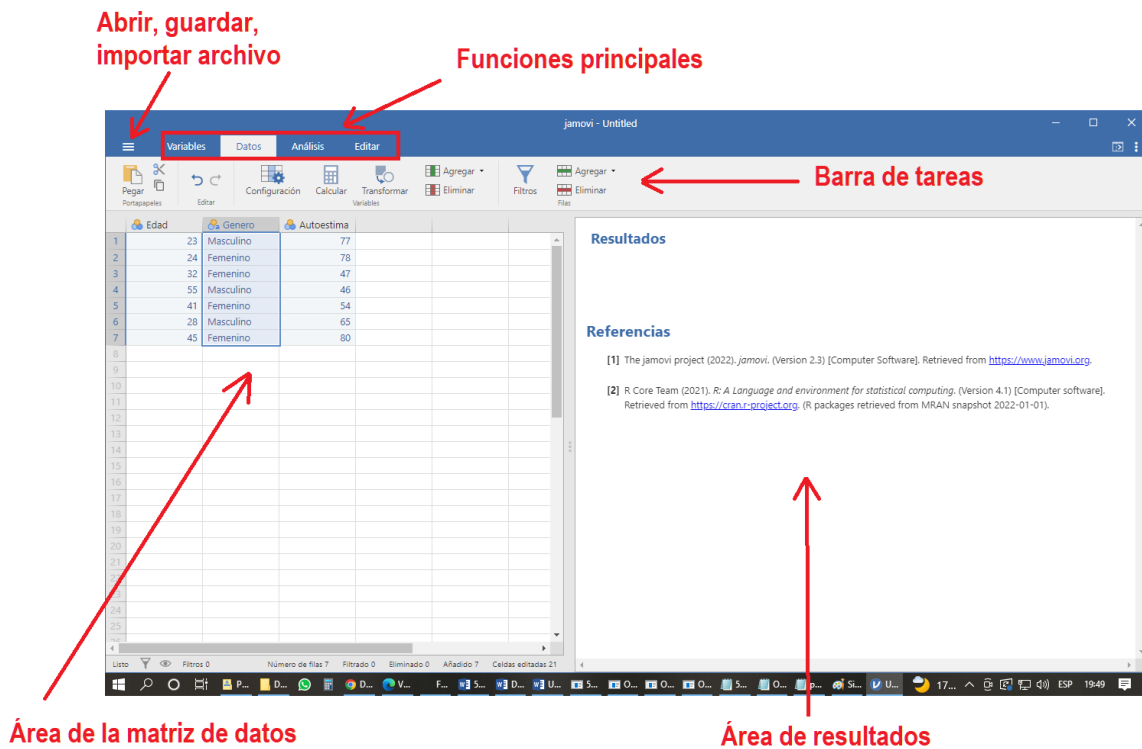


Figura 2: Captura de pantalla del programa estadístico Jamovi

Una vez instalado el software podrás ver, al abrir Jamovi, una imagen similar a la que mostramos arriba.

En la parte superior izquierda encontrarás las opciones de Abrir, Guardar, Importar y Exportar. A la derecha verás cuatro pestañas: Variables, Datos, Análisis y Editar; cada una de estas pestañas tiene una barra de tareas propia que te servirá para ejecutar todas las funciones que permite el programa. Finalmente, verás que la mayor proporción de la pantalla se divide en dos, a la izquierda podrás encontrar el área de la matriz de datos, las propiedades de las variables con las que trabajarás o las configuraciones necesarias para el procesamiento de los análisis que deseas calcular; a la derecha siempre estarán los resultados de los análisis que realices en formas de tablas o gráficos.

A lo largo de la asignatura, conforme vayamos avanzando te mostraremos imágenes y algunas explicaciones de los cálculos y resultados que obtengamos mediante Jamovi

Tabla de frecuencias, intervalos de clase; frecuencias absoluta, relativa y porcentual.

La **Tabla de frecuencias** permite mostrar los resultados de una forma organizada y facilitar análisis posteriores de mayor complejidad.

Es una tabla estadística de doble entrada donde se presentan los datos resumidos, de tal manera que se puede tener una visión general del comportamiento de la/s variable/s. En las filas se ubican los valores que ha obtenido la variable y en las columnas un conjunto de estadísticos de resumen; como, por ejemplo, **Frecuencia absoluta**, **Frecuencia relativa**, **Frecuencia absoluta acumulada**, **Frecuencia relativa acumulada**, **Porcentaje**, **Porcentaje acumulado**.

La **Frecuencia absoluta** indica el número de veces que una característica o valor se repite en una variable determinada. Se suele escribir como n_i , y la suma de todas las frecuencias absolutas debe coincidir con el total de elementos estudiados (N).

Por lo tanto, si deseamos calcularlo, lo que tenemos que hacer es ordenar los datos (de manera creciente cuando éstos son numéricos) y luego contar cuántas veces se repite cada uno de ellos.

Veamos un ejemplo, tengo la siguiente matriz de datos:

	Profesión	Edad	Género	EstadoCivil	Ciudadde...	¿Conqui...	Sobrehijos
1	Kinesiólogo/a	32	Femenino	Casado/a	Cordoba	Con su familia	Si tengo y viv.
2	Kinesiólogo/a	39	Femenino	Separado/a	Alta gracia	Con su familia	Si tengo y viv.
3	Kinesiólogo/a	37	Femenino	Comprometi...	Cordoba	Sólo/a	No tengo hijo
4	Otros	32	Femenino	Casado/a	parana entre ...	Con su familia	Si tengo y viv.
5	Kinesiólogo/a	48	Femenino	Soltero/a	Córdoba Capi...	Con su familia	Si tengo y viv.
6	Kinesiólogo/a	50	Femenino	Separado/a	Cordobs	Con su familia	Si tengo y viv.
7	Kinesiólogo/a	45	Femenino	Casado/a	Córdoba	Con su familia	Si tengo y viv.
8	Kinesiólogo/a	46	Femenino	Casado/a	Cordoba	Con su familia	Si tengo y viv.
9	Kinesiólogo/a	48	Femenino	Soltero/a	Córdoba Capi...	Con su familia	Si tengo y viv.
10	Kinesiólogo/a	49	Femenino	Casado/a	Córdoba . Ca...	Con su pareja	Si tengo y viv.
11	Bioquímico/a	50	Femenino	Casado/a	Ushuaia	Con su familia	Si tengo y viv.
12	Bioquímico/a	60	Femenino	Divorciado/a	La Punta	Con su familia	Si tengo y viv.
13	Médico/a	54	Femenino	Casado/a	San Luis	Con su pareja	Si tengo, per..
14	Bioquímico/a	31	Femenino	Casado/a	San Luis	Con su familia	Si tengo y viv.
15	Otros	50	Femenino	Casado/a	San Luis	Con su familia	Si tengo y viv.
16	Psicólogo/a	50	Femenino	Unión libre o ...	Trelew	Con su familia	Si tengo y viv.
17	Bioquímico/a	33	Femenino	Soltero/a	San Luis	Sólo/a	No tengo hijo
18	Bioquímico/a	49	Masculino	Separado/a	San Luis	Sólo/a	Si tengo, per..
19	Bioquímico/a	61	Femenino	Casado/a	Coronel Mold...	Con su familia	Si tengo, per..
20	Bioquímico/a	50	Femenino	Casado/a	Junín de los ...	Con su familia	Si tengo y viv.
21	Bioquímico/a	46	Femenino	Casado/a	San Luis	Con su familia	Si tengo y viv.
22	Bioquímico/a	46	Femenino	Casado/a	San Luis	Con su familia	Si tengo y viv.
23	Bioquímico/a	49	Femenino	Casado/a	Villa mercedes	Con su familia	Si tengo y viv.
24	Bioquímico/a	36	Femenino	Unión libre o ...	Villa Mercede...	Con su pareja	No tengo hijo
25	Bioquímico/a	33	Femenino	Casado/a	Loncopue	Con su familia	Si tengo y viv.

Figura 3: Matriz de datos para analizar

En esta matriz de datos se encuentra toda la información obtenida de una investigación que realizamos durante la pandemia y con la cual buscábamos indagar el efecto en el plano emocional de diversos profesionales de la salud. Se trata de un estudio realizado con 169 participantes y numerosas variables.

Ahora, para ir entendiendo los conceptos, analizaremos las Frecuencias absolutas en las variable *Género* (cualitativa) y *Edad* (cuantitativa).

Como habrás visto arriba, para calcular las frecuencias absolutas lo que debemos hacer es contar cuantas veces se repiten los valores iguales; por ejemplo, para la variable *Género* cuanto se repite el dato "Femenino" y cuanto se repite el dato "Masculino". A continuación, lo haremos con Jamovi...

Frecuencias de Género			
Género	Frecuencias	% del Total	% Acumulado
Femenino	145	85.8 %	85.8 %
Masculino	22	13.0 %	98.8 %
No binario	2	1.2 %	100.0 %

Figura 4: Captura de pantalla del análisis de frecuencia realizado mediante el programa estadístico Jamovi

Como puedes observar, el resultado del análisis nos informa que la muestra está constituida por 145 personas de género femenino, 22 de género masculino y 2 personas no binarias. Seguramente habrás observado que también se ofrecen los resultados de Porcentaje (% del total) y de Porcentaje acumulado (% acumulado)... Bueno, por ahora quedémonos sólo con las frecuencias acumuladas, en un instante te explicaremos lo demás.

Y, ¿qué pasará con la variable Edad?... veamos... le solicitamos a Jamovi que la analice y nos ofrece lo siguiente...

Tabla 1.

Análisis de frecuencias para la variable Edad

Edad	Frecuencias	% del Total	% Acumulado
25	5	3.0 %	3.0 %
27	1	0.6 %	3.6 %
28	1	0.6 %	4.1 %
29	3	1.8 %	5.9 %
30	6	3.6 %	9.5 %
31	2	1.2 %	10.7 %
32	6	3.6 %	14.2 %
33	8	4.7 %	18.9 %
34	7	4.1 %	23.1 %
35	5	3.0 %	26.0 %
36	9	5.3 %	31.4 %
37	6	3.6 %	34.9 %
38	4	2.4 %	37.3 %
39	3	1.8 %	39.1 %
40	9	5.3 %	44.4 %
41	2	1.2 %	45.6 %
42	5	3.0 %	48.5 %
43	8	4.7 %	53.3 %
44	2	1.2 %	54.4 %
45	8	4.7 %	59.2 %
46	8	4.7 %	63.9 %
47	1	0.6 %	64.5 %
48	5	3.0 %	67.5 %
49	4	2.4 %	69.8 %
50	12	7.1 %	76.9 %
51	3	1.8 %	78.7 %
52	4	2.4 %	81.1 %
53	7	4.1 %	85.2 %
54	7	4.1 %	89.3 %
55	1	0.6 %	89.9 %
56	4	2.4 %	92.3 %
57	1	0.6 %	92.9 %
59	1	0.6 %	93.5 %
60	2	1.2 %	94.7 %
61	2	1.2 %	95.9 %
62	3	1.8 %	97.6 %
63	2	1.2 %	98.8 %
66	1	0.6 %	99.4 %
69	1	0.6 %	100.0 %

Notarás que, al existir más variabilidad de datos en Edad, el análisis se hace más complejo; no obstante, podremos identificar que existieron 5 participantes que tenían 25 años, 1 participante que tenía 27 años, etc. etc. etc....

*Nota: en la unidad que veremos a continuación conocerás otras estrategias que podemos utilizar para analizar los datos de las variables cuantitativas y que brindarán información importante acerca de cómo se distribuyen los datos de una muestra.

Bueno, ahora le pediremos a Jamovi que nos haga algunos gráficos con ambas variables

Gráfico 1.

Frecuencias para la variable Género

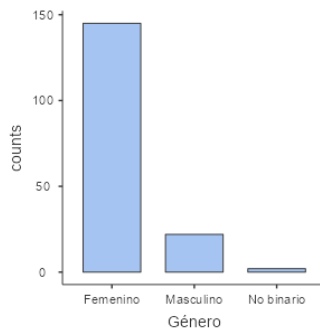
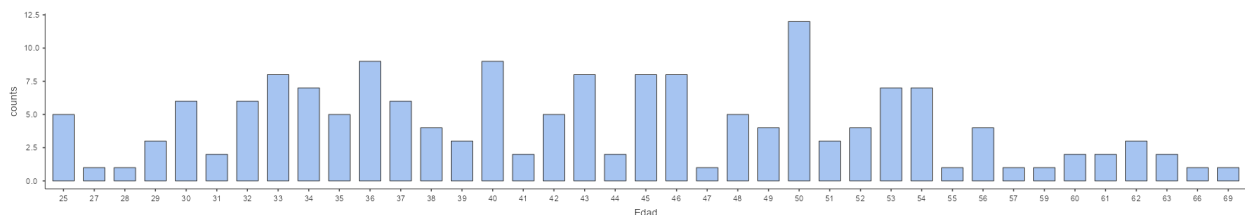


Gráfico 2.

Frecuencias para la variable Edad



Como podrás observar, con el análisis de Frecuencias absolutas ya tenemos información que nos permite describir nuestra muestra (o población, según el caso). ¿Te animas a intentarlo?... vamos!!! podrías decir que “atentos a los valores de frecuencia, la muestra está conformada mayoritariamente por personas de género femenino”, verdad?... muy bien... has hecho un gran progreso...

Ahora sigamos...

IMPORTANTE: siempre ten en cuenta el tipo de variable que estudiarás (cualitativa -nominal u ordinal- o cuantitativa) para elegir qué tipo de análisis puedes implementar. Presta especial atención siempre a esto.

Ahora, profundizaremos el conocimiento con otros estadísticos, pero para asimilarlos adecuadamente lo haremos pensando y calculándolo con lápiz y papel.

Por favor, presta atención al siguiente cuadro resumen. Léelo con cuidado y ve pensando cómo se calcula y cuál es su posible utilidad. Iniciaremos con la Frecuencia absoluta y luego iremos pensando los demás, ya que todos los otros son derivaciones del cálculo de frecuencias

Estadístico	Símbolo	¿Qué es? ¿Cómo se calcula?	¿Para qué sirve?
Frecuencia Absoluta	n_i	Es el número de veces que una característica o valor se repite en una variable. Se calcula ordenando los datos y luego sumando cada valor que se repite	Te informa el número de participantes que poseen una característica determinada
Frecuencia Absoluta Acumulada	N_i	Es la suma de las frecuencias absolutas de los valores menores o iguales a X_i . Acá vas sumando por orden todos los valores de la frecuencia absoluta hasta un determinado valor. Formula: $N = n_{i1} + n_{i2} \dots + n_{ix}$	Te dice cuántos elementos o participantes hay hasta un determinado valor de la variable.
Frecuencia Relativa	f_i	Es la proporción de la frecuencia absoluta. Se calcula dividiendo la frecuencia absoluta por el número total de elementos N .	Te informa la proporción de participantes que poseen una característica determinada

		<i>Formula: $f_i = n_i / N$</i>	
Frecuencia Relativa Acumulada	F_i	Es la <u>suma de las frecuencias relativas</u> . Se calcula sumando por orden todos los valores de la frecuencia relativa hasta un determinado valor. <i>Formula: $F_i = f_{i1} + f_{i2} + \dots + f_{ik}$</i>	Te informa la proporción de participantes que hay hasta un determinado valor de la variable.
Porcentaje	%	Es la <u>frecuencia relativa expresada en porcentajes (%)</u> . Se calcula tomando la frecuencia relativa y luego multiplicándola por 100. <i>Formula: $\% = f_i * 100$</i>	Te da el porcentaje de participantes que poseen una característica determinada
Porcentaje Acumulado	$\%_A$	Es el porcentaje de las frecuencias relativas acumuladas. Se calcula tomando la frecuencia relativa acumulada y luego multiplicándola por 100. <i>Formula: $\%_A = F_i * 100$</i>	Te da el porcentaje de participantes que hay hasta un determinado valor de la variable.

Cuadro resumen 1: Descripción y explicación de los estadísticos descriptivos de frecuencia absoluta, relativa y porcentaje.

***Nota:** tanto la Frecuencia Absoluta Acumulada, la Frecuencia Relativa Acumulada y el Porcentaje Acumulado tiene sentido calcularlo cuando la variable que estamos estudiando tiene al menos un nivel de medición ordinal. En una variable Nominal la información que da, carece de sentido.

Vamos a tomar un ejemplo que tenga en cuenta la variable Edad.

En la tabla de frecuencias que verás abajo, notarás que ya se ha calculado la Frecuencia absoluta, ¿te animas a completar los cálculos de los otros estadísticos? Inténtalo!!! va a ser una linda forma de comprenderlos. La solución te la mostramos al finalizar este documento ;)

Tabla 2.

Tabla de frecuencias para la variable Edad (Práctica).

Edad						
	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada	Porcentaje	Porcentaje Acumulado
12	1	0,0625	1	0,0625	6,25	6,25
13	1	0,0625	2	0,125	6,25	12,5
14	2	0,125	4			
16	1					
17	1					
21	2					
23	2					
27	1					
32	1					
41	1					
44	2					
45	1					
Total	16					

Intervalos de clase

Los intervalos de clase son particularmente útiles cuando las variables poseen muchos valores, se los suele utilizar como resúmenes de los datos. Proporcionan una mejor posibilidad de comunicar el patrón que poseen los datos y también facilitan, luego, los análisis más complejos de los mismos.

Vamos con un ejemplo: Supongamos que estas estudiando, en un centro de salud, la cantidad de *semanas que llevan de tratamiento* los pacientes con algún problema psicomotriz. Reúnes un conjunto de pacientes y obtiene los siguientes datos: 3, 15, 24, 28, 33, 35, 38, 42, 43, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13 (semanas).

Como puedes ver, esta información se muestra confusa, extensa y poco útil, por lo que podemos pensar en agrupar los datos en clases.

¿Cómo lo hacemos?

1. Lo primero que debes hacer es identificar el valor mínimo y máximo de la distribución. En este caso son 3 y 48.
2. Luego debes tomar el valor máximo y restarle el valor mínimo ($48 - 3 = 45$). Con eso obtendremos el rango o el recorrido
3. A continuación, si hace falta, debemos aumentar el número obtenido para que cumpla dos condiciones: a) que sea un número entero y no tenga decimales y, b) que sea divisible por el número de intervalos que deseamos hacer.
4. Una vez que hagamos eso, el número obtenido lo dividiremos en la cantidad de intervalos que deseemos establecer.
5. Posteriormente se forman los intervalos teniendo presente que el límite inferior de una clase pertenece al intervalo, pero el límite superior no pertenece al intervalo.
6. Y finalmente, construimos una tabla de frecuencias de clase.

Siguiendo el ejemplo inicial, si deseamos que nos queden agrupados los datos en 5 intervalos, tomaremos el 45 ($48 - 3 = 45$) (como es un número entero y éste es divisible por 5, no le sumaremos nada) y lo dividiremos por 5, lo que nos dará como resultado 9 (este será la cantidad de elementos que se incluirá en cada intervalo, por lo que esa cantidad le tendremos que sumar al valor mínimo de cada clase).

Bien, empezamos con el primer intervalo. El valor mínimo que tiene la variable Edad es 3 y le sumaremos 9 lo que nos dará 12, ¿cierto?... muy bien! Entonces el primer intervalo incluirá será [3-11]... recuerda lo que señalamos arriba "el límite inferior de una clase pertenece al intervalo, pero el límite superior no pertenece al intervalo" ¿qué quiere decir esto?, que cuando computes los datos tendrás que considerar que ese intervalo incluirá todos los valores MENORES a 12...

El segundo intervalo incluirá desde los 12 años, hasta los 20 y así sucesivamente....

*Nota: suele ser confuso esto del límite superior y límite inferior del intervalo de clase. Despeja tus dudas con los profes si necesitas ayuda

Tabla 3.

Tabla de frecuencias para los Intervalos de clase de la variable Edad.

Intervalo	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada	Porcentaje	Porcentaje Acumulado
[3-12]	3	0,075	3	0,075	7,5	7,5
[12-21]	6	0,15	9	0,225	15	22,5
[21-30]	8	0,2	17	0,425	20	42,5
[30-39]	15	0,375	32	0,8	37,5	80
[39-48]	8	0,2	40	1	20	100
Total:	40				100	

¿Puedes analizar la información que presenta esta tabla? Intenta sacar las conclusiones más importantes y consúltalas con los profes ;)

Representación gráfica de una distribución de datos

Si bien ya hemos visto algunos gráficos a modo de ejemplos, tienes que tener en cuenta que la representación gráfica de una distribución de frecuencias depende del tipo de datos que la constituya y la finalidad, es decir, lo que nosotros queramos poner en evidencia.

Los gráficos, son siempre un excelente complemento de las tablas y ofrecen un análisis rápido, directo y comprensible para un conjunto amplio de personas, independientemente de su grado de conocimiento, porque ofrecen una lectura inicialmente intuitiva.

Como verás abajo hay distintos tipos de gráficos y, como dijimos, cada uno de ellos resulta conveniente para un propósito. Por ejemplo, si deseamos graficar la proporción de participantes, probablemente usaremos un gráfico de sectores, mientras que, si deseamos graficar las frecuencias de edad, usaremos diagramas de barra; y si deseamos graficar la altura de los participantes, usaremos un histograma.

Tipos de representaciones gráficas

Las representaciones gráficas son muy variadas, hay diversos tipos y, a su vez, distintos estilos de presentación. A continuación, te ofreceremos sólo tres de ellos, que son los que más frecuentemente los que se asocian a los análisis que aprenderemos a lo largo del desarrollo de esta asignatura.

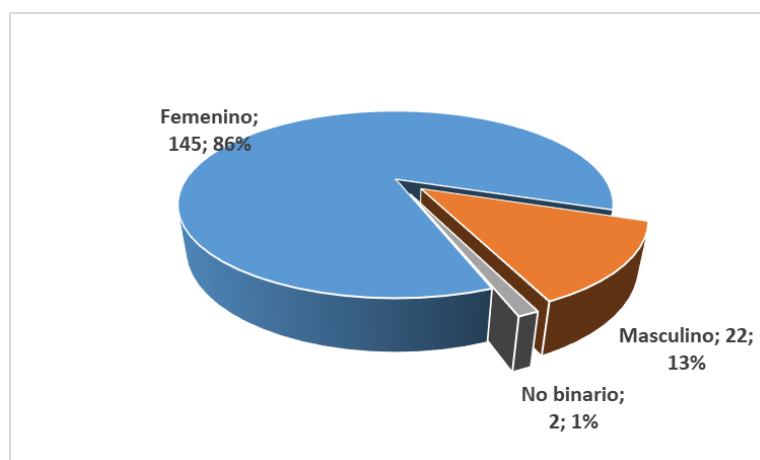
- **Gráficos de sectores:** circulares o de torta, dividen un círculo en porciones proporcionales según el valor de las frecuencias relativas. Para entenderlo puedes pensar en una pizza o un bizcochuelo. La totalidad de la torta representa el 100%, es decir la totalidad de los datos que se han considerado al momento de realizar el gráfico. Una porción de esa torta, será una parte de esa totalidad, por lo que cuando lo observas la figura verás las distintas categorías de los datos y el valor del porcentaje y/o las frecuencias de las mismas.
- **Diagramas de barras:** muestran los valores de las frecuencias absolutas sobre un sistema de ejes cartesianos. Se utiliza cuando la variable es cualitativa o numérica discreta (nivel de medición). Cada barra se representa separada una de otras para dar cuenta justamente de la naturaleza de la variable.
- **Histogramas:** es una forma especial del diagrama de barras que se usa cuando la variable presenta una distribución numérica continua (variables cuantitativas). Cada barra se representa junta una a otra, para dar cuenta de la naturaleza de la variable.

***Nota: En el caso de los Gráficos o diagramas de barras e Histogramas, en el eje horizontal (eje X) se suelen mostrar las variables que estamos midiendo, mientras que en el eje vertical (eje Y) se indica el número de observaciones que toma la variable representada.**

Tomaremos como ejemplo la matriz de datos de más arriba para graficar algunas variables, lo haremos mediante Excel:

Gráfico 3.

Frecuencia y porcentaje para la variable Género



***Nota: observa cómo se presentan las imágenes (son simples, no estás sobrecargadas de información o de colores) y donde se colocan los títulos (tienen un formato específico según las normas APA)**

BIBLIOGRAFÍA

- Gorgas-García, J., Cardiel-López, N. & Zamorano-Calvo, J. (2011). *Estadística básica para estudiantes de ciencias*. Departamento de Astrofísica y Ciencias de la Atmosfera Facultad de Ciencias Físicas Universidad Complutense de Madrid. Recuperado el 22/2/19 de http://webs.ucm.es/info/Astrof/users/jaz/ESTADISTICA/libro_GCZ2009.pdf
- Salazar, C. & Castillo, S. (2018). *Fundamentos básicos de estadística*. Recuperado el 22/2/19 de <http://www.dspace.uce.edu.ec/bitstream/25000/13720/3/Fundamentos%20B%C3%A1sicos%20de%20Estad%C3%ADstica-Libro.pdf>
- The jamovi project (2022). *Jamovi (Version 2.3.13) [Computer Software]*. <https://www.jamovi.org>.

SIMULACIONES

Las simulaciones te permitirán comprender en mayor profundidad los detalles en la medida que vayas modificando los valores. ¡Atrévete a jugar y reforzar tus conocimientos!!!

Percentiles: <https://www.bioestadistica.uma.es/analisis/percentil/>

SOLUCIONES

Ejercicio 1

Tabla 4.

Tabla de frecuencias para la variable Edad (Respuestas).

Edad						
	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada	Porcentaje	Porcentaje Acumulado
12	1	0,0625	1	0,0625	6,25	6,25
13	1	0,0625	2	0,125	6,25	12,5
14	2	0,125	4	0,25	12,5	25
16	1	0,0625	5	0,3125	6,25	31,25
17	1	0,0625	6	0,375	6,25	37,5
21	2	0,125	8	0,5	12,5	50
23	2	0,125	10	0,625	12,5	62,5
27	1	0,0625	11	0,6875	6,25	68,75
32	1	0,0625	12	0,75	6,25	75
41	1	0,0625	13	0,8125	6,25	81,25
44	2	0,125	15	0,9375	12,5	93,75
45	1	0,0625	16	1	6,25	100
Total	16				100	



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)