9. Programas informáticos para análisis estadístico de datos cuantitativos

El análisis estadístico y la investigación cuantitativa no serían tan populares sin los programas informáticos. La función de los programas informáticos en la investigación es ejecutar análisis con cantidades de datos considerables, por lo que son imprescindibles en la investigación de datos cuantitativos.

Los principales programas informáticos para el análisis estadístico de datos cuantitativos y usados en ciencias sociales son: Excel, SPSS, PSPP, Stata y R, entre muchos otros. Cada uno tiene sus ventajas y desventajas, y la decisión de usar uno u otro depende de los recursos disponibles y la familiaridad del investigador con programación.

En este capítulo presentamos el software PSPP que fue creado a imagen y semejanza de SPSS pero con código abierto. SPSS es el software más popular en ciencias sociales debido a su alta usabilidad y enorme capacidad de análisis, pero con un coste de licencia elevado, solo al acceso de universidades o empresas. Por ello, PSPP fue creado mediante software libre que permite la distribución y acceso libre. Véase que PSPP son las siglas invertidas de SPSS.

Descarga del software PSPP: https://www.gnu.org/software/pspp/

Manual de instalación de PSPP en español:
http://www2.uned.es/socioestadistica/Practicas_%20ejercicios_guia/Practica2.pdf

Tutorial de procesamiento de datos en PSPP:
http://www2.uned.es/psiped-psicologia-social/ecodilema/acceso/orientaciones_analisis.pdf

En Youtube existen decenas de tutoriales de cada una de las funciones y análisis estadísticos de PSPP. Puede consultarlos si desea conocer más sobre este software libre.

El programa informático ejecuta las órdenes del investigador, por lo que se debe tener claro qué análisis corresponde en cada momento según el tipo y número de variables, como veremos en el próximo capítulo.

Bibliografía sugerida

Salmerón Gómez, Romelio (2015) (Mini)Manual de PSPP, alternativa libre a SPSS. Disponible en: https://softtcm.files.wordpress.com/2014/04/pspp.pdf

Ejercicio

Descargue las bases de datos de la Encuesta Mundial de Valores y Latinobarómetro en formato ASCI, SPSS, Stata o R. Revise cómo está guardada la información de los casos y las variables.

10. Análisis de datos: tipos de análisis estadísticos

La estadística es una ciencia de las matemáticas que se usa en la investigación para analizar datos cuantitativos.

El análisis de los datos depende de dos grandes factores:

A. el tipo de variables que analizamos
B. el número de variables que analizamos simultáneamente

Las variables equivalen a las preguntas del cuestionario o los indicadores identificados

A. Tipos de variables

Hay 3 grandes tipos de variables según las categorías de respuesta:

- Nominales: son aquellas variables cuyas categorías de repuesta no tienen un orden preestablecido ni jerarquía interna. Es decir, las categorías de respuesta pueden variar de orden y mantienen su sentido. Ejemplos:
  - Variable Género: 1- Masculino, 2-Femenino
  - Variable Estado civil: 1-Soltero, 2-Casado, 3-Vive en pareja, 4-Divorciado, 5-Viudo, 6-Otros
- Ordinales: son aquellas variables cuyas categorías de respuesta sí tienen un orden preestablecido o jerarquía interna. Las categorías de repuesta deben mantener un orden. Ejemplos:
  - Variable Nivel educativo: 1-Sin estudios, 2-Estudios primarios, 3-Estudios secundarios, 4-Estudios universitarios
  - Variable Grado de conformidad con el presidente: 1-Totalmente de acuerdo, 2-De acuerdo, 3-Ni de acuerdo ni en desacuerdo, 4-En desacuerdo, 5-Totalmente en desacuerdo
  - Variable grupo de edad: 1-Menos de 18 años, 2-Entre 18 y 35 años, 3-Entre 36 y 50 años, 4-Entre 51 y 65 años, 5-Más de 65 años
- Escalar: son aquellas variables cuyas categorías de respuestas sí tienen un orden preestablecido o jerarquía interna, y el espacio entre una categoría y otra es el mismo.
  - Variable Año de Nacimiento (pregunta abierta):
  - Variable Número de veces que ha asistido al cine el último mes (pregunta abierta):

Cuando la edad se pregunta de forma abierta es una variable escalar, cuando se pregunta por intervalos es una variable ordinal, ya que, al preguntarlo de forma abierta, el espacio que hay entre 18 y 19 años (365 días) es el mismo que hay entre 35 y 36 años (también 365 días), y entre 65 y 66 años (también 365 días). Todas las variables o preguntas que hacen referencia a cantidades y no están preguntadas en forma de intervalos son variables escalares.

B. Número de variables a analizar

Según el número de variables a analizar simultáneamente, hay 3 grandes tipos de análisis:

- Análisis descriptivo univaridado: se analiza una sola variable, y este tipo de análisis tiene una finalidad netamente descriptiva, es decir, exponer los resultados de una variable.

- Análisis bivariado: se analizan dos variables, y este tipo de análisis tiene una finalidad de comprobar hipótesis relacionales (causales o asociativas), es decir, se comprueba si dos variables están relacionadas o no.
- Análisis multivariado (o multivariante): se analizan más de dos variables, y este tipo de análisis tiene una finalidad de comprobar hipótesis relacionales (causales o asociativas), es decir, se comprueba si varias variables están relacionadas entre sí.

Técnicas estadísticas

Las principales técnicas de análisis estadístico en cada uno de los tipos de análisis son:

Tabla 5. Técnicas estadísticas principales

Tipo de análisis	Técnicas estadísticas
Análisis univariado	▪ Tablas de frecuencias ▪ Media, desviación típica
Análisis bivariado	▪ Tablas de contingencia ▪ Correlaciones bivariadas ▪ ANOVA de un factor
Análisis multivariado	▪ Regresión lineal múltiple ▪ Regresión logística (binaria y multiple) ▪ Modelos logit y probit ▪ Análisis discriminante ▪ Análisis factorial ▪ Análisis de clústers ▪ Escalado multidimensional ▪ ANOVA de dos factores ▪ MANOVA ▪ Ecuaciones estructurales

Fuente: elaboración propia

En los siguientes capítulos se explicará cómo usar las técnicas estadísticas para llevar a cabo estos tipos de análisis. La estadística permite describir y hacer inferencias. La estadística descriptiva busca sintetizar y visualizar los resultados de estudio. La estadística inferencial buscar sacar conclusiones o predicciones a partir del análisis de los datos obtenidos en un estudio.

Bibliografía sugerida

Estadística para torpes. Disponible en:
https://unidadinvestigacionhvn.files.wordpress.com/2010/11/estadistica-para-torpes.pdf

Ejercicio

Seleccione diversos artículos de investigación relacionados con su pregunta de investigación o su tema de interés. Compruebe qué técnicas estadísticas emplean para comprobar las hipótesis o responder a la pregunta de investigación.

11. Análisis univariado descriptivo

Los estadísticos usados para analizar una sola variable con la finalidad de describir los datos recolectados dependen del tipo variable: nominal, ordinal o escalar.

Si la variable que deseamos analizar es *nominal*, se usará una tabla de frecuencias y un gráfico circular o de barras.

Si la variable que deseamos analizar es *ordinal*, se usará una tabla de frecuencias y un gráfico de barras

Si variable que deseamos analizar es *escalar u ordinal de más de cinco categorías*, se usarán los estadísticos de media (promedio) y desviación estándar (típica), y un histograma como gráfico. Ojo, no se recomienda analizar los datos mediante una tabla de frecuencias cuando la variable es de tipo escalar.

Tabla de frecuencia

La tabla de frecuencias es un análisis donde se expresan las respuestas de los datos recolectados en la cantidad de veces que ocurren (frecuencia absoluta) y el porcentaje que representan estas respuestas (frecuencia relativa). La frecuencia relativa (o porcentaje de respuestas) se calcula dividiendo el número de respuestas de una categoría entre el total de casos.

Ejemplo de análisis descriptivo univariado mediante tabla de frecuencias y gráfico de barras

Variable: *"¿Alguna vez en su vida ha usado Ud. correo electrónico o se ha conectado a Internet?"*

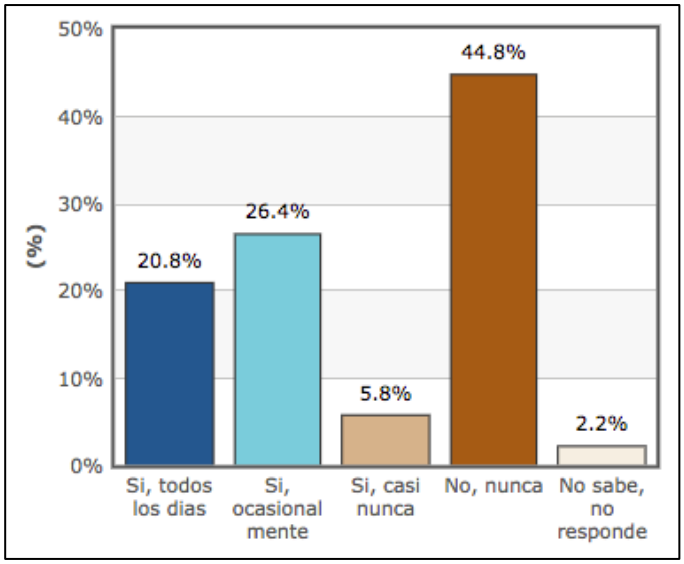
Casos: personas habitantes en Perú

Tabla 6. Ejemplo de tabla de frecuencias

	Frecuencia absoluta (número de casos)	Frecuencia relativa (porcentaje)
Sí, todos los días	250	20,8%
Sí, ocasionalmente	317	26,4%
Sí, casi nunca	70	5,8%
No, nunca	537	44,8%
No sabe, no contesta	26	2,2%
N (total de casos)	1200	100%

Fuente: elaboración propia a partir de Latinobarómetro (2015)

Figura 5. Ejemplo de gráfico de barras



Fuente: elaboración propia a partir de Latinobarómetro (2015)

Todo análisis debe ir acompañado de un texto donde se interpreten los resultados, es decir, se destaquen los principales valores o categorías de respuesta y se haga una lectura sobre ello. Ejemplo:

> La mayoría de encuestados en Perú (44,8%) nunca ha usado correo electrónico o se ha conectado a Internet en 2015. Sorprende este resultado en la era digital, lo que manifiesta la baja extensión de las tecnologías de la información en Perú.

Media (promedio)

La media (también denominada popularmente promedio o en lenguaje estadístico, media aritmética) es el valor que representa y sintetiza un conjunto de datos. Es una medida de la tendencia central de una variable. Se calcula a partir de la suma de los valores de las respuestas dividido por el número total de casos.

Desviación estándar

La desviación estándar (también denominada desviación típica) es el valor que mide la dispersión de valores o respuestas de una variable. Es el promedio de las distancias de los valores de cada caso respecto a la media. Es decir, se calcula una media de cuánto se alejan las respuestas de cada actor respecto a la media. Cuanto más alta sea la desviación estándar más dispersas son las respuestas obtenidas, es decir, más heterogénea es la opinión o comportamiento de los actores analizados.

La desviación estándar al cuadrado es la varianza. Tanto la desviación estándar como la varianza son usados ampliamente en los análisis multivaridos.

El histograma es el gráfico usado para visualizar los resultados de una variable escalar, y se caracteriza por incluir la curva de normalidad. La curva normal, también conocida como campana de Gauss, es una representación de la distribución de los datos siguiendo

Ejemplo de análisis descriptivo univariado mediante media y desviación estándar.

Variable: *Con una escala de 1 a 10, le pedimos evaluar cuán democrático su país. El "1" quiere decir que "su país no es democrático" y el "10" quiere decir que "su país es totalmente democrático" ¿Dónde pondría Ud. a su país?*

Se realizó esta pregunta en tres países diferentes. Debido a que es una variable ordinal de más de 5 categorías se asemeja a una variable escalar y se calcula la media y desviación estándar.

Resultados:

Tabla 7. Ejemplo de tabla comparación de medias y desviación estándar

País	Media	Desviación estándar
Perú	5,26	2,05
Chile	5,78	1,98
México	5,00	2,46

Fuente: elaboración propia a partir de Latinobarómetro (2015)

La interpretación de estos resultados sería:

> Según las opiniones de los encuestados en sus respectivos países, el país más democrático es Chile (media = 5,78), seguido de Perú (media = 5,26), y el menos democrático de los tres analizados es México (media = 5,00). Aunque en México es donde más divergen las opiniones (desviación estándar = 2,46).

Para decir que una media y una desviación estándar son altas, media o bajas debemos comparar los resultados entre los grupos de otra variable. En el ejemplo anterior, hemos comparados los resultados de la pregunta sobre "qué tan democrático es su país" en función del país. Para saber si estas diferencias son significativas llevaremos a cabo técnicas más complejas como ANOVA de un factor, que veremos en capítulos siguientes.

En conclusión, el tipo de análisis para describir una variable depende del tipo de variable

Tabla 8. Tipo de variables y análisis descriptivo univariado

Variables	Tabla de frecuencias	Medida de tendencia central	Medidas de dispersión	Gráficos
Nominal	Sí	*Moda**	*Ratio de variación**	Sectores o barras
Ordinal	Sí	*Mediana**	*Rango**	Barras
Escalar **	No	Media	Desviación típica	Histograma

* Tienen un uso limitado, por lo que no los presentamos ya que no aportan mucho a la descripción de la variable. ** Las ordinales de más de 5 categorías pueden ser tratadas como variables escalares.

Fuente: elaboración propia a partir de Latinobarómetro (2015)

Ejercicio

Revise la Encuesta Mundial de Valores en su opción de "Online Analysis". Escoja una sección o tema y analice todas las variables de esa sección. Presente los resultados del análisis descriptivo univariado para cada variable. Tenga en cuenta el tipo de variable para seleccionar el tipo de análisis (tabla de frecuencias, media y desviación estándar) y el tipo de gráficos (sectores, barras o histograma).

12. Cómo hacer y analizar tablas de contingencia

Las tablas de contingencia (también llamadas a veces tablas dinámicas, tablas cruzadas, tablas de control o *crosstabs* en inglés) son posiblemente la técnica estadística más utilizada en análisis de datos. Las tablas de contingencia son una forma de presentar los datos de dos variables, y pueden ser usadas para analizar la asociación entre dos variables mediante las frecuencias relativas (porcentajes).

Condiciones para usar tablas de contingencia:

- Solo se pueden relacionar dos variables, por ello son una técnica de análisis bivariado.[3]
- Esta técnica de análisis se usa especialmente con variables nominales y ordinales. Las variables nominales son las que no tienen orden interno establecido (p.ej. género o estado civil), y las variables ordinales son aquellas que sí tienen un orden interno establecido y el paso de una categoría a otra no es igual (p.ej. nivel educativo (sin estudios-primarios-secundarios-universitarios) o interés en la política (mucho interés-algo de interés-no muy interesado-nada interesado)). Las tablas de contingencia no se usan para analizar relaciones de variables escalares como la edad ya que si se usara la tabla sería inmensa e ilegible. Si queremos usar la edad como variable en una tabla de contingencia debemos recodificarla por rangos (p.ej. 18-35 años, 36-64 años, más de 64 años). Al recodificar una variable escalar como la edad por rangos, deja de ser escalar y pasa a ser ordinal, y por tanto sí se puede incluir un análisis de tablas de contingencia. Ejemplo:

Tabla 9. Tabla de contingencia: interés en la política según grupos de edad

		Grupos de edad			Total
		< 35	35-65	> 65	
Interés en la política	Muy interesado	53 10,0%	225 21,9%	148 30,5%	426 20,8%
	Algo interesado	218 41,1%	445 43,3%	188 38,7%	851 41,6%
	No muy interesado	160 30,2%	278 27,0%	130 26,7%	568 27,8%
	Nada Interesado	99 18,7%	80 7,8%	20 4,1%	199 9,7%
	Total	530 100,0%	1028 100,0%	486 100,0%	2044 100,0%

Fuente: elaboración propia a partir de Encuesta Mundial de Valores

[3] También se pueden usar las tablas de contingencia para analizar tres variables, pero son poco empleadas.

Qué variable en filas y qué variable en columnas

Debido a que se estudia una variable en función de otra, el investigador ha de distinguir entre la variable dependiente (o a explicar) y la variable independiente (o explicativa). Esta distinción entre variable independiente y dependiente es importante porque la variable independiente se sitúa en columnas, y la variable dependiente en filas. El investigador es el que decide cuál variable es independiente (o explicativa) y cuál dependiente (o a explicar). Si nuestra hipótesis es causal o explicativa, la variable independiente (o explicativa) se coloca en columnas, y la variable dependiente (o a explicar) en filas. Si nuestra hipótesis es asociativa (y no causal o explicativa), el investigador debe decidir qué variable se coloca en filas y cuál en columnas. Las tablas de contingencia permiten comprobar hipótesis asociativas y causales (o explicativas). Aunque para afirmar causalidad deberían llevarse a cabo análisis más avanzados como tablas de contingencia de control (con 3 variables) o regresiones.

Las tablas de contingencia están compuestas de las categorías de respuesta de la variable en filas, de las categorías de respuesta de la variable en columnas, y el valor de las casillas. El valor de las casillas corresponde a la cantidad de casos que se adhieren a las respectivas categorías, y el porcentaje que representan estos casos.

Cómo se comprueba en una tabla de contingencia si dos variables están relacionadas

Dos variables están relacionadas (asociadas) si la distribución de frecuencias de una variable (la que se coloca en filas) es diferente según las categorías de respuesta de la otra variable (la que se coloca en columnas). Es decir, si los resultados de una variable son diferentes en las categorías de respuesta de la otra variable es que las dos variables están relacionadas. En cambio, si los resultados de una variable son similares en las categorías de respuesta de la otra variable es que las variables son independientes (no están asociadas).

Para poder llevar a cabo esta comprobación es necesario que se calculen los porcentajes por columnas. Los porcentajes por columnas se obtienen de dividir el valor de una celda por el valor total de la columna. Por ejemplo, en la anterior tabla de contingencia donde se cruzan los datos de “interés en la política” y “grupos de edad”, hay 53 personas son menores de 35 años y están muy interesados en la política. Para calcular qué porcentaje representan, se divide 53 entre 530, que son el total de personas menores de 35 años. Así se obtiene el 10%, que representa el total de menos de 35 años que están muy interesados en la política. En la casilla de abajo, el 41,1% es el total de menores de 35 años que están algo interesados en la política, y se ha obtenido de dividir 218 (personas menores de 35 años algo interesadas en la política) entre 530 (total personas menores de 35 años). Este cálculo lo ejecutan los diversos programas informáticos, aunque siempre es el investigador el que debe indicar qué porcentaje se calcula y ser consciente de los datos que está interpretando.

Recomendación: las tablas de contingencia se leen fila por fila, y de derecha a izquierda. Esta es la mejor manera de poder averiguar si los valores de la variable en filas, se repiten por igual o de manera diferente en las categorías de respuesta de la variable en columnas. Si los valores (porcentajes) son muy diferentes es que la asociación entre las variables es fuerte. Si los valores son un poco diferentes es que la asociación entre las variables es débil. Y si los valores son muy similares o iguales es que no hay asociación entre variables. Se recomienda hacer una lectura fila por fila, y posteriormente una interpretación global de la asociación entre las dos variables.

Ejemplo en 3 pasos:

1. Hipótesis de partida y selección de las dos variables

Quiero analizar si la creencia en Dios explica el interés en la política. Mi hipótesis de partida es que las personas creyentes en Dios tienen más interés en la política. Por tanto, tengo dos variables a relacionar: "interés en la política" y "creencia en Dios".

- La variable “interés en la política” fue preguntada en cuatro categorías de respuesta: 1-muy interesado, 2-algo interesado, 3-poco interesado, 4-nada interesado.
- La variable “creencia en Dios” tiene dos categorías: 1-sí creo, 2-no creo.

Voy a probar esta hipótesis usando la Encuesta Mundial de Valores realizada en Alemania en 2013. Antes de realizar la tabla de contingencia, estas son las tablas de frecuencias de cada variable por separado, sin haberlas cruzado o relacionado entre sí todavía.

Tabla 10. Tabla de frecuencias: Interés en la política

	Frecuencia	Porcentaje
Muy interesado	426	20,8%
Algo interesado	852	41,6%
No muy interesado	568	27,8%
Nada Interesado	199	9,7%
Total	2045	100,0%

Tabla 11. Tabla de frecuencias: Creencia en Dios

	Frecuencia	Porcentaje
Sí creo	1286	65,1%
No creo	690	34,9%
Total	1976	100,0%

2. Construcción de la tabla de contingencia

Según mi hipótesis inicial es la creencia en Dios lo que explica el interés en la política, por tanto, la variable “creencia en Dios” será la variable independiente (o explicativa) y la variable “interés en la política” será la variable dependiente (o a explicar). La variable creencia en Dios irá en columnas y la variable interés en la política en filas.

Muy importante, al relacionar las dos variables en una tabla de contingencia, se calculan los porcentajes por columnas. ¿Cómo se calculan estos porcentajes por columnas? Normalmente estos cálculos los hace el programa informático con el que estemos

trabajando. Se detalla nuevamente cómo se calculan los porcentajes por columnas para que se pueda entender mejor el proceso y el resultado.

- Al construir una tabla de contingencia, en cada celda se coloca el total de casos que cumplen las categorías donde se cruzan. Por ejemplo, en la primera casilla, 283 es el número de personas que manifestaron que "sí creen en Dios" y además que dijeron tener "mucho interés en la política".
- Para calcular los porcentajes por columnas, se divide el número de casos de cada casilla entre el total de casos de la columna. Por ejemplo, se divide el número de personas que "sí creen en Dios" y tienen "mucho interés en la política" entre el total de personas que "sí creen en Dios", y se multiplica por 100 para expresarlo en porcentaje. En nuestro ejemplo: (283 / 1286)*100 = 22,0%
- Seguidamente se calcula el porcentaje de todas las otras casillas. Por ejemplo, el número de personas que "no creen en Dios" y tienen "mucho interés en la política" se divide entre el total de personas que "no creen en Dios". En nuestro ejemplo: (122 / 691) * 100 = 17,7%. Y así con todas las casillas de la tabla.

Tras calcular los porcentajes por columnas, el resultado es la siguiente tabla de contingencia.

Tabla 12. Tabla de contingencia: Interés en la política según creencia en Dios

		Creencia en Dios		Total
		Sí creo	No creo	
Interés en la política	Muy interesado	283 22,0%	122 17,7%	405 20,5%
	Algo interesado	528 41,1%	306 44,3%	834 42,2%
	No muy interesado	356 27,7%	190 27,5%	546 27,6%
	Nada Interesado	119 9,3%	73 10,6%	192 9,7%
	Total	1286 100,0%	691 100,0%	1997 100,0%

Fuente: elaboración propia a partir de Encuesta Mundial de Valores

Una vez que la tabla de contingencia que relaciona dos variables está expresada en porcentajes ya podemos pasar a su lectura e interpretación. Si no están calculados los porcentajes, NO se puede leer la tabla de forma correcta ya que el número de casos no es igual en cada columna. Se deben calcular y presentar los porcentajes por columnas para analizar una tabla de contingencia.

4. Lectura e interpretación de la tabla de contingencia

Recordemos que la tabla de contingencia se lee fila por fila, y de derecha a izquierda. Siguiendo con el ejemplo anterior, presentamos la manera correcta de describir esta tabla e interpretar los resultados.

El 20,5% de los encuestados en Alemania están muy interesados en la política. Este porcentaje es mayor en las personas que sí creen en Dios (22% de los que sí creen en Dios están muy interesados en la política, frente al 17,7 de los que no creen en Dios). Dijeron estar algo interesados en la política, el 42,2% de los encuestados, y este porcentaje es mayor para los que no creen en Dios. El 27,6 de los encuestados manifestaron estar poco interesados en la política, siendo este porcentaje casi igual para creyentes y no-creyentes en Dios (27,7 % y 27,5% respectivamente). Finalmente el 9,7% de los encuestados afirmaron no estar nada interesados en la política, siendo este porcentaje levemente mayor en los no-creyentes en Dios (9,3% de las personas que sí creen en Dios no están nada interesadas en política, frente al 10,6% de las que no creen en Dios). Por tanto, la creencia en Dios sí explica el interés en la política, aunque es una relación débil ya que las diferencias entre creyentes y no-creyentes en Dios, solo son medianamente considerables en la categoría de los muy interesados. En el resto de categorías del interés en la política, las diferencias entre los que sí creen en Dios y no creen en Dios, son pequeñas o inexistentes. Estos resultados nos ayudan a entender que el interés por la política no depende tanto de cuestiones de fe religiosa, y que el debate religioso no tiene mucha influencia sobre la movilización política. Futuros análisis deberían profundizar en otros aspectos para entender el interés en la política, tal vez el nivel educativo o el nivel de ingresos.

No hay una medida estándar para señalar si la diferencia entre porcentajes es alta o baja ya que las diferencias dependen del tamaño de la muestra y el número de categorías. Por ello, existen una serie de estadísticos denominados de contraste que sintetizan en un solo número si las dos variables están asociadas. Los estadísticos de contraste comprueban si las diferencias observadas para cada par de categorías de respuesta (casillas) difieren de las esperadas en caso de que las variables fueran independientes. El principal estadístico de contraste es chi-cuadrado de Pearson. Cuando la significación de chi-cuadrado es menor de 0,05 es que las dos variables están asociadas.

5. Conclusiones:

a. La tabla de contingencia es una técnica de análisis bivariado ya que relaciona dos variables y trata de averiguar si una variable explica la otra.
b. Las tablas de contingencia permiten relacionar variables nominales y ordinales, pero no variables escalares (a no ser que las recodifiquemos por rangos).
c. Hay que distinguir entre la variable que quiero explicar (variable dependiente) que irá en filas, y la variable explicativa (o independiente) que irá en columnas.
d. Se debe calcular el porcentaje por columna para poder leer la tabla de contingencia.
e. La tabla se lee fila por fila y de derecha a izquierda.
f. Lo importante es averiguar si los porcentajes de la variable a explicar (la que va en filas) se diferencian mucho, poco o nada entre las categorías de la variable explicativa (la que va en columnas). Si hay altas diferencias de porcentajes las dos variables están relacionadas (asociadas), una variable explica la otra. Si no hay diferencias de porcentajes es que no hay relación entre las variables. Y si la

diferencia es pequeña u ocurre solo en algunas categorías es que la relación explicativa entre las variables es débil.

Bibliografía sugerida

Sánchez Ramos, M. Á. (2005). Uso metodológico de las tablas de contingencia en la ciencia política. *Espacios Públicos*, *8*(16), 60-84.

Ejercicio

En la siguiente tabla de contingencia se presentan los resultados de cruzar las variables "Confianza en las grandes empresas" y "Género" basados en la encuesta Latinobarómetro realizada en Perú en 2012. Escriba un comentario e interpretación de la tabla donde se incluya:

- si hay asociación o no entre las variables,
- si existe asociación, señale si es fuerte, moderada o débil
- y qué tendencia o dirección tiene la asociación

Tabla 13. Tabla de contingencia: Confianza en las grandes empresas según género

		Género		Total
		Hombre	Mujer	
Confianza en las grandes empresas	Mucha confianza	57 9,6%	54 9,3%	111 9,5%
	Algo de confianza	176 29,7%	173 29,9%	350 29,8%
	No mucha confianza	214 36,1%	206 35,5%	420 35,8%
	Nada de confianza	145 24,5%	147 25,3%	292 24,9%
	Total	593 100,0%	580 100,0%	1173 100,0%

Solución: No existe asociación entre confianza en las grandes empresas y género. El 9,5% de los encuestados tienen mucha confianza en las grandes empresas, este porcentaje es casi idéntico para hombre y mujeres. El 29,8% del total de encuestados dicen tener algo de confianza en las empresas, y esta pauta es igual en hombres y mujeres. También es similar el porcentaje de respuestas entre los hombres y las mujeres que dicen tener no mucha confianza en las grandes empresas, y los que dicen no tener nada de confianza. En consecuencia, la confianza respecto a las grandes empresas es muy similar en hombres y mujeres, y por tanto no existe asociación significativa entre confianza en las grandes empresas y género. Otras variables explicativas que sí pueden estar asociadas a la confianza en las grandes empresas son el nivel de ingresos y el nivel educativo.

13. Correlaciones bivariadas

La correlación bivariada es una técnica estadística destinada a averiguar:

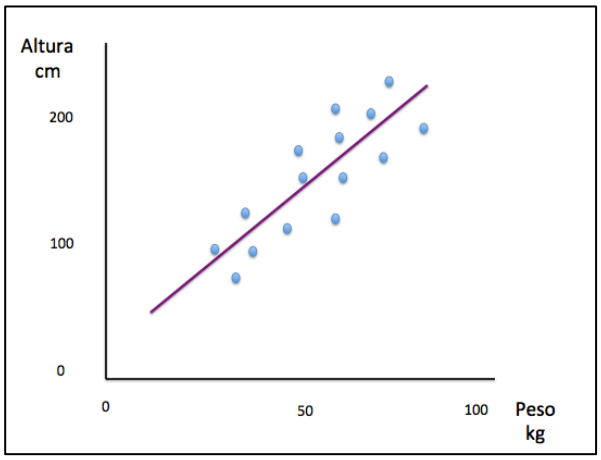
a) si dos variables tienen relación entre sí
b) si la relación es fuerte, moderada o débil y
c) qué dirección tiene la relación

Las coincidencias muchas veces esconden asociaciones entre fenómenos. La correlación es la técnica más usada para medir asociación lineal en todas las ciencias.

El análisis de correlaciones indica asociación o relación entre dos variables, no implica causalidad.[4]

La correlación está basada en la asociación lineal, es decir, que cuando los valores de una variable aumentan los valores de la otra variable pueden aumentar o disminuir proporcionalmente. Por ejemplo, la altura y el peso corporal tienen una relación lineal positiva, a medida que aumenta la altura aumenta el peso. Si realizamos un gráfico de puntos con ambas variables, la nube de puntos se asemejará a una diagonal, lo cual indica que hay correlación entre esas dos variables.

Figura 6. Correlación entre altura y peso



Fuente: elaboración propia

Existen 2 grandes tipos de correlaciones: correlación de Pearson y correlación de Spearman. Ambas están basadas en la misma información, aunque usan fórmulas diferentes. La correlación de Pearson es más adecuada cuando las variables siguen la curva normal. La correlación de Spearman es más conveniente usarla cuando las variables no siguen la curva normal. Por lo general, no suele haber muchas diferencias entre los resultados, aunque pueden variar sobre todo cuando se trabaja con muestras pequeñas.

[4] Se pueden realizar análisis de correlaciones con tres variables, son denominadas correlaciones parciales, las cuales sí permiten indicar cierta causalidad. Aunque para demostrar causalidad se suelen usar análisis multivariados.

Se usa la correlación en análisis estadístico de datos cuando trabajamos con variables ordinales o escalares. Las variables ordinales y escalares son aquellas que sus categorías tienen un orden interno. Si incluimos una variable nominal debemos recodificarla a variable dummy. Una variable dummy es aquella de solo dos categorías o valores, 1 y 0, la categoría o valor 1 indica presencia del fenómeno, y la categoría o valor 0 indica ausencia del fenómeno.

Cómo analizar la correlación bivariada en 2 pasos

La gran ventaja de la correlación es que toda la información de existencia de relación, fortaleza y dirección, aparece sintetizada en un coeficiente de correlación (r) y un nivel de significación (sig.).

1. El nivel de significación indica si existe o no relación entre dos variables. El nivel de significación más usado es 0,05, lo cual hace referencia al 95% de nivel de confianza, que es la probabilidad de que el resultado no se deba al azar. Cuando la significación obtenida de correlacionar dos variables es menor de 0,05 sí existe correlación significativa entre esas dos variables. Si existe correlación significativa debemos pasar al paso 2.

2. El coeficiente de correlación (r) indica que tan fuerte o débil es una correlación. Este coeficiente puede oscilar entre -1 y +1. Cuanto más se aleja de 0, más fuerte es la relación entre las dos variables. Cuanto más cerca de 0 es que la relación entre las variables es débil. Si el coeficiente de correlación es muy próximo a 0 es que ambas variables no están correlacionadas. Por otro lado, el signo (positivo o negativo) del coeficiente de correlación indica la dirección de la relación.

Varios ejemplos para entenderlo mejor:

Ejemplo 1:

La muestra (N) son: 2249 encuestados en Colombia (World Values Survey 2005)

Analizamos la relación entre “Ideología” e “Importancia de Dios en la vida”

- Ideología es una escala de 1 a 10, donde 1 es extrema izquierda y 10 es extrema derecha.
- Importancia de Dios en la vida es una escala donde 1 es nada importante y 10 muy importante.

		Ideología
Importancia de Dios en la vida	Correlación de Pearson (r)	0,124
	Significación	0,000
	N	2249

Existe correlación significativa entre Ideología e Importancia de Dios en la vida ya que la significación es 0,000 y por tanto menor de 0,05, por lo que este resultado no se debe al azar. La correlación de Pearson (r = 0,124) señala que se trata de una relación débil al estar próxima a 0. El signo positivo de esta correlación señala que a cuanto más de derecha son las personas en Colombia, más importancia le dan a Dios en la vida. Y lo mismo si

lo leemos de forma inversa, cuanta menos importancia de dan a Dios en la vida, más de izquierda son las personas colombianas.

Advertencia: hay que tener muy en cuenta cómo están ordenadas las categorías de las variables, ya que la lectura de la dirección de la relación está basada en el orden de las categorías. Es decir, el signo positivo de una correlación nos señala que los resultados se asemejan a una diagonal hacia arriba, pero es el investigador el que debe comprobar cómo se interpretan los resultados. Por ejemplo, la importancia de Dios en la vida fue preguntado en una escala de 1 a 10 donde 1 es nada importante y 10 es muy importante, y la ideología en una escala del 1 al 10 donde 1 es extrema izquierda y 10 es extrema derecha. El valor de la correlación de Pearson (r = 0,124) al ser positivo indica que, a más importancia de Dios en la vida, más de derechas son las personas, o lo que es lo mismo, a menor importancia de Dios en la vida, más tendencia a ser de izquierdas. Pero si la variable importancia de Dios en la vida hubiera sido preguntado en una escala de 1 a 10, pero donde el 1 es muy importante y el 10 es nada importante (al revés de como se hizo originariamente), el valor de la correlación de Pearson entre importancia de Dios en la vida e ideología hubiera sido (r = –0,124), es decir, mismo valor, pero con signo negativo. Esto indica que al aumentar los valores de la variable importancia de Dios en la vida, descienden los valores de la variable ideología. Como ideología está preguntado de izquierda a derecha, e importancia de Dios en la vida de muy importante a nada importante, la lectura sería: a más importancia de Dios en la vida, más de derechas son las personas, o bien menor importancia de Dios en la vida, más de izquierdas son las personas. Es decir, el resultado es el mismo, pero hemos de estar muy atentos al orden de las categorías de las variables para leer correctamente la dirección de una correlación significativa.

Ejemplo 2:

La muestra (N): 3017 personas en Colombia (World Values Survey 2005)

Analizamos la correlación entre “Edad” e “Interés en la política”

- La edad es una variable escalar
- Interés en la política es una variable ordinal donde las categorías son: 1-mucho interés, 2-bastante interés, 3-poco interés, 4-nada de interés

		Interés en la política
Edad	Correlación de Pearson (r)	0,013
	Significación	0,467
	N	3017

No hay correlación significativa entre edad e interés en la política ya que la significación es mayor de 0,05 (Sig. = 0,467). A medida que aumenta la edad no crece o decrece el interés en la política. Por tanto, deberíamos buscar otras variables si queremos

comprender con qué se relaciona el interés en la política, ya que la edad no correlaciona con el interés en la política.

El uso de la correlación es útil para caracterizar y extraer perfiles. Por ejemplo, permitiría identificar quiénes son los que más le dan importancia a Dios en la vida. De momento ya sabemos que los que tienen tendencia política más a la derecha. ¿Qué otras variables correlacionan con ésta? Las variables que correlacionen con la de nuestro interés serán las que nos permita identificar perfiles. Además, las correlaciones permiten analizar relaciones entre fenómenos (o variables). Por ejemplo, ¿hay relación entre la inversión en educación y la reducción de crímenes? ¿A más turismo extranjero más reducción de la pobreza? ¿A más número de becas concedidas menor satisfacción con el gobierno? Investigar es descubrir relaciones entre fenómenos, y las correlaciones son imprescindibles para ello.

Bibliografía sugerida

Llopis Pérez, J. (2012, noviembre 30). Tema 5: Correlación. Recuperado a partir de https://estadisticaorquestainstrumento.wordpress.com/2012/11/30/tema-4-correlacion/

Ejercicio

El siguiente cuadro presenta los resultados de calcular las correlaciones entre varias variables. A pesar que se presenten todas las variables en un cuadro, la correlación es una técnica bivariada donde se comprueba la asociación entre solo dos variables. Las variables analizadas y sus respectivas categorías de respuesta son:

- Nivel de felicidad: 1. Nada feliz, 2. Algo feliz, 3. Bastante feliz, 4. Muy feliz
- Importancia de Dios en la vida: 1. Nada importante, 2, 3, 4, 5, 6, 7, 8, 9, 10. Muy importante
- Interés en la política: 1. Muy interesado, 2. Algo interesado, 3. Poco interesado, 4. Nada interesado
- Nivel educativo: 1. Sin estudios, 2. Primarios, 3. Secundarios, 4. Universitarios

Revise la tabla y comente todos resultados. Debe incluir para cada análisis de correlación respuesta a estas preguntas:

a) ¿Existe asociación (o correlación) significativa entre las 2 variables?
b) Si existe asociación, ¿es una relación fuerte, moderada o débil?
c) Si existe asociación, ¿qué dirección tiene esta correlación?

Tabla 14. Tabla de correlaciones de Pearson

		Nivel de felicidad	Importancia de Dios en la vida	Nivel educativo	Número de hijos
Importancia de Dios en la vida	Correlación de Pearson	0,035			
	Significación	0,230			
	N	1199			
Nivel educativo	Correlación de Pearson	0,168	-0,116		
	Significación	0,000	0,000		
	N	1202	1207		
Número de hijos	Correlación de Pearson	-0,085	0,104	-0,358	
	Significación	0,003	0,000	0,000	
	N	1197	1202	1205	
Interés en la política	Correlación de Pearson	0,059	-0,062	0,213	-0,114
	Significación	0,041	0,033	0,000	0,000
	N	1193	1197	1200	1195

Solución:

- No existe correlación significativa entre nivel de felicidad e importancia de Dios en la vida ya que el nivel de significación es mayor de 0,05.
- Sí existe correlación significativa entre nivel de felicidad y nivel educativo ya que el nivel de significación es menor de 0,05. La correlación de Pearson es 0,168. Comparando este valor de correlación con los otros valores de correlación de la tabla, podemos señalar que la correlación entre nivel de felicidad y nivel educativa es moderada (ni muy fuerte ni tampoco muy débil). El signo positivo de la correlación de Pearson indica que a más nivel educativo más nivel de felicidad, o lo que es lo mismo, a menor nivel educativo, menor nivel de felicidad.
- Sí existe asociación (o correlación) significativa entre nivel de felicidad y número de hijos ya que la significación es menor de 0,05. El valor de la correlación de Pearson es muy próximo a 0 y comparándolo con otros valores de correlación de esta misma tabla nos indica que es una relación débil. El signo negativo nos indica que a mayor número de hijos, menor nivel de felicidad, o lo que es lo mismo, a menos o ningún hijo, más nivel de felicidad, aunque cabe tener en cuenta que es una relación débil.

14. Qué es y cómo analizar ANOVA de un factor

ANOVA de un factor (también llamada ANOVA unifactorial o one-way ANOVA en inglés) es una técnica estadística que señala si dos variables (una independiente y otra dependiente) están relacionadas en base a si las medias de la variable dependiente varían en las categorías o grupos de la variable independiente. Es decir, señala si las medias entre dos o más grupos son similares o diferentes.

ANOVA son las siglas de Analysis of Variance. Hay varios subtipos de ANOVA. Nos centramos en ANOVA de un factor ya que es un tipo de análisis bivariado ya que comprueba si dos variables están asociadas. Se le denomina ANOVA de un factor porque a la variable independiente se le conoce como factor.

1. ¿Cuándo usar ANOVA de un factor?

Usamos ANOVA de un factor cuando queremos saber si las medias de una variable son diferentes entre los niveles o grupos de otra variable. Por ejemplo, si queremos comparar el número promedio de hijos entre los grupos o niveles de clase social: clase baja, clase trabajadora, clase media-baja, clase media-alta y clase alta. Es decir, vamos a comprobar mediante ANOVA si la variable "número de hijos" está relacionada con la variable "clase social". Concretamente, se analizará si la media del número de hijos varía según el nivel de clase social a la que pertenece la persona.

Condiciones:

- En ANOVA de un factor solo se relacionan dos variables: una variable dependiente (o a explicar) y una variable independiente (o explicativa, que en esta técnica se suele llamar factor)
- La variable dependiente es cuantitativa (escalar, u ordinal de más de 5 categorías) y la variable independiente es categórica (nominal u ordinal).
- Se pide que las variables sigan la distribución normal, aunque como siempre esto es difícil de cumplir en investigaciones sociales.
- También que las varianzas (es decir, las desviaciones estándar al cuadrado) de cada grupo de la variable independiente sean similiares (fenómeno que se conoce como homocedasticidad). Aunque esto es lo ideal, en la realidad cuesta de cumplir, e igualmente se puede aplicar ANOVA, aunque los resultados deben ser interpretados con mayor precaución.

2. ¿En qué se basa ANOVA de un factor?

ANOVA de un factor compara las medias de la variable dependiente entre los grupos o categorías de la variable independiente. Por ejemplo, comparamos las medias de la variable "Número de hijos" según los grupos o categorías de la variable "Clase social".

Si las medias de la variable dependiente son iguales en cada grupo o categoría de la variable independiente, los grupos no difieren en la variable dependiente, y por tanto no hay relación entre las variables. En cambio, y siguiendo con el ejemplo, si las medias del número de hijos son diferentes entre los niveles de la clase social es que las variables están relacionadas.

3. ¿Qué estadísticos se calculan en ANOVA?

Al aplicar ANOVA de un factor se calcula un estadístico o test denominado F y su significación. El estadístico F o F-test (se llama F en honor al estadístico Ronald Fisher) se obtiene al estimar la variación de las medias entre los grupos o categorías de la variable independiente y dividirla por la estimación de la variación de las medias dentro de los grupos. El cálculo del estadístico F es algo complejo de entender, pero lo que hace es dividir la variación entre los grupos por la variación dentro de los grupos. Si las medias entre los grupos varían mucho y la media dentro de un grupo varía poco, es decir, los grupos son heterogéneos entre ellos y similares internamente, el valor de F será más alto, y por tanto, las variables estarán relacionadas. En conclusión, cuanto más difieren las medias de la variable dependiente entre los grupos o categorías de la variable independiente, más alto será el valor de F. Si hacemos varios análisis de ANOVA de un factor, aquel con un valor de F más alto indica que hay más diferencias, y por tanto una relación más fuerte entre las variables. Para decir que el valor de F es alto o medio o bajo, debemos compararlo con otro análisis realizado con los mismos datos, o datos similares.

4. ¿Cómo se interpreta el test de F y la significación?

La significación de F se interpreta como la probabilidad de que este valor de F se deba al azar. Siguiendo un nivel de confianza del 95%, el más utilizado en ciencias sociales, cuando la significación de F sea menor de 0,05 es que las dos variables están relacionadas.

Hemos de analizar e interpretar al aplicar ANOVA de un factor:

- **Significación**: si es menor de 0,05 es que las dos variables están relacionadas y por tanto que hay diferencias significativas entre los grupos.
- **Valor de F**: cuanto más alto sea F, más están relacionadas las variables, lo que significa que las medias de la variable dependiente difieren o varían mucho entre los grupos de la variable independiente.

5. Ejemplos de ANOVA de un factor

Ejemplo 1:

Quiero averiguar si el número de hijos varían según la clase social. Para ello comparo las medias de números de hijos entre los diversos niveles, grupos o categorías de clase social: clase baja, clase trabajadora, clase media baja, clase media-alta, y clase alta. Utilizo los datos de la Encuesta Mundial de Valores realizada entre 2010 y 2014 en 58 países del mundo.

Introduciré las dos variables que quiero analizar:

- "Número de hijos": es una variable cuantitativa que va de 0 hasta 8 o más hijos. Esta será la variable dependiente o a explicar
- "Clase social": variable ordinal. Es la variable independiente o factor. Las categorías o grupos de la clase social son: clase baja, clase trabajadora, clase media baja, clase media-alta, y clase alta.

El resultado es el siguiente:

En la tercera columna se observan las medias para cada grupo de clase social. Si nos fijamos en las medias del número de hijos en cada grupo de clase social, podemos observar que a medida que aumenta la clase social desciende la media del número de

hijos. Las personas de clase baja tienen de media 2,17 hijos, las de clase trabajadora 1,93 hijos, las de clase media-baja 1,82 hijos en promedio, las de clase media-alta tienen 1,74 hijos de media, y las de clase alta tienen de media 1,76 hijos.

Tabla 15. Descriptivos: Número de hijos según clase social

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mín	Máx
					Límite inferior	Límite superior		
Clase alta	1938	1,76	1,798	0,041	1,68	1,84	0	8
Clase media-alta	17889	1,74	1,735	0,013	1,71	1,76	0	8
Clase media-baja	31389	1,82	1,729	0,01	1,8	1,84	0	8
Clase trabajadora	24333	1,93	1,76	0,011	1,91	1,96	0	8
Clase baja	10760	2,17	2,014	0,019	2,14	2,21	0	8
Total	86309	1,88	1,783	0,006	1,87	1,89	0	8

Tabla 16. ANOVA de un factor

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	1526,009	4	381,502	120,622	0
Intra-grupos	272961,208	86304	3,163		
Total	274487,217	86308			

El valor de F es 120,622 y la significación es 0,000. Al ser la significación menor de 0,05 es que las diferencias de la media de hijos entre los grupos de la clase social son significativas. Aunque aparentemente podamos pensar que las diferencias no son exageradas, la decisión de si las diferencias son significativas no depende de nuestro criterio, sino de la significación de F. Este es el objetivo de aplicar ANOVA de un factor: valorar estadísticamente si las diferencias de medias son significativas o no.

Ejemplo 2:

Relacionamos la variable "Número de hijos" con la variable "¿Es una persona religiosa?". Se les preguntó a las personas si eran religiosos y se les dio 3 opciones de respuesta:

1. soy es una persona religiosa
2. no soy una persona religiosa
3. soy ateo

Comparamos la media de hijos por los grupos de la variable "¿Es una persona religiosa?". En la tabla de descriptivos nos fijamos en la columna de las medias, y observamos que

las personas religiosas tienen de media 2,06 hijos, las no religiosas tienen de media 1,56 hijos y los ateos 1,3 hijos en promedio.

Tabla 17. Descriptivos: Número de hijos según es una persona religiosa

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mín	Máx
					Límite inferior	Límite superior		
Soy una persona religiosa	57781	2,06	1,878	0,008	2,04	2,07	0	8
No soy una persona religiosa	22186	1,56	1,545	0,01	1,54	1,58	0	8
Soy ateo	4199	1,3	1,289	0,02	1,27	1,34	0	8
Total	84166	1,89	1,789	0,006	1,87	1,9	0	8

Tabla 18. ANOVA de un factor

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	5496,42	2	2748,21	876,775	0
Intra-grupos	263804,841	84163	3,134		
Total	269301,262	84165			

La significación de F es 0,000, al ser menor de 0,05 es que hay relación significativa entre las variables, y el valor de F es 876,775.

Podemos comparar el valor de F de este ejemplo con el del ejemplo anterior ya que están basados en los mismos datos:

- "Número de hijos" según "Clase Social" F = 120,622
- "Número de hijos" según "¿Es una persona religiosa?" F = 876,775

El valor de F es más grande en la relación de "Número de hijos" y "¿Es una persona religiosa?" porque las diferencias de medias del número de hijos son más grandes que las diferencias de medias según la clase social. Cuanto más alto sea el valor de F, más diferencias de medias habrá entre los grupos o categorías de la variable independiente (o factor) y por tanto más fuerte es la relación entre las variables.

6. Conclusión

ANOVA de un factor se utiliza muchísimo en las ciencias sociales. Es muy popular en psicología y en análisis comparativo y experimental para poder saber si las diferencias de un grupo respecto a otro son significativas y qué fortaleza tienen.

A la hora de presentar ANOVA en un informe, artículo de investigación o tesis, se debe presentar la tabla de las medias y desviaciones típicas, y seguidamente el estadístico F y su significación.

Sin ANOVA de un factor, las diferencias entre un grupo y otro quedarían a juicio subjetivo del observador, y donde una persona ve diferencias otra quizás no las verías. Es mejor usar la estadística para saber si hay similitud o diferencia entre los grupos, y ANOVA es excelente para este propósito.

Ejercicio

En la Encuesta Mundial de Valores se preguntó a las personas en qué nivel toleran la prostitución en una escala del 1 al 10, donde 1 es nunca lo justificaría y 10 es siempre lo justificaría. Se analizan las diferencias de opinión respecto a la "Tolerancia con la prostitución" según la "clase social" de los encuestados, y posteriormente según el "género" de los encuestados. Se aplicó ANOVA de un factor, y estos son los resultados.

Comente cada análisis de ANOVA y responda después qué variable independiente (o factor) está más relacionada con la "Tolerancia con la prostitución".

Tabla 19. Descriptivos: Tolerancia con la prostitución según clase social

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
Clase alta	1414	2,91	2,672	0,071	2,77	3,05	1	10
Clase media-alta	12769	2,92	2,565	0,023	2,88	2,97	1	10
Clase media-baja	22310	2,81	2,493	0,017	2,77	2,84	1	10
Clase trabajadora	17913	2,64	2,429	0,018	2,6	2,67	1	10
Clase baja	8268	2,56	2,428	0,027	2,5	2,61	1	10
Total	62674	2,75	2,489	0,01	2,73	2,77	1	10

Tabla 20. ANOVA de un factor: Tolerancia con la prostitución según clase social

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	1023,008	4	255,752	41,39	0
Intra-grupos	387240,042	62669	6,179		
Total	388263,05	62673			

Tabla 21. Descriptivos: Tolerancia con la prostitución según género

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
Hombre	30613	2,88	2,578	0,015	2,85	2,91	1	10
Mujer	33293	2,65	2,407	0,013	2,62	2,67	1	10
Total	63906	2,76	2,493	0,01	2,74	2,78	1	10

Tabla 22. ANOVA de un factor: Tolerancia con la prostitución según género

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	860,138	1	860,138	138,68	0
Intra-grupos	396352,437	63904	6,202		
Total	397212,575	63905			

Solución:

- Las respuestas del total de personas encuestadas en cuanto a la tolerancia con la prostitución en una escala del 1 al 10 donde 1 es nunca y 10 es siempre, se sitúan en promedio en 2,75 (desviación estándar 2,489). Existen diferencias según la clase social. Las clases alta y media-alta tienen más tolerancia hacia la prostitución ya que el promedio es mayor que el de las clases baja y trabajadora. Estas diferencias de medias son significativas según el análisis de ANOVA de un factor ($F = 41,39$; $Sig. = 0,000$).
- Si analizamos las diferencias en la tolerancia con la prostitución según el género de los encuestados, observamos que los hombres tienden a ser más tolerantes que las mujeres en promedio, 2,88 los hombres y 2,65 las mujeres. Estas diferencias son significativas según el análisis de ANOVA de un factor ($F = 138,68$; $Sig. = 0,000$).
- El valor más alto de F indica que las diferencias respecto a la tolerancia con la prostitución según el género son mayores que según la clase social. Es decir, hay más variación de opiniones según el género que según la clase social de las personas. Aunque tanto género como clase social están relacionadas con la tolerancia con la prostitución, el género es más explicativo que la clase social para entender las diferencias de opiniones respecto a la prostitución.