Documento 01

El coeficiente de correlación

1. Qué mide el coeficiente de correlación

El coeficiente de correlación *r de Pearson* expresa *en qué grado los sujetos tienen el mismo orden en dos variables*.

Si los sujetos más altos pesan más y los más bajitos pesan menos, entre peso y altura tendremos una correlación *positiva: a mayor altura, mayor peso.*

Si los de más edad corren más despacio y los más jóvenes corren más deprisa, entre edad y velocidad tendremos una correlación *negativa; a mayor edad, menor velocidad.*

Los coeficientes de correlación pueden ser por lo tanto *positivos* o *negativos*. Lo que expresan estos coeficientes se entiende bien mediante su representación gráfica, los *diagramas de dispersión* en los que las dos variables están simbolizadas con las letras X e Y (figura 1).

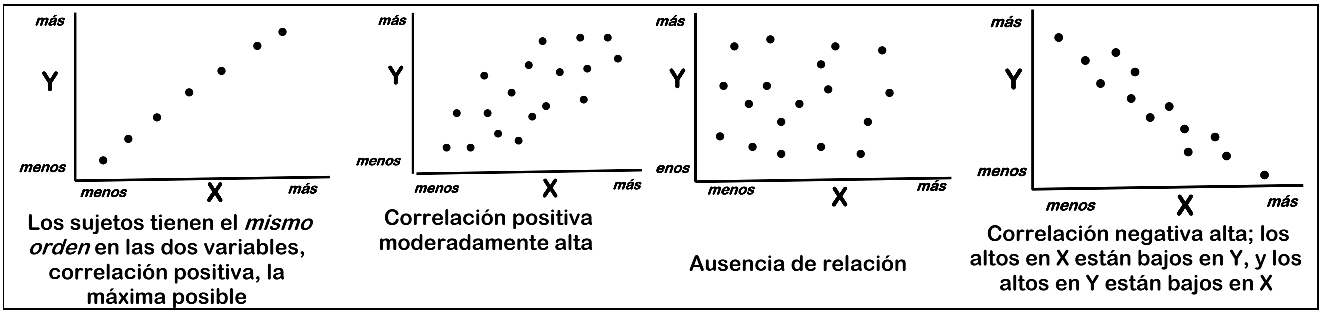

Figura 1

El valor del coeficiente de correlación oscila entre 0 y ±1; una correlación igual a 0 significa *ausencia de relación*. Si de los mejores profesores unos investigan mucho y otros nada o muy poquito y de los profesores que no son tan buenos docentes unos investigan y publican mucho y otros no investigan nada, entre productividad en investigación y calidad de la docencia *no hay relación* (que no hay que confundir con *relación negativa*).

Los coeficientes de correlación se calculan con toda facilidad en programas informáticos (como EXCEL y el SPSS).

2. Importancia de las diferencias para detectar relaciones

Para que se dé una relación (*distinta de cero*) tiene que haber sujetos distintos en las dos variables. No podemos ver si entre los jugadores de baloncesto los más altos encestan más y los más bajos encestan menos si todos son de la misma altura.

Podemos verlo de manera intuitiva en la figura 2, donde tenemos el número de respuestas correctas de cuatro alumnos en cinco exámenes.

Sujeto	Historia	Lengua	Física	Matemáticas	Arte
1	10	9	9	4	6
2	10	9	8	3	7
3	10	9	3	2	8
4	10	9	2	1	9

Figura 2

Entre Historia y Lengua la correlación es igual a *cero* porque no hay diferencias, lo mismo que entre cualquiera de estas dos asignaturas y las demás.

Entre Física y Matemáticas tenemos la máxima relación (r = 1) porque el *orden* de los sujetos es el mismo en las dos asignaturas; la correlación expresa *semejanza en el orden, NO en valores absolutos*

Entre Arte y Matemáticas o Física la correlación es también la máxima posible pero *negativa* (r = -1); lo mejores en Matemáticas y Física son los peores en Arte y viceversa.

Una consecuencia práctica es que, para detectar relaciones, en nuestros instrumentos de recogida de datos debemos intentar recoger las diferencias que de hecho existen; no es lo mismo una pregunta con respuestas *sí* o *no* (en números, 1 ó 0) que si las respuestas son *mucho, bastante, poco* y *nada* (dando a estas respuestas los valores de 4, 3, 2 y 1).

3. Los coeficientes de correlación uniendo o separando subgrupos

Cuando en nuestra muestra de sujetos hay *subgrupos* (carreras, cursos, sexos, etc.) las relaciones que podemos encontrar pueden ser muy distintas si las calculamos *en cada grupo* o en *todos los sujetos*. En estos casos podemos *probar* calculando las correlaciones que nos interesen en cada subgrupo y en cada submuestras.

Podemos verlo de manera intuitiva en las siguientes figuras. En la figura 3 tenemos gráficamente la relación entre *peso* y *altura* en un grupo de niños de 7 años (A) y en otro de 12 años (B) (datos ficticios). En cada grupo r = 0; el ser un poco más alto o un poco más bajo no significa pesar más o menos; en cambio uniendo los dos grupos la relación es alta; los más altos pesan más y los más bajos pesan menos.

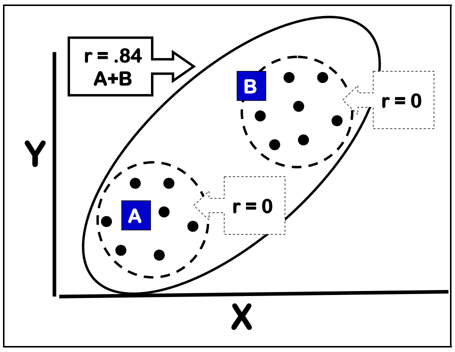

Figura 3

En la figura 4 tenemos la correlación entre notas en *matemáticas* (X) y un test de *inteligencia* (Y) en dos secciones de un mismo curso. En ambos grupos la relación es grande; los más inteligentes tienen mejores notas, pero si unimos los dos grupos en uno solo la correlación baja notablemente. ¿Qué puede estar sucediendo? El test sí está relacionado con rendimiento en matemáticas, pero un profesor califica más alto y otro más bajo.

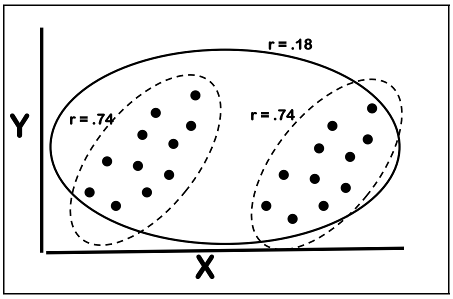

Figura 4

En un caso como el de la figura 4 podríamos calcularla correlación entre el test y rendimiento por separado en cada una de los dos casos y después calcar la correlación media (ponderando cada coeficiente por su número de sujetos (*correlación media* = (N_1r_2+N_2r_2/N_1+N_2)).

En la figura 5 (ejemplo también ficticio) X es *nota media al finalizar una carrera universitaria* e Y es el resultado de una *prueba de admisión* en la Universidad muy exigente. El subgrupo encerrado en el círculo con trazo continuo es el de los *admitidos* que terminaron la carrera; en este grupo la relación entre test de admisión y resultado final la relación es cero. En trazo discontinuo tenemos la correlación que *hubiéramos obtenido si todos hubieran sido admitidos*.

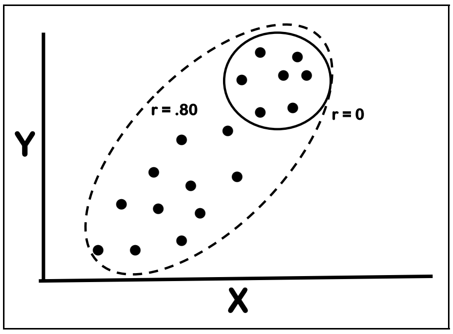

Figura 5

La figura 5 representa de manera exagerada una situación que puede ser relativamente común. Cuando hay un proceso de selección exigente la *muestra admitida* queda *homogeneizada,* entre los sujetos *admitidos* hay menos diferencias en la prueba de admisión y la correlación entre esta prueba y cualquier criterio de éxito posterior será inferior a la que se hubiera obtenido si todos hubieran sido admitidos (podría ser no ya inferior, sino negativa). En estos casos la conclusión no debería ser que *la prueba de admisión es inútil porque no predice el éxito*; hay modos de estimar la correlación que *hubiéramos obtenido* si todos hubieran sido admitidos.

4. Influjo en los coeficientes de correlación de sujetos con puntuaciones extremas

Unos pocos sujetos *con puntuaciones muy extremas* en una o en las dos variables pueden *desvirtuar la información* sobre el grado de relación entre dos variables como puede apreciarse en las figuras 6 y 7 (ejemplos exagerados para mayor claridad).

En la figura 6 en el grupo encerrado en trazo continuo hay una correlación positiva y alta entre X e Y, pero un solo sujeto con una puntuación muy alta en una variable (X) y muy baja en la otra (Y) hace bajar la correlación que en este ejemplo es incluso negativa (trazo discontinuo).

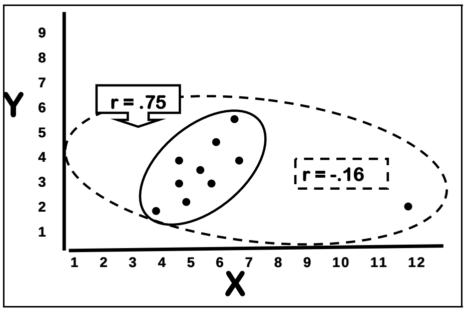

Figura 6

En la figura 7 vemos en el grupo encerrado en trazo continuo una correlación de cero (ausencia de relación) pero un solo sujeto con una puntuación muy alta las dos variables hace que la correlación entre X e Y sea positiva y muy alta (trazo discontinuo).

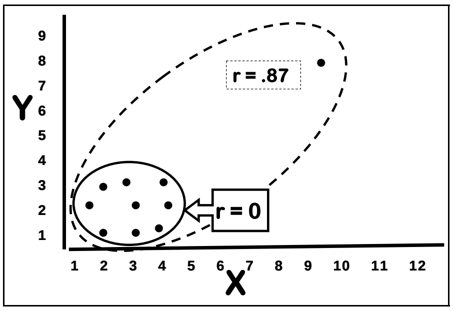

Figura 7

¿Qué se puede hacer cuando hay sujetos con puntuaciones muy extremas? Una solución es prescindir de estos sujetos, sobre todo si el énfasis está en verificar relaciones entre variables más que en analizar una muestra concreta de sujetos. También se pueden calcular las correlaciones *con y sin* estos sujetos (que serán pocos) o examinar qué caracteriza a estos sujetos (quizás son de una edad muy distinta a la de la mayoría y *no representan a la misma población*). En algunos programas informáticos (como el SPSS) está previsto el poder prescindir de los sujetos con puntuaciones extremas (por ejemplo los que se apartan de la media en más de 1.5 desviaciones típicas).

5. Cómo interpretar los coeficientes de correlación

5.1. Correlación y causalidad.

Lo primero que hay que advertir es que un coeficiente de correlación nunca se puede interpretar como *una relación de causa a efecto;* el que dos variables *covaríen,* tiendan a ser altas o bajas simultáneamente (peso y altura, actitud hacia el estudio y rendimiento académico, etc.) no significa que una sea causa de la otra, y aunque así fuera, un coeficiente de correlación no es prueba de causalidad. Una correlación de .99 entre *tiempo dedicado al estudio* y *rendimiento académico* no se puede interpretar como que el buen rendimiento *se debe* al tiempo dedicado al estudio aunque sí puede proponerse como *hipótesis;* el tema de la causalidad es complejo, para probar causalidad hay que poder *excluir otras explicaciones;* la correlación simplemente expresa *asociación de hecho.*

5.2. Cuándo un coeficiente de correlación es *estadísticamente significativo*

Para interpretar un coeficiente de correlación *el primer paso* es verificar si ha sido *casual* (si es o no es *estadísticamente significativo*) y en segundo lugar podemos ya valorar su magnitud.

Qué entendemos por *estadísticamente significativo* lo vemos fácilmente con un ejemplo. No tienen ningún sentido esperar que el día de nacimiento (de 1 a 31 de cualquier mes) esté relacionado con el número de teléfono (en ese caso los nacidos a finales de mes tendrían números de teléfono más altos, y los nacidos a primeros de mes tendrían números más bajos). Sin embargo si hacemos el cálculo con dos sujetos nada más podemos encontrar una correlación de 1, la máxima posible por pura casualidad; en cambio con 100 sujetos eso no sucedería. El primer paso es por la tanto *descartar el azar* como explicación de la correlación obtenida, es entonces cuando decimos que la correlación es *estadísticamente significativa*; dicho en términos más académicos la correlación es muy improbable que se explique por el *error muestral* en el supuesto de que no haya relación entre las dos variables.

Una correlación estadísticamente significativa quiere decir que en una muestra semejante encontraríamos una correlación entre las dos variables *distinta de cero* (pero no necesariamente de una magnitud parecida). Podemos extrapolar el *hecho* de la relación, no su magnitud.

Este *descartar el azar* depende ya del número de sujetos.

Con muestras pequeñas

Un coeficiente de correlación grande puede ser simplemente *casual* y no nos permite afirmar que *en general* se da esa relación con una razonable seguridad de no equivocarnos (aunque *sí* existe en *esa* muestra, los sujetos están ordenados de manera parecida en las dos variables, pero puede ser pura coincidencia…); por ejemplo con N = 4 una correlación de .90 (grande) puede ser puramente casual que no nos permite afirmar que las dos variables están relacionadas.

Con muestras grandes

Un coeficiente pequeño puede ser superior a lo que se puede esperar por azar; con N = 150, una correlación de .20 (pequeña) nos saldría menos de 5 veces de cada 100 por azar (exactamente el 1.4 % de las veces); podemos afirmar con seguridad que sí existe esa relación aunque no sea grande.

Encontraremos con más facilidad correlaciones estadísticamente significativas en muestras relativamente grandes.

El límite convencional para descartar el azar es el 5%: afirmamos que una correlación es *estadísticamente significativa* cuando se podría explicar por factores aleatorios solamente 5 veces o menos de cada cien.

Este *umbral* del 5% es convencional pero es el aceptado habitualmente y se denomina *nivel de confianza*; suele expresarse α = .05 (probabilidades de error al afirmar la relación) o también (en positivo) *nivel de confianza* del 95% (probabilidades de acertar al afirmar la relación). La expresión habitual es p < .05, o probabilidad de error inferior al 5% al afirmar que se da un relación (con niveles de confianza mas estrictos pondríamos p < .01 o p < .001).

Los valores críticos de los coeficientes de correlación para distintos valores de N (número de sujetos) los tenemos en tablas de los textos de estadística (hasta N = 100) y en programas de Internet.

5.2. Cómo valorar la magnitud de los coeficientes de correlación

Una vez que determinamos que un coeficiente de correlación es *estadísticamente significativo*, como los valores mínimo y máximo son 0 y ±1 podemos valorar su magnitud: .20 será intuitivamente una relación baja y .85 indicará una relación que ya podemos considerar grande.

En términos absolutos. A partir de .30 ya es una magnitud que se puede considerar *apreciable* (e incluso *grande* para algunos autores) dado que en las Ciencias Sociales suelen ser bajas por la falta de precisión en nuestras medidas (preguntas, tests) y además no siempre se recogen los datos en circunstancias óptimas (respuestas rápidas, cansancio o falta de interés de los que responden, etc.). Una correlación baja (pero *estadísticamente significativa*) puede ser la punta del *iceberg;* lo que calculamos es lo que somos capaces de *cuantificar* con los datos que obtenemos con nuestros instrumentos, pero en la realidad la relación puede ser mucho mayor.

En términos relativos también se puede (y conviene) valorar estos coeficientes, comparando el coeficiente de correlación con otros obtenidos en el mismo contexto, y se puede advertir al comentarlo (por ejemplo, r = .35, relación alta *en este contexto*, o *la mayor* de todas las vistas, etc.).

Otra manera de valorar la magnitud de un coeficiente de correlación consiste en elevarlo al cuadrado; r^2 se denomina *coeficiente de determinación* e indica la *proporción de variabilidad común*. Si entre un test de inteligencia y notas en matemáticas encontramos una correlación de r = .50, esto quiere decir que el 25% ($.50^2$) de las diferencias en matemáticas (propiamente el 25% de la varianza) se explica (o está determinado) por las diferencias en *ese* test de inteligencia.

5.4. Cómo valorar coeficientes de cierta magnitud pero no estadísticamente significativos

A veces encontramos coeficientes de correlación que:

1º Son de cierta magnitud (.30 o bastante mayores),
2º No son estadísticamente significativos al nivel de confianza habitual ($p< .05$) pero tampoco es muy probable que sean *casuales* (p entre .06 y .09…),
3º Obtenidos en muestras pequeñas,
4º Que además vemos que responden a *hipótesis plausibles*, que tienen su lógica.

Este señalar *como hipótesis*, que necesitarían confirmarse en muestras mayores, correlaciones no significativas puede tener un interés añadido cuando todos los coeficientes están en la misma dirección (son del mismo signo); por ejemplo cuando los coeficientes de correlación entre una serie de opiniones (grado de acuerdo de 1 a 6) y la variable sexo tienen el mismo signo.

En estos casos una conclusión que se puede proponer *como hipótesis* es que en muestras mayores sí obtendremos un coeficiente de correlación estadísticamente significativo.

6. Entre qué variables se puede calcular un coeficiente de correlación

1º Cuando *las dos variables son continuas*, es decir, admiten una serie de valores (como edad, curso, respuestas de 1 a 5 en un cuestionario, etc.); los números *tienen que tener sentido*. En principio el coeficiente *r de Pearson* está pensado para este tipo de variables.

2º Cuando *una variable es dicotómica y la otra continua*. Variables *dicotómicas* son las que solamente admiten dos valores, *unos y ceros* (los dos sexos, jornada matutina o jornada vespertina, nacional o extranjero, o también respuesta correcta o incorrecta, etc.). *Unos* y *ceros* equivalen de alguna manera a *presencia* o *ausencia* de un rasgo o característica y tiene sentido utilizarlos como *números*.

Una correlación significativa y *positiva* significa que los sujetos codificados con un *uno* tienen en la variable continua una media mayor que los sujetos codificados con un *cero*; si la correlación es *negativa*, la media mayor en la variable continua corresponde a los sujetos codificados con un *cero*.

El término académico del coeficiente de correlación cuando *una variable es dicotómica y otra continua* es *correlación biserial-puntual*.

En este caso, correlación entre una variable dicotómica y otra continua, un análisis alternativo que nos llevaría a las mismas conclusiones es el contraste de medias (*t de Student*); de hecho disponemos de una sencilla fórmula para convertir el valor de la t de Student en un coeficiente de correlación.

3º Cuando *las dos variables son dicotómicas*, las dos codificadas con *unos y ceros*. Pueden ser dos preguntas respondidas *bien o mal*, o dos características de la persona que sólo admiten dos posibilidades (como sexo y estudia Derecho o Psicología). No se trata propiamente del coeficiente *r de Pearson*, pero el cálculo en los programas informáticos (como EXCEL o el SPSS que no entienden de *conceptos*, sólo, en este caso, de *unos y ceros*) es el mismo. El término apropiado de este coeficiente es ϕ (*phi*, letra *efe griega*) y está asociado al *ji cuadrado*.

Un coeficiente positivo quiere decir que los sujetos tienden a estar clasificados o en los *dos unos* o en los *dos ceros*; un coeficiente negativo quiere decir que la tendencia es a estar en *uno* en una variable y en *cero* en la otra variable.

El que haya *nombres y fórmulas* distintos para designar los coeficientes de correlación calculados con distintos tipos de variables se debe en parte a que la estadística se desarrolló antes que las calculadoras y los programas informáticos, pero también es verdad que la interpretación hay que *modularla* en función del tipo de datos con los que calculamos estas correlaciones.

Cuando *no se puede calcular un coeficiente de correlación* es cuando una de las dos variables admite más de dos categorías de clasificación (como Psicología, Derecho e Ingeniería) aunque estas categorías estén codificadas con números que en este caso son meros símbolos de conceptos y no números en sentido propio (no expresan orden o cantidad) aunque en las hojas de cálculo se introducen con números.

7. Cómo presentar los resultados de un análisis correlacional

El *output* o salida de un programa informático es una *matriz de correlaciones* en la que aparecen los coeficientes de correlación de cada variable con todas las demás. Un ejemplo lo tenemos en la tabla de la figura 8[1].

N = 20	*Edad*	*Sexo*	*1 Aborto*	*2. Eutanas.*	*3 Pena Mte.*	*4. Evad imp.*	*5. Sobornar*
Edad	1						
Sexo	-0,270	1					
1. Aborto	-0,390	0,620	1				
2. Eutanasia	-0,352	0,206	0,304	1			
3. Pena de muerte	-0,353	0,124	0,655	-0,032	1		
4. Evadir impuestos	-0,242	-0,207	-0,181	-0,144	0,225	1	
5. Sobornar autoridad	-0,396	-0,039	0,328	0,349	0,460	0,218	1

Figura 8

Veinte alumnos universitarios (de entre 18 y 21 años) han expresado su grado de aprobación a cinco cuestiones de tipo ético-legal (seis respuestas, de 1, *nunca o casi nunca*, a 6, *en muchas circunstancias*); además tenemos el sexo (Varón = 1, Mujer = 0) y la edad.

Observaciones sobre las matrices de correlaciones.

1. En la *diagonal* vemos *unos* porque la correlación de cada ítem consigo mismo es perfecta; la atabla es simétrica (la correlación del ítem 1 con el 3 es la misma que la del 3 con el 1) por lo que sólo aparecen resultados en la mitad de la tabla (es lo que dan directamente los programas informáticos).

2. En los coeficientes de correlación son suficientes *tres decimales* (o incluso dos).

3. Estas tablas hay que *arreglarlas* sin limitarse a copiarlas directamente de un programa informático, de manera que la información quede muy clara y estéticamente presentable; es recomendable *copiar la redacción de los ítems* (puede ser en forma abreviada) en los encabezados de filas y/o columnas.

4. También es recomendable recordar en la misma tabla el número de sujetos (N = 20) aunque se supone que está indicado en otro lugar.

[1] Datos de Andrea Marroquín, Marcela Pereira y Virna Zamora, trabajo del Diplomado en Investigación (Guatemala, Universidad Rafael Landívar, 2011).

5. Lo primero que hay que hacer es indicar *de manera visible* qué coeficientes son estadísticamente significativos y poner fuera de la tabla los valores de referencia (los buscamos en tablas o en Internet). En este caso y para N = 20 estos valores (que se podrían indicar debajo de la tabla) son .444 (p = .05), .56 (p = .01) y .679 (p = .001).

En este caso (figura 8), como nuestro nivel de confianza es de .05, se destacan en blanco con fondo oscuro los coeficientes estadísticamente significativos (pues superan .44); también es habitual ponerlos en *negrita* o añadir asteriscos a los coeficientes (p < .05*, p< .01** y p<.001***), por ejemplo y en este caso, pondríamos .460*, .620** y .655**.

6. Estas tablas se pueden *subdividir en tablas menores* para mayor claridad en la presentación de los resultados y en interpretación, sobre todo en tablas muy grandes. En este caso hubieran sido preferibles dos tablas: una con las con las correlaciones entre los ítems y otra con las correlaciones de los cinco ítems con *edad* y *sexo*.

8. Cómo interpretar los resultados de un análisis correlacional

Al interpretar y comentar los coeficientes de correlación conviene *buscar cierta estructura.* En el ejemplo de la tabla de la figura 8:

Con respecto a las relaciones entre los ítems:

Los que mantienen una postura más liberal con respecto a la *pena de muerte* también la tienen con respecto al *aborto* (es la relación *mayor en este contexto* y no pequeña en términos absolutos) y con respecto a *sobornar a la autoridad para conseguir algo importante* (ésta es la formulación completa del ítem). El resto de las correlaciones tienen valores dentro de lo aleatorio.

Con respecto a las relaciones de los ítems con sexo y edad

1) *Relación con sexo*. Vemos una relación alta (.62) entre *sexo* y *aborto*. Como los varones están codificados con un *uno* y las mujeres con un *cero*, esta relación muestra que *los varones son más liberales* que las mujeres con respecto al aborto. En un contraste de medias, encontraremos que la media de los varones en permisividad con el aborto es significativamente más alta que la de las mujeres (la conclusión será la misma en los dos análisis). Esta conclusión es compatible con que las dos medias sean *muy bajas en términos absolutos* (pero una será mayor que la otra…). En un estudio más completo figurarán también los datos descriptivos de las dos muestras (aquí nos limitamos a los coeficientes de correlación).

2) *Relación con edad*. No hay ningún coeficiente estadísticamente significativo (ninguno llega a .44 que es el valor crítico). No podemos afirmar que *en la población representada por esta muestra* haya una relación *distinta de cero*, pero (según lo advertido en el apartado 5.4) observamos que:

1º La muestra es muy pequeña (20 sujetos) y en muestras tan pequeñas no es fácil encontrar coeficientes de correlación estadísticamente significativos.

2º La magnitud de las correlaciones no es *tan pequeña*, oscilan entre .24 y .39 (casi .40) y la mayoría superan .30

3º Todos los coeficientes tienen *el mismo signo menos* (lo que indica que *los más jóvenes están menos de acuerdo con posturas más liberales en estos temas*).

Estos datos nos sugieren que se puede proponer *como hipótesis* que *los más jóvenes son más conservadores* en estos temas (ciertamente lo son en esta muestra), pero esta hipótesis habría que verificarla *en una muestra mayor* (y quizás mas representativa, no sólo mayor) para salir de dudas.

8. Temas de ampliación

Hay temas de ampliación que quedan abiertos y que pueden consultarse en la obra de referencia *(Morales Vallejo, P. (2008). Estadística aplicada a las ciencias sociales. Madrid: Universidad Pontificia Comillas)*

1) ¿Cuál es la *fórmula* del coeficiente r, por qué mide *relación* o *semejanza en el orden* de los sujetos en dos variables y por qué su valor máximo es ±1? El entender la *fórmula original* (no otras derivadas para simplificar los cálculos antes de generalizarse el uso de calcadoras y programas informáticos) ayuda a interpretar correctamente estos coeficientes (cap. 5, apartado 2.1).

2) En el caso de la figura 5 ¿Cómo podemos *estimar* la correlación que hubiéramos obtenido entre una prueba de admisión y resultados finales *si todos hubieran sido admitidos?* (cap. 5, apartado 4.2).

3) Una correlación *estadísticamente significativa* quiere decir que en la población representada por esa muestra hay una relación *distinta de cero* (es decir, *hay relación* aunque sea pequeña); pero *¿qué magnitud tiene esa relación?* El valor de esa correlación en la población no lo sabemos, pero sí podemos estimar entre qué límites máximo y mínimo se encuentra (cap. 5, apartado 5.6; está clara la *teoría,* en la *práctica* se resuelve fácilmente en programas de Internet).

Por ejemplo; hemos visto que la correlación entre la postura hacia el *aborto* y hacia *la pena de muerte* es de .655 en una muestra de 20 sujetos. Con un nivel de confianza del 95 % en la *población* representada por esa muestra oscilaría entre .316 y .855; el valor más bajo probable es superior a *cero,* por eso la correlación es estadísticamente significativa (> 0 en la población). En cambio entre *edad* y *aborto* la correlación que en esta muestra es r = -.390, en la población estaría entre -.709 y +.063; entre estos valores se encuentra el *cero,* por eso este coeficiente de -.390 no es estadísticamente significativo.

4) Si tenemos tres variables, X, Y y Z muy relacionadas entre sí ¿Cuál será correlación entre X e Y igualando a todos los sujetos en Z (neutralizando las diferencias en Z)? (cap. 5, apartado 5; se resuelve fácilmente en programas de Internet).

Por ejemplo entre *sexo* (X) y *postura ante al aborto* (Y) tenemos una correlación de .62; por otra parte vemos que entre *sexo* (X) y *edad* (Z) tenemos una correlación de -.27 y entre *aborto* (Y) y *edad* (Z) la correlación es de -.39 ¿Cuál será la correlación estimada entre *sexo* y *aborto si todos fueran de la misma edad*? En este caso esta relación baja ligeramente de .62 a .58, lo que nos hace pensar que la *edad* está pesando en la relación entre *sexo* y *aborto* (la edad media de las mujeres es 18.9 y de los varones es 18.4).

5) En muestras de *dos carreras distintas* calculamos la correlación entre *autoeficacia* y *probabilidad percibida de obtener calificaciones altas* ¿Es esta relación significativamente más alta en una carrera que en otra? (no está explicado en el texto de referencia, pero la teoría es análoga a la del contraste de dos medias; se resuelve fácilmente en Internet).

9. La correlación en Internet

Son bastantes los programas de Internet que resuelven fácilmente todo lo relacionado con los coeficientes de correlación; dos especialmente útiles son:

The Chinese University of Hong Kong, Department of Obstetrics and Gynaecology, http://department.obg.cuhk.edu.hk/index.asp?scr=1024 (valores críticos del coeficiente de correlación para cualquier valor de N).

Lowry, Richard, VassarStats: Web Site for Statistical Computation,Vassar College, Poughkeepsie, NY, USA; http://faculty.vassar.edu/lowry/VassarStats.html (para las cuestiones planteadas antes en 3, 4 y 5).