



Serie de Posts
Número 5

**Los 'Grupos de Calidad' en las Universidades Cubanas:
¿cuánto más se puede hacer por la Autoevaluación? (3° parte)**



En la primera parte de este Número 5 nos centramos en la importancia de ampliar las escalas de evaluación, con la utilización de reactivos que intenten aislar los errores que más frecuentemente se presentan en el cumplimiento de las exigencias curriculares universitarias (tanto en lo instructivo, como en lo formativo). En la segunda parte tratamos la cuestión del análisis de las propiedades psicométricas de las pruebas de logro y de los cuestionarios de contexto; lo mismo si estos utilizan escalas dicotómicas, o de créditos, o graduadas (como la 'Escala Likert').

Este otro tema es trascendente y la utilización de la Autoevaluación Universitaria para ilustrarlo ha sido solo un pretexto para referirlo. En realidad, todos los

científicos sociales (psicólogos, docentes, sociólogos, administrativos, políticos, etc.) deberíamos conocerlo y aplicarlo, pues a diferencia de los físicos, los químicos, los ingenieros y los médicos, entre otros, los que nos ocupamos del estudio científico de cualidades personalógicas de los individuos no tenemos instrumentos precisos para medirlas. Sí, es cierto, los creamos por cientos y hasta por miles, pero pocas veces sabemos cuán válidos son; es una condición que, desde la teoría, la damos por verdadera.

Ahora, ya lo decíamos en la primera parte del Número: ¡hay que conectar los contenidos de esos dos temas precedentes! Claro, si comprendemos la necesidad de ajustarnos al rigor de la ciencia hasta el final; porque de poco vale desgastarnos en crear instrumentos para esas cualidades subjetivas y aplicarlos con tan alto costo material, financiero y humano en una representación de las unidades de análisis del objeto de investigación, si al final la interpretación de sus resultados queda a merced de los siempre ambiguos pensamientos especulativos y empíricos. De eso es, precisamente, de lo que hablaremos hoy.

La necesidad de superar los períodos ‘primitivo’ y ‘medievo’ de la Estadística aplicada a la investigación científica

Uno de los atributos más importantes de la actuación científica es la permanente búsqueda de ‘la mayor objetividad posible’. Eso no debería ser difícil de entender por los científicos sociales cubanos, pues estamos dotados de una Epistemología (ciencia del pensamiento verdadero; o lo que es igual, ‘un paso’ más adelante de la Gnoseología) que establece como uno de sus pilares básicos la objetividad de la práctica, en contraste con la actividad intelectual que intenta conocerla (para transformarla, y no solo para describirla y modelarla, como se ha puesto tan de moda últimamente).

Así, para tensionar la reflexión sobre esta raigal cuestión de la Metodología de la Investigación Científica, me gusta insistirles a mis talentosos estudiantes del grupo nacional de ‘alto desempeño’ (que ha conformado el Ministerio de Salud Pública y gestiona mi Universidad de Ciencias Médicas de La Habana; por cierto, ¡una magnífica iniciativa!): “Un resultado científico solo existe en nuestras cabezas y en la comunicación que establecen los investigadores en torno a él”. Otra cosa es acudir a la técnica y a la innovación para introducirlos en la práctica y producir las transformaciones anheladas.

En efecto, un resultado científico es siempre un producto subjetivo y, por supuesto, sometido a todos los ‘avatares’ con que ese plano de producción convive: ‘lo que sé’, ‘lo que me gustaría’, y/o ‘lo que me conviene’ del objeto de investigación); es un acto reflejo de esa parte de la práctica que se estudia; la cual, dada la naturaleza inagotable de la realidad objetiva, no puede ser ‘atrapada’ en su totalidad; a lo que hay que añadir las ‘ataduras’ del condicionamiento histórico-concreto de toda investigación (es decir, lo que es conocido del objeto o fenómeno estudiado hasta ese momento, la tecnología disponible para estudiarlos, etc.).

Por tanto, son muchos los ‘riesgos’ a los que están sujetos los investigadores al realizar su labor; posiblemente, el peor de todos es no saber ‘apartar’ su subjetividad lo más posible ante cada paso del proceso investigativo. Pero, ¿cómo puede lograrse eso?... Solo hay una respuesta: elevando al máximo el rigor de la aplicación del método científico, y asumiendo a la práctica misma como punto de partida y ‘juez supremo’ de la veracidad de su resultado final. Desde nuestra Epistemología, no se podrá admitir como verdadero un resultado investigativo que sea contradicho por la práctica. **Ni expertos, ni tribunales saben de la realidad más que ‘ella misma’.**

Hay mucho más que decir en torno a estos trascendentes temas de la investigación científica, pero no ‘cabén’ este Post. Sin embargo, concomitando con lo dicho (y con lo pendiente) está la cuestión de los recursos de la Estadística aplicada a su campo. Lo he dicho otras veces, pero toca repetirlo: para una buena parte de los científicos sociales cubanos, la utilización de los más recientes avances de la Estadística sigue teniendo la peligrosa condición de ‘**asignatura pendiente**’, especialmente en el caso de los ‘estudios explicativos’; lo más intentados y necesarios.

Y eso ocurre en el medio de grandes **paradojas**; por un lado: **[i]** cuando cada día se tienen más claras evidencias de que la realidad objetiva es sumamente compleja y dinámica, e incluso perturbada (esto es, que ante la acumulación considerable de hechos extremos se pueden producir cambios críticos, que incluye la expansión combinada de efectos que hasta entonces se manifestaban de forma independiente); por otro lado: **[ii]** que muy a pesar del desarrollo vertiginoso de las **TIC**, que hacen cada vez más accesibles y eficientes a los software estadísticos; tanto que ya casi se reduce su utilización al saber elegir el recurso estadístico apropiado para la investigación que se ejecuta, y el poder interpretar el significado de los resultados devueltos, generalmente ya en forma gráfica.

Sin embargo, se siguen acudiendo a recursos simplistas (en comparación con los posibles y necesarios) que, aun sin ánimo de molestar a nadie, no puedo clasificarlos sino como herramientas de los períodos ‘primitivos’ y ‘medieval’ del decursar histórico de la Estadística aplicada a la investigación científica. Me estoy refiriendo a los recursos de la **Estadística Descriptiva** (con su peculiar presentación de resultados con gráficos de barras y tablas de por cientos), en el primer caso, y a las **pruebas de hipótesis** en el segundo (con la famosa frase de ‘estadísticamente significativo’, que ni siquiera se entiende cabalmente, con no poca frecuencia).

No importa que con la Estadística Descriptiva solo podamos ‘hablar’ de lo ‘visto’ en la muestra elegida (aunque los hay que, sin el menor pavor, hablan a seguidas de ‘generalización’ de sus resultados), ni que las pruebas de hipótesis tan solo nos digan si muy probablemente dos (o más) medidas estadísticas similares obtenidas con la muestra (o que la comparación de una de ellas con un valor alcanzado en la población consistentemente, o que la comparación de la distribución empírica con otra teórica) difieran o no. ¡Así de simple!... ¡Nada de cómo se relacionan entre sí las muchas variables que subyacen en la manifestación del objeto de

investigación, o de en qué medida algunas de ellas influyen (“impactan”) a otras con las que están asociadas, de alguna forma, y que tanto ayudaría a explicar causas de efectos!

La desaparecida pero autorizada Dra. Martha Martínez Llantada tenía una alerta muy pertinente para desatinos como estos, en la Investigación Científica; la **parfraseo**: “Podemos pensar que estamos avanzando hacia la esencia del objeto de estudio; pero lo que estamos haciendo es alejarnos cada vez más de ella”.

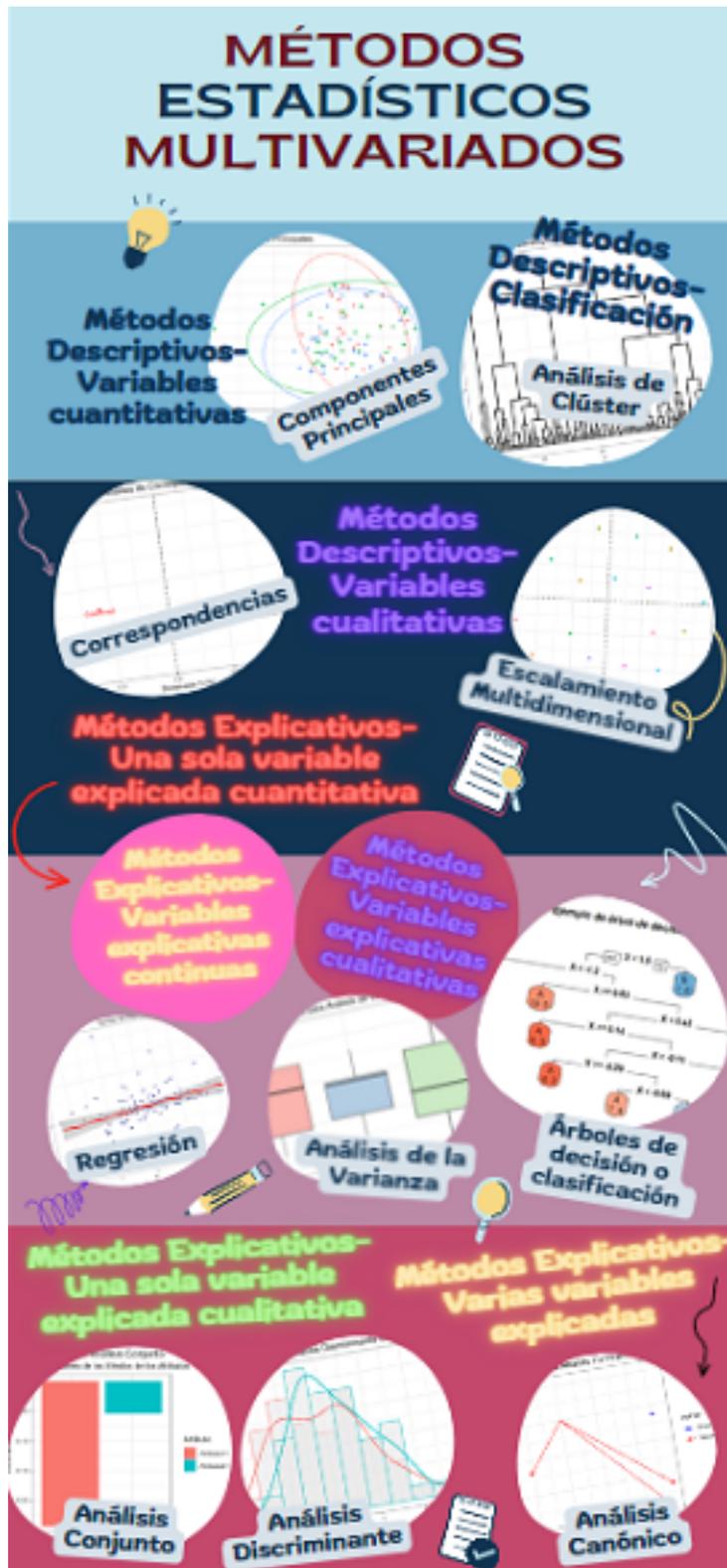
Una parte considerable, si bien no todos, de esos desaciertos metodológicos de la investigación científica en el ámbito social, en nuestro país, parecen estar asociados con errores conceptuales y con el insuficiente dominio de las técnicas del ámbito de la Estadística aplicada (y por extensión con los software creados para facilitarlas); en particular, con los **métodos del análisis multivariado y multinivel de la Estadística Inferencial**.

Para contribuir, al menos modestamente, a ofrecer algo ‘de luz’ en torno a ellos, he resumido en la infografía que aparece más abajo los diez métodos fundamentales del **análisis multivariado** descritos por César Pérez (2004), en su magnífico texto: “**Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS®**”.

- 1. Análisis Factorial y de Componentes Principales:** cuando se requieren métodos descriptivos y se disponen de variables cuantitativas.
- 2. Correspondencias:** como método descriptivo, pero con variables cualitativas no ordinales.
- 3. Escalamiento Multidimensional:** también como método descriptivo, pero cuando las variables son cualitativas ordinales.
- 4. Análisis de Clúster:** igualmente como método descriptivo, pero con fines de clasificación.
- 5. Regresión:** como método (ahora) explicativo; con solo una variable explicada y cuantitativa, mientras que la (las) variable(s) explicativa(s) es (son) también cuantitativa(s).
- 6. Análisis de la Varianza:** también como método explicativo; pero con solo una variable explicada y cuantitativa, y variables explicativas cualitativas.
- 7. Segmentación (o Árbol de decisión, clasificación y regresión):** método explicativo, ‘a medio camino’ entre las técnicas en las que la variable explicada es cuantitativa o cualitativa, pero las variables explicativas son cualitativas.
- 8. Análisis Conjunto:** también método explicativo con solo una variable explicada, cualitativa como las variables explicativas.
- 9. Análisis Discriminante:** como el caso anterior, solo que las variables

explicativas son cuantitativas.

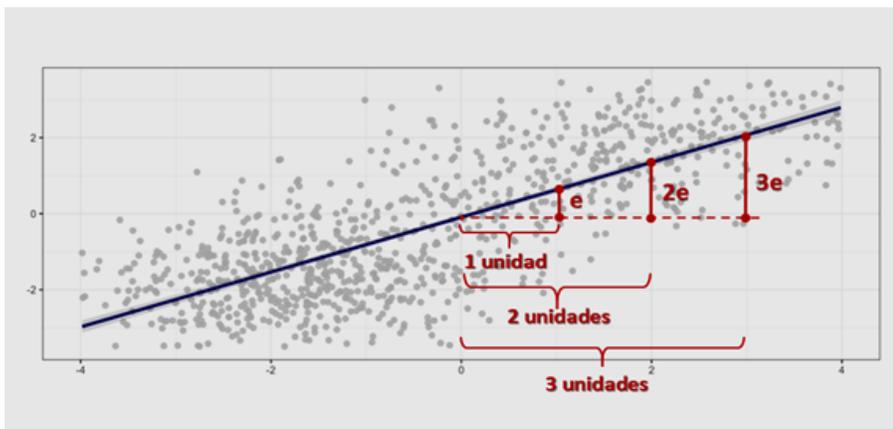
10. Análisis Canónico: método explicativo, pero donde se tienen más de una variable explicada.



Como se puede apreciar, ni 'son demasiadas', ni conforman alternativas de elección; cada método responde a una situación de investigación, y de tipos de variables derivadas de su diseño, diferentes.

La marcada potencia del método multivariante de Regresión

La técnica estadística de **Regresión** (lineal o binomial, entre otras) dispone de una doble potencialidad: una **capacidad predictiva** y otra **explicativa**. Ciertamente, por un lado, es capaz de ‘hablarnos’ de estimaciones de ‘efectos’ que se ubican más allá del rango de los datos provenientes del ‘trabajo de campo’ de la investigación; mientras que, por el otro, nos permite estimar además una medida del ‘impacto’ (pudiera decirse; mejor de ‘verdadero impacto’ y no de ese que tanto se habla con una ligereza desenfadada y desconcertante) de las variables explicativas sobre la variable explicada, que se asumen en el modelo estadístico. Ese valor de ‘impacto’ estimado por el modelo crece proporcionalmente, por cada incremento en una unidad de la variable predictora, como se muestra en la imagen que sigue.



El inconveniente con la regresión estadística es que, como se señaló arriba, esa variable explicada debe ser única y además numérica (o cuantitativa). Sin embargo, ya vimos en la parte segunda de este Número de la Serie de Posts, que esas dos exigencias no constituyen problemas. Recuerden, con los avances de los recursos psicométricos apoyados en la **Teoría de Respuesta al Ítem (TRI)** podemos reducir a un único **índice estadístico** (numérico) las respuestas a instrumentos compuestos por escalas cualitativas (como repasamos al inicio: lo mismo sean dicotómicas, con créditos, que graduadas). De modo que el esfuerzo requerido por esa transformación adicional bien vale la pena. A ello hay que añadir que los modelos de regresión aceptan también variables explicativas cualitativas, con tal de que al menos una de ellas sea cuantitativa. Nada, ¡algo genial para las investigaciones en ciencias sociales!

Regresemos al ejemplo de la autoevaluación educativa organizada por el ‘Grupo de Calidad’ de una Facultad Universitaria

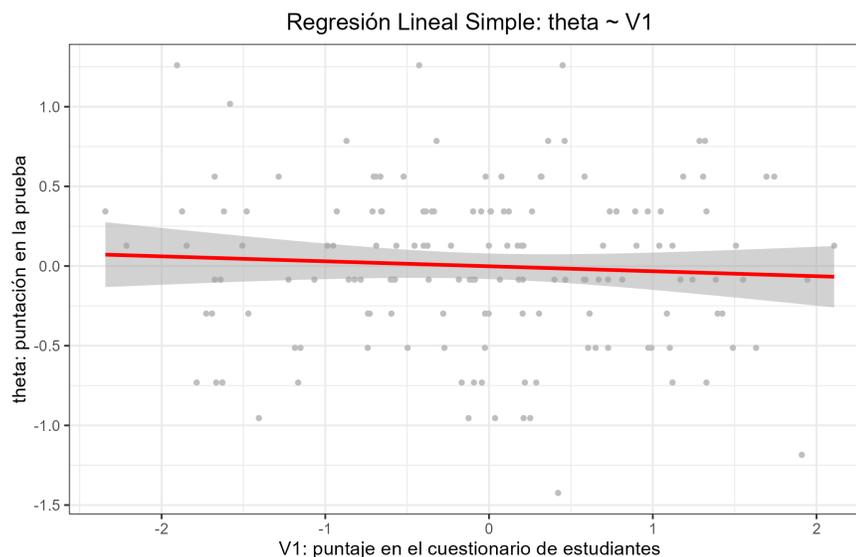
Efectivamente, el lector asiduo recordará que en la segunda parte del presente Número 5 simulamos la realización de un proceso de **autoevaluación**, previo a otro de **Acreditación Universitaria**, a cargo del ‘Grupo de Calidad’. Recuerden que dimos por hecho que ese estudio exploratorio se centraría en la evaluación de una

disciplina curricular, en un año académico determinado y en una carrera específica de la Facultad.

Supusimos también que para ello fue aplicada a 150 estudiantes del Año una **prueba de logro cognitivo** de doce reactivos (la mayoría de ellos de selección múltiple, pero con una única respuesta correcta y tres distractores en cada uno; mientras que los dos últimos constituyeron preguntas de desarrollo y tres créditos diferentes de calificación). Con las respuestas de los estudiantes fueron generados puntajes propios que conformaron una distribución numérica, siendo esta la **variable a explicar**.

Pero más adelante, se refirió que también fue aplicada una **encuesta de contexto** (o también llamada de 'factores asociados al logro'). Esta fue administrada a 200 informantes (entre estudiantes, docentes y directivos), pero que solo 171 de ellos respondieron todas sus preguntas, de modo que hubo que excluir a los restantes de la base de datos creada, por contener **missing** (o 'datos perdidos'); es decir, se optó por no '**impugnar la base de datos**' (esto es, no arriesgarnos a perder precisión, estimando esos valores desconocidos de los datos faltantes, en vez de excluir los registros correspondientes).

Supongamos ahora que ninguno de los 150 estudiantes evaluados en la prueba de logro cognitivo está entre los informantes que dejaron preguntas sin responder en el cuestionario, de modo que podemos articular una 'saludable' base de datos con **dos variables numéricas**: el puntaje de la prueba (**variable a explicar**) y el **índice estadístico**, también numérico, construido con las respuestas a las preguntas del cuestionario (**variable explicativa**). Como se ilustra en la imagen de la portada del presente Post, tenemos condiciones para aplicar el método estadístico de **Regresión**. Al considerar una única variable explicativa, diremos que se trata de una **Regresión lineal simple**. Esa imagen 'nos dice' además que el procesamiento estadístico lo haremos con nuestro 'genio de la lámpara': **R-Project**.



En la imagen anterior está la salida gráfica proporcionada por **R**. Al fondo, en color gris claro, está la '**nube de puntos**' que se genera al mapear la combinación

de **V1** (variable explicativa) y **theta** (variable a explicar). La línea roja es el resultado del modelo diseñado, acompañada de la banda de intervalos de confianza de esos puntos alineados (construida con un 95% de confianza al inferir). El resultado no es bueno, en el sentido de que en la medida que nos movemos a lo largo del eje horizontal (el de la variable predictora **V1**) casi no cambia la estimación de su efecto sobre la variable representada en el eje vertical (o sea, la variable explicada **theta**).

Podría decirse, que **V1** no ‘impacta’ (casi nada) a **theta**; vamos, son independientes una de otra. Esto se entiende mejor en la siguiente imagen, que presenta la salida analítica (tabular) de la estimación del efecto (**estimate**) y de su nivel de significación estadística (**p.value**). El primero, casi cero; el segundo mayor que 0.05.

Estimaciones de los coeficientes en el modelo de regresión lineal simple

term	estimate	std.error	statistic	p.value
(Intercept)	0.00	0.04	-0.03	0.97
V1	-0.03	0.04	-0.76	0.45

Para no dejar ‘lagunas’ por el camino, el parámetro ‘(Intercept)’; o sea, el **intercepto de la recta de regresión** (con la recta paralela al eje vertical por $x=0$) no es aquí trascendente, pues solo se refiere al valor predicho por el modelo cuando aún no ha comenzado a influir la variable predictora **V1**.

R-Project nos devuelve también, con sus poderosas ‘librerías’ especializadas, los valores de un grupo de **parámetros de control de calidad del modelo estadístico** diseñado y ‘fijado’. En la siguiente imagen se presentan los más importantes.

Estimaciones de los parámetros del modelo de regresión lineal simple

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual	nobs
0	0	0.5	0.57	0.45	1	-107	220.1	229.1	36.6	148	150

El valor del parámetro de **R-cuadrado** (**‘r.squared’**), y que sin redondear es 0.00384, nos indica que apenas el 0.38% de la variabilidad de **theta** es explicada por el modelo asumido. Mientras que el parámetro de **R-cuadrado ajustado** (**‘adj.r.squared’**), y cuyo valor sin redondear es -0.00289, se ajusta por el número de variables predictoras incluidas (aquí una sola, **V1**), y es tan cercano a cero que refuerza la idea de que el modelo no explica bien la variabilidad de los datos a los que trata de ajustarse. Por otro lado, un valor p (**‘p.value’**) de 0.45, para el **estadígrafo F**, expresa que el modelo fijado no es significativamente mejor que un modelo sin predictor alguno.

Resumiendo, ni el efecto de la variable predictora **V1** es grande, ni estadísticamente significativo como para asumirlo en la población toda; ni los parámetros de control de calidad del modelo ‘hablan’ a favor de un buen ajuste del modelo a los datos.

Es importante destacar que estos resultados no ocurren necesariamente por razones estadísticas, o psicométricas. Generalmente, ellos tienen sus causas en la omisión del análisis de la **validez del instrumento en relación con el constructo** (es decir, que la habilidad o inclinación general que se pretendió evaluar con la encuesta a los estudiantes no estuvo suficiente ajustada en relación con el desenvolvimiento de la asignatura que se evaluó con la prueba de logro cognitivo), o por errores en la **operacionalización de la variable** que pretende reflejar ese constructo (lo más probable, por no partir de asumir los atributos o **definiens** de la definición ‘conceptual’ o teórica de esa misma variable antes); es decir, más por razones metodológicas (de la investigación científica) que **docimológicas o estadísticas**, como señalamos.

Otra razón probable (a la par o no de las anteriores) es la **decisión de trabajar con pocas variables predictoras**, ¡siendo tan compleja y multifacética la realidad objetiva cuya esencia se trata de ‘atrapar’!

Vamos entonces a suponer que el ‘Grupo de Calidad’ previó ese inconveniente y no se limitó a administrar solo una encuesta de contexto a los estudiantes, sino que aplicó además **otra encuesta** relacionada con la marcha del proceso de enseñanza de esa asignatura, **pero entre los profesores** que la imparten en el Año académico; los resultados fueron integrados en otro índice estadístico, que en lo adelante se referirán como variable predictora **V2**.

También se entrevistaron con los directivos que han monitoreado el desarrollo de la **preparación de la asignatura** y que les han realizado **controles a clases** de la asignatura a esos docentes, pudiendo calificar a cada uno de ellos con una escala compuesta por tres categorías descendentes (“metodológicamente muy eficaz”, “medianamente eficaz en el orden metodológico” e “insuficientemente eficaz en el orden metodológico”). Esas calificaciones, integradas en un único valor, fueron agregadas en la base de datos a los estudiantes de la brigada estudiantil a la que le imparte docencia cada docente; se tiene así una tercera **variable predictora V3**.

Resumiendo, ahora el ‘Grupo de Calidad’ dispone de una base de datos más robusta, integrada por cuatro variables: la explicada ‘**theta**’ y las tres predictoras; de las cuales dos son **numéricas** y la última (**V3**) es **categorica ordinal**. Como la variable de salida ‘**theta**’ es numérica se puede pensar en emplear nuevamente el **método estadístico de Regresión**; solo que ahora no se podrá trabajar con un modelo de regresión lineal simple, sino **múltiple**.

Puesto que se tienen varias variables predictoras, se pueden suponer diferentes formas de relacionarlas, pero eso sería solo para el **análisis de un ‘modelo base’** que se corresponda con las relaciones originales de los datos, pues ya sabemos

que al tomar decisiones sobre la base de la especulación o la experiencia estaremos cometiendo un error epistemológico de consideración.

Lo más frecuente es partir, para el **modelo basal**, de suponer que la **suma de efectos** de esas (tres) variables es lo que mejor explica la variabilidad de la variable a explicar (θ). Pero si se tienen razones suficientemente plausibles, se pueden añadir otras relaciones, como las de **'ajustada por'** (que supone la división de los valores de un par de variables predictoras), o la de **'interactuadas'** (que se traduce en la multiplicación de los valores de un par de variables predictoras).

En nuestro ejemplo, además de sumar los efectos individuales de las tres variables predictoras, vamos a añadir el ajuste de la variable **V2** por la variable **V3**, teniendo en cuenta que la variable **V2** se refiere a la inclinación de los docentes por los atributos de calidad de la conducción del proceso docente de la asignatura de referencia, mientras que **V3** hace referencia a la eficacia del trabajo metodológico que ellos realizan para prepararse en esa dirección, aunque esté calificada desde la perspectiva de los cuadros administrativos que los monitorean; es decir, **apoyados en el hecho comprobado de que la calidad del trabajo metodológico ajusta la calidad de la dirección del proceso educativo.**

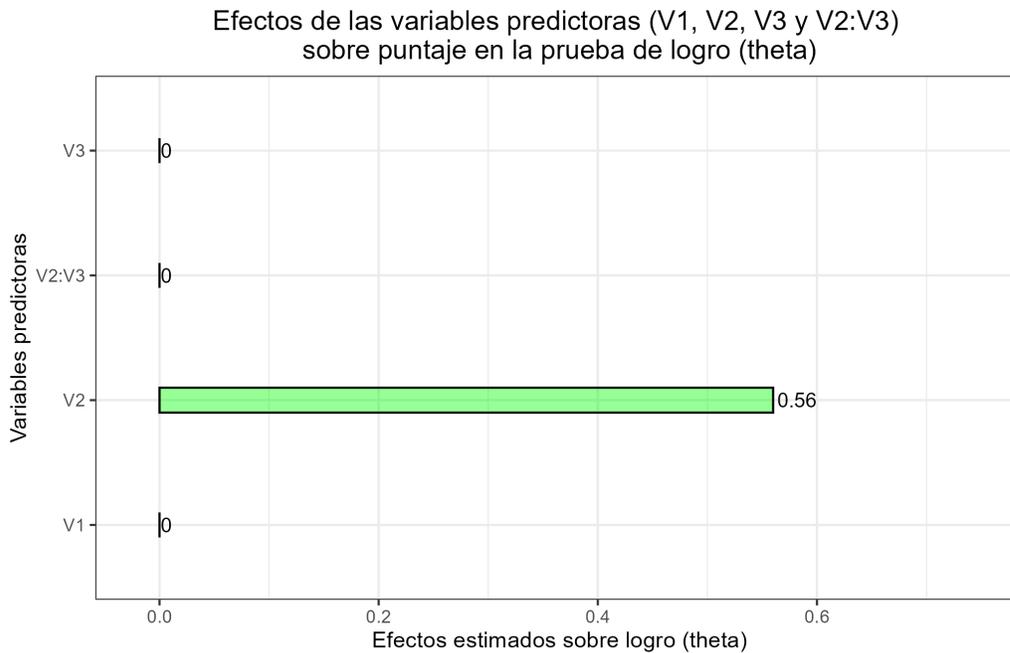
De modo que nuestro **'modelo basal'** múltiple tendría el diseño estadístico siguiente: **"la variabilidad de θ es explicada por la suma de los efectos individuales de V1, V2 y V3, más el efecto del ajuste de V2 por V3"** (lo digo así porque me sigo negando a escribir ecuaciones matemáticas que pocos entienden y terminan asustando al lector). Veamos ahora los resultados de la corrida del modelo así diseñado.

Estimaciones de los coeficientes en el modelo de regresión lineal múltiple

term	estimate	std.error	statistic	p.value
(Intercept)	-0.17	0	-609.38	0.00
V1	0.00	0	-0.78	0.44
V2	0.56	0	2057.70	0.00
V3	0.00	0	0.28	0.78
V2:V3	0.00	0	-2.50	0.01

Aunque es solo un modelo de partida, para salir a la búsqueda de otros **'mejores'**, se aprecian ciertos avances, pues las variables predictoras **V2** y la combinación **V2:V3** resultan estadísticamente significativas (con **'valores p'** inferiores incluso al 1% de error; es decir, con significaciones estadísticas superiores al 99% de confianza); mientras que, al parecer **V1** y **V3** están **'empujando'** al modelo a un **'sobreajuste'** (o sea, el modelo se adapta tanto a los datos de la muestra que no generaliza bien hacia otros datos de la población de la que se extrajo aquella). Por otra parte, solo **V2** está mostrando un efecto considerable sobre la variabilidad de θ , pues el de la combinación **V2:V3**, sin redondear a las

centésimas, es de solo **-0.000495**; casi 0. Gráficamente, tenemos los siguientes resultados devueltos por R-Project.



Esta vez no acudiremos a la nube de puntos y a una recta de regresión pasando ‘por su centro’, pues se trata de un modelo de regresión lineal múltiple y eso significa que el efecto de la segunda (y después de la tercera, y de la cuarta variables) se obtienen cuando queda ajustado el efecto de la variable predictiva que le precede, de modo que utilizaremos aquí un gráfico de barras; la longitud de las barras de los efectos de las variables **V1** y **V3**, así como de la combinación **V2:V3** (recuérdese, “**V2 ajustada por V3**”) son tan pequeños que no se divisan en el gráfico generado.

Veamos ahora el comportamiento de los parámetros de calidad del modelo basal.

Estimaciones de los parámetros del modelo de regresión lineal múltiple

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual	nobs
1	1	0	4075525	0	4	764.9	-1518	-1500	0	145	150

Los valores de **R-cuadrado** (**‘r.squared’**) y **R-cuadrado ajustado** (**‘adj.r.squared’**) son sospechosamente perfectos (recuerden que ellos nos hablan de la proporción de la variabilidad de la variable de salida (aquí, **theta**) que es explicada por el modelo, lo cual refuerza la suposición de un efecto de ‘sobreajuste’. Por su parte, el valor p del estadístico F es muy pequeño, casi cero (sin redondear y en notación científica: **2.2*10-16**); de modo que se puede rechazar la hipótesis nula de que todos los coeficientes de regresión son iguales a cero (lo que equivaldría a que el modelo no tiene utilidad predictiva).

Podemos concluir que, definitivamente, este modelo basal no puede resultar el mejor modelo posible a crear con esas variables predictoras. En fin, debemos

seguir trabajando.

Pero nuestro ‘genio de la lámpara’ (R-Project) es tan potente que dispone de ‘librerías’ (conjunto de códigos de programación) que son capaces de sorprendernos, además de asistirnos. Le pedí a seguidas que me generara todos los modelos de regresión lineal que (con esas variables inicialmente asumidas) pudieran constituir **candidatos a ‘mejor modelo’**. Claro, previamente tuve que programar dos condiciones: **(i) verificar si existía colinealidad (alta correlación) entre pares de variables predictivas, en cuyo caso habría que excluirlas de los modelos candidatos, y (ii) aunque puede llegar a generar cientos (y hasta miles) de modelos, que no sobrepasaran el 10% del número de registros disponibles en la base de datos (aquí 15, o sea el 10% de los 150 estudiantes en torno a los cuales se proyectan θ , V1, V2 y V3)**. Veamos el caso de la exclusión de las relaciones de colinealidad entre variables predictoras en la imagen siguiente.

Exclusión de relaciones de colinealidad

V1	V2	V3
NA	NA	NA
TRUE	NA	NA
TRUE	TRUE	NA

En esa tabla de salida, ‘NA’ significa que el valor no está disponible (bien por obvio, como sucede en la diagonal descendente de la matriz, dado que la correlación de una variable con ella misma es siempre perfecta [o sea, tiene coeficiente 1], o bien porque los valores del triángulo superior coinciden con sus simétricos en el triángulo inferior); mientras que **TRUE** (verdadero) significa que con esa pareja de variables se puede seguir adelante; en fin, con las tres: **V1, V2 y V3**.

Hecho esto, le pedí a R-Project que generara los **candidatos a mejor modelo**. La siguiente tabla contiene los diez modelos de más calidad, ordenados por **AICc** (o sea, por el **Coeficiente de Información de Akaike ajustado**) y, por tanto, por **Delta** (es decir; por la diferencia del valor del **AICs** de un modelo con relación al que le sigue, en orden ascendente). Se aprecian también las combinaciones de variables en cada modelo (donde **NA** nuevamente significa ‘valor no disponible’; o sea, variable no incluida en esa combinación).

Selección de los modelos posibles, a partir del basal

V1	V2	V3	V2:V3	df	logLik	AICc	delta	weight
NA	0.56	0.00	0	5	764.6	-1518.8	0.00	0.47
0.00	0.56	0.00	0	6	764.9	-1517.3	1.55	0.22
NA	0.56	NA	NA	3	761.3	-1516.5	2.30	0.15
0.00	0.56	NA	NA	4	761.7	-1515.2	3.64	0.08
NA	0.56	0.00	NA	4	761.3	-1514.4	4.41	0.05
0.00	0.56	0.00	NA	5	761.8	-1513.1	5.73	0.03
NA	NA	NA	NA	2	-107.3	218.7	1737.55	0.00
-0.03	NA	NA	NA	3	-107.0	220.2	1739.06	0.00
NA	NA	0.03	NA	3	-107.2	220.5	1739.28	0.00
-0.03	NA	0.03	NA	4	-106.9	222.2	1740.98	0.00

Pero sucede que Delta nos alerta que no es correcto asumir un único ‘mejor modelo’, pues la diferencia de los valores de AICc en la pareja superior no es mayor que 2. De nuevo tenemos que seguir trabajando; ¡hay que promediar los dos primeros!

Pero, ¡nada por qué preocuparse!... ¡Ahí está nuestro ‘genio de la lámpara!... Veamos los resultados del ‘modelo promedio’ que él nos devolvió con una rapidez impresionante; están presentes las estimaciones de los coeficientes (‘estimate’) y el valor p (‘p.value’) de la significación estadística de esas estimaciones.

Estimaciones y p_value del Modelo Promedio, según basal

term	estimate	std.error	statistic	p.value
(Intercept)	-0.17	0	-1.240e+15	0
V2	0.56	0	4.186e+15	0
V3	0.00	0	8.230e+11	0
`V2:V3`	0.00	0	-5.153e+12	0
V1	0.00	0	-7.973e+12	0

En principio, no difiere mucho (en este caso) del ‘modelo basal’ que asumimos; sin embargo, se trata de un nuevo modelo, porque ahora todas estimaciones de efectos de las variables predictoras sí son estadísticamente significativas, lo que excluye la posibilidad de un ‘sobre ajuste’ del modelo a los datos. Los valores de los efectos (sin redondear a las centésimas) de V2 y V3 son 0.556 y 0.0000722, respectivamente; ambos son positivos luego ‘trabajan’ a favor del crecimiento del puntaje promedio de theta.

Definitivamente, el comportamiento profesional de los profesores que imparten la asignatura evaluada habla a favor de un mejoramiento de los resultados de la calidad de su aprendizaje, aunque el efecto de V3 (impregnado de la percepción de los cuadros institucionales acerca de ese desempeño) es muy bajo. En cambio, las opiniones de los estudiantes sobre la calidad de su propio aprendizaje en esa asignatura no son muy certeros (o no fueron bien preguntado, como dijimos antes).

Bueno, ¿qué?... ¿Terminamos?... Pudiera parecer suficiente, pero tenemos aún un inconveniente metodológico que atender. Y es que ya hemos hablado en esta Serie de los inconvenientes de depender solo de una muestra de ocasión. En rigor, lo que más interesa a la ciencia no son los resultados investigativos aislados, sino la consistencia de los hallazgos relacionados con ellos.

Aprovecho así para nuevamente llamar la atención que una verdadera teoría científica emerge de resultados prácticos, cuando menos, similares en muchas investigaciones aplicadas, ejecutadas en momentos y en lugares distintos. ¡Desconfiemos de las construcciones teóricas que suelen ser denominadas ‘científicas’ y que, en realidad, son el resultado de reiteradas reflexiones teóricas, basadas más en el pensamiento especulativo y empírico (propio de los sujetos), que en datos emanados directamente de la práctica y replicados suficientemente!

De modo que, daremos un paso más hacia adelante e intentemos escapar de los efectos engañosos de una única muestra, por más que la consideremos aleatoria y representativa de la población de la cual se extrajo. Pero, ¡tranquilos!... Ahí tenemos a R-Project listo para ayudarnos en ese otro paso del proceso investigativo.

Para replicar los resultados de nuestro ‘modelo promedio’ vamos a utilizar una técnica muy cercana al primer nivel de la Inteligencia Artificial, al Machine Learning (o aprendizaje autónomo de los equipos de cómputo). Ella es conocida como validación cruzada (‘crossvalidation’).

En esencia, consiste en dividir la base de datos en dos grandes sub-bases de datos (a razón de 80% y 20%, o de proporciones similares). La parte mayor la utilizaremos para el ‘entrenamiento del modelo’ (training) y la segunda para su validación definitiva (testing). Al mismo tiempo, ambas sub-bases de datos son divididas en ‘pliegos’ (‘fold’) que van cambiando constantemente tras replicar el proceso varias veces con cada uno de ellos.

En nuestro código de **R** programamos esta parte del proceso de manera que se dividiera la sub-base de entrenamiento en cinco ‘pliegos’ y con cada uno de ellos se ejecutarán cincuenta corridas del modelo promedio, antes de someterlo al ‘pliegue’ de testeo.

En la siguiente tabla se pueden apreciar los quince primeros parámetros de calidad del modelo generados con los resultados de las repeticiones sucesivas.

Parámetros del entrenamiento del modelo, al remuestrear la muestra particionada

RMSE	Rsqared	MAE	Resample
3e-16	1	3e-16	Fold1.Rep01
1e-16	1	1e-16	Fold2.Rep01
2e-16	1	2e-16	Fold3.Rep01
3e-16	1	2e-16	Fold4.Rep01
3e-16	1	2e-16	Fold5.Rep01
4e-16	1	3e-16	Fold1.Rep02
7e-16	1	5e-16	Fold2.Rep02
3e-16	1	3e-16	Fold3.Rep02
3e-16	1	3e-16	Fold4.Rep02
4e-16	1	3e-16	Fold5.Rep02
3e-16	1	3e-16	Fold1.Rep03
6e-16	1	5e-16	Fold2.Rep03
2e-16	1	2e-16	Fold3.Rep03
3e-16	1	2e-16	Fold4.Rep03
4e-16	1	3e-16	Fold5.Rep03

El promedio y desviación estándar de las 250 repeticiones, por ejemplo, del parámetro de calidad R-cuadrado son impresionantemente perfectos.

Resúmenes de la validación cruzada del Modelo Promedio

mean_R2	sd_R2
1	0

Entiéndase bien, hemos simulado con **R** la repetición de la fijación de nuestro ‘mejor modelo’ (asumido como promedio de los dos mejores candidatos), con 250 muestras más, a parte de la ocasional (o sea, de la elegida realmente) y hemos podido apreciar que la calidad del modelo resulta suprema; luego el modelo y las estimaciones de los coeficientes de sus variables explicativas son muy confiables. Tenemos resultados con una elevada significación estadística (y, por tanto, con una alta posibilidad de generalización a nuevos datos), así como un control

preciso del error que se pudo haber cometido al obtener esas estimaciones, y todo con datos obtenidos directamente de la práctica educativa.

Bueno, este es el final del **Número 5**. Seguiremos en unos días con un nuevo tema sobre los procesos universitarios 'sustantivos'. ¡Los espero!...



Dr. Cs. Paul Antonio Torres Fernández

Profesor e Investigador Titular; Bioestadístico

Facultad de Ciencias Médicas 'Salvador Allende'

Universidad de Ciencias Médicas de La Habana

orcid.org/0000-0002-7862-2737

Para profundizar:

Torres, P. A. (2022a). *Serie: ¿Cómo se comporta la investigación educativa en Cuba, cinco años después? Los grandes aliados de la investigación cuantitativa en el análisis de los datos: la estadística multivariante y el entorno 'R-Project' (1º parte)*.

https://drive.google.com/file/d/1WG7AzX0Sq_OFt1w15SbbYcJZigcr9bKR/view?usp=sharing

_____ (2022b). *Serie: ¿Cómo se comporta la investigación educativa en Cuba, cinco años después? Los grandes aliados de la investigación cuantitativa en el análisis de los datos: la estadística multivariante y el entorno 'R-Project' (2º parte)*.

https://drive.google.com/file/d/1_5wL4P-o19IUOIMhk6J6PgXKdf_P4uQr/view?usp=sharing

Creado con **Canva**