

Sumando Peras y Manzanas: Análisis de los Tests de Admisión Universitarios

Jorge Eduardo Mahecha Gómez

jorge.mahecha@udea.edu.co

Profesor, Instituto de Física, Universidad de Antioquia

Medellín, 18 de enero, 2016

“Las joyas no se pesan en una báscula de tienda”

Carl Friedrich Gauss

Resumen

Los *rankings* de todo tipo están de moda. Se soportan en la creencia de que la actividad humana es susceptible de medirse, que cada individuo y cada institución deben adelantar un mejoramiento continuo de su ubicación en el *ranking*. La educación y la investigación son algunas de las actividades más afectadas por la presión del *ranking*. El MIDE del Ministerio de Educación; la clasificación de grupos de investigación e investigadores por Colciencias; *rankings* de revistas, artículos y libros; las pruebas Saber (3°, 5°, 9°, 11), Saber Pro; exámenes de admisión a las universidades; acreditación de “alta calidad”, registro calificado, transformación curricular, Sistema de Aseguramiento de la Calidad ... En el ámbito internacional: Los *rankings* de universidades del tipo Shanghai, los del “impacto” de las revistas, Test of English as a Foreign Language (TOEFL), The Graduate Record Examination (GRE), las Pruebas Pisa, etc. Suponen que el intelecto se puede “medir”, estandarizan, califican con base en un número único que se obtiene al “Sumar Peras y Manzanas”. Luego usan el número para “repartir, por orden descendente, la calidad” de instituciones, programas académicos, grupos de investigación y personas. En el presente texto se discute la imposibilidad técnica de dichas pretenciones, haciendo énfasis en la elaboración de un *ranking* para escoger las personas que merecen ingresar a la Universidad de Antioquia y descartar a quienes no lo merecen. Busca dar respuesta a las siguientes preguntas: (1) ¿El sistema permite hacer un *ranking* válido de los aspirantes, basado en los puntajes obtenidos? (2) ¿El sistema realmente es capaz de establecer una frontera válida entre los puntajes de los admitidos y los no admitidos?

1 La suma de números de diferente naturaleza

La suma de 3 Manzanas más 2 Peras da, aparentemente, “5”. ¿Qué representa realmente ese “5”? Es algo que no tiene conexión con los objetos que lo originaron, porque puede obtenerse de $5M + 0P$, $4M + 1P$, $3M + 2P$, $2M + 3P$, $1M + 4P$, $0M + 5P$. Por lo tanto ni siquiera son 5 nuevos objetos. Si esa suma es extraña, igualmente lo es el promedio. En cualquiera de las parejas vale 2.5.

¿Cómo es posible que el promedio de 5M y 0P, por ejemplo, valga lo mismo que el de 0M y 5P? Si se acepta el promedio, nada impide hacer ponderación con factores de peso diferentes de 0.5. Alguien puede decir: Las Manzanas merecen un factor de peso mayor que las peras, porque una manzana fue lo que le dio Eva a Adán, o porque la manzana representa una marca de celulares muy caros. ¿Entonces qué peso asignarles? ¿Tal vez 60% a las manzanas y 40% a las peras? ¿O, mejor aún, 69.89% a las manzanas y 30.11% a las peras? El promedio simple de Manzanas y de Peras da 2.5, y si se usan pesos, se obtiene $0.6989 \times 3 + 0.3011 \times 2 = 2.6989$, o cualquier otro número que cambiará cuando se cambien los factores de peso. En el caso del *ranking*, los pesos son una propiedad del “modelo”, definidos caprichosamente por quien lo diseña, al cambiarlos cambiará el ordenamiento, el cual, por lo tanto, no expresa cualidades intrínsecas de las personas o instituciones que se pretenden clasificar, sino los prejuicios del autor del “modelo”. Quien está dispuesto a sumar Peras y Manzanas, querrá igualmente multiplicar Manzanas por Peras, incluso elevar Peras al cuadrado, o calcularles el logaritmo.

Dado que $\ln(x+1) = x - x^2/2 + x^3/3 + \dots$, usando una expansión en series, si x fuera cierto número de peras, $\ln(x+1)$ daría una cosa formada por peras, peras al cuadrado, peras al cubo, etc. Si x fuera “metro” se obtendría “metro - metro²/2 + metro³/3 + ...”. Es decir, sumar metro lineal, con metro cuadrado, con metro cúbico, ... La observación no es irrelevante: En el centro del “modelo” de clasificación de grupos de investigación por Colciencias está el cálculo de $\ln(x+1)$ donde x es “cierto número de Peras”, y el “1” no tiene unidades, es una Manzana.

La suma 2 yenes + 3 euros si se puede hacer, incluso evaluar el promedio, porque las monedas son convertibles. Sin embargo, la imposibilidad de calcular “log yen” subsiste. Es similar a la de calcularle el logaritmo a la unidad metro. Igualmente, 2 millas + 3 kilómetros no da “5”, pero con las conversiones adecuadas de unidades se obtiene 3.86 millas o 6.22 kilómetros. La imposibilidad de sumar Peras y Manzanas está fundamentada en el análisis dimensional. “Manzanas” y “Peras” tienen “unidades” diferentes. No está definido cómo convertir Manzanas en Peras o Peras en Manzanas. Sumar Peras y Manzanas, Elefantes y Libros, ... es como sumar metros y kilogramos.

No tiene sentido, por lo tanto, decir que el aspirante que respondió 2 preguntas correctas de “lógica” y 3 de “lenguaje” merece la misma calificación que quien respondió 3 preguntas correctas de “lógica” y 2 de “lenguaje”, o 5 preguntas correctas de “lógica” y 0 de “lenguaje” ... Una pérdida de información (que puede expresarse con la desviación estándar) es inevitable cuando se reemplazan las calificaciones originales por el promedio.

2 ¿La “medición” del talento para qué?

Se pretende darle soporte “objetivo” a los prejuicios asignando números¹. El asunto esencial es: ¿Cómo llevar los prejuicios al terreno normativo, administrativo o disciplinario sin ser acusados de cometer arbitrariedad o de violar los derechos humanos, más aún, reclamando soporte “técnico” y “científico”? En el caso de la educación y la investigación se tiene un recurso muy astuto: Sustentar la discriminación con una supuesta medición de la “calidad de los individuos”, específicamente de “las capacidades”. Se dice que en este terreno no hay lugar para la democracia, que definitivamente no todos son iguales, que la investigación y la educación universitaria requieren personas altamente talentosas y por lo tanto se necesita, de manera prioritaria, desplegar medios para “medir” los talentos. El advenimiento de la gran industria a partir del primer cuarto del siglo XX estuvo emparejado con el desarrollo de la estandarización para producir objetos con calidad similar y susceptibles de fabricarse con obsolescencia programada. Por otra parte, la gran industria se desarrolla por medio de líneas de producción que siguen rígidos protocolos. ¿Si fue posible la estandarización de la calidad en la gran industria, por qué no hacer lo mismo con la “calidad” de las personas? Este propósito brinda un beneficio político colateral nada despreciable: Una sociedad formada por individuos maquinizados en vez de ciudadanos completos que puedan pensar por sí mismos. La medición del “talento” es un filtro que permite ajustar el currículo de educación a lo que las empresas y el comercio requieren. No importa que eso implique desestimar la importancia de los estudios tradicionales de ciencias básicas, humanidades y artes. La educación al servicio de las corporaciones simplemente necesita las habilidades básicas de la alfabetización y la aritmética, que están al alcance de todos y cada uno de los seres humanos. “Las corporaciones no necesitan sabios”. Sólo necesitan personas: (1) Dispuestas a someterse a la disciplina corporativa, (2) Dispuestas a fabricarse altos “indicadores” del “talento” estandarizado sin importar los medios requeridos, (3) Libres de compromisos y distractores que les impidan dedicarse por completo a los fines de las corporaciones. Son estas tres cosas y no “el talento” ni “las capacidades” lo que importa. Lo que sucede es que no es políticamente correcto reconocerlo, es prioritario camuflarlas en las supuestas medidas técnicas de “la calidad”.

Claudine Haroche (en *O Inavaliável em uma Sociedade de Desconfiança*), presenta citas de Yves Charles Zarka (de *Qu'est-ce que Tyranniser le Savoir?* y *L'évaluation: un Pouvoir Suposto Saber*). Refiriéndose a las mediciones del saber, escribió: “Zarka menciona que “la evaluación es siempre subjetiva y relativa”, pero que esta “subjetividad ... debe permanecer escondida” y que, por este motivo, “procura esconderse detrás de una matemática”. Insiste sobre el hecho de que “la evaluación (...) mide otra cosa”. Pensemos que esa otra cosa corresponde, entre otras, a la parte imposible de observar que protege la interioridad del individuo: La evaluación realmente realiza ... “una inquisición de la

¹Véase: Normose, a doença da “normalidade” no mundo acadêmico, Renato Santos de Souza, 01/07/2014. <http://www.uninomada.co/inicio/index.php/capcog/108-lanormosis-academica>.

*interioridad”, con el fin de gobernar tanto a los individuos como sus comportamientos, actuaciones y reflexiones,...”*².

3 El *ranking* de grupos de investigación de Colciencias

Actualmente está abierta la convocatoria de Colciencias para establecer *rankings* de grupos de investigación y de personas que participan en investigaciones. En la p. 69 del llamado documento conceptual de la medición de grupos, <http://goo.gl/m9BJYC>³, se da una fórmula para determinar la “calidad”, basada en los “indicadores de producción” que se definen a partir de la p. 58, un ejemplo clarísimo de Suma de Peras y Manzanas.

La “calidad” sería función de los diferentes productos, los grados de divulgación y apropiación social, cohesión y cooperación, en el caso de los grupos. Definen indicadores de productos de “nuevo conocimiento” de tipos A1, A y B; de formación de investigadores de tipos A y B; de divulgación y apropiación social; en fin, de cohesión y cooperación. Asignan factores de peso: 40% a documentos de tipo A1, 25% de tipo A y 10% de tipo B; 10% a formación A y 5% a formación B; 2% a divulgación y apropiación social; 4% a cohesión interna de los grupos y 4% a cooperación entre grupos. Véase también la p. 70. Como si fuera poco la ponderación de Manzanas, Peras, etc., se atreven a asignarles pesos caprichosos. Además se inventan algo extremadamente curioso: Los índices de “producción” y “apropiación-divulgación” los determinan luego de evaluar logaritmos. En la p. 57 del mencionado documento de Colciencias se lee que el “conteo escalado de los productos en cada subtipo ... se denota por $\ln()$ la función logaritmo natural, que es usada para escalar los valores de la producción ...”. Luego aparece la fórmula $\ln(\text{Número de productos en el subtipo} / \text{Período de observación} + 1)$. Como comentó el profesor Carlos Vasco⁴: Comete “... el error matemático de sumar los índices de peras con manzanas para obtener una letra “A” vacía de sentido”. La Suma de Peras y Manzanas aparece dos veces, en el argumento de la función logaritmo y al totalizar el índice del grupo con base en las mencionadas ponderaciones.

El argumento de la función logarítmica, $\{\text{Número de productos en el subtipo}\} / \{\text{Período de observación}\}$, es una cantidad con unidades, cuyo valor depende de su escogencia; son las Peras. En tanto que “+1”, aparentemente, es un número puro, sin unidades, representa las Manzanas. Los autores definen “período de observación” igual a cierto número de años. Si la unidad de tiempo es 10 años, un grupo con 10 productos del subtipo dado en ese período, obtendrá

²http://www.scielo.br/scielo.php?pid=S1517-97022011000300014&script=sci_arttext. Los enlaces fueron consultados el 16 de enero de 2016.

³*Modelo de Medición de Grupos de Investigación, Desarrollo Tecnológico o de Innovación y de Reconocimiento de Investigadores del Sistema Nacional de Ciencia, Tecnología e Innovación, Año 2015*. Colciencias, Bogotá, 18 de diciembre de 2015.

⁴Mencionado en *El veto a Colciencias*, Jorge Iván González, periódico La República, 10/09/2015.

un “conteo escalado” de $\ln(10 + 1) = 2.39$. Pero si la unidad de tiempo hubiera sido 1 año, tendría 10 productos en 10 años y el “conteo escalado” valdría $\ln(1 + 1) = 0.69$. Esta escogencia puede hacer cambiar el ordenamiento. Una demostración de por qué no es bueno sumar Peras y Manzanas. Es triste que los directivos de las universidades (incluidos los de la Universidad de Antioquia), aceptaron el requerimiento de Colciencias de crear una burocracia paralela para aportar los datos necesarios “para que Colciencias les evalúe el logaritmo” ...

4 Modelo de admisión de la Universidad de Antioquia

El modelo de admisión a la Universidad de Antioquia se basa igualmente en Sumas de Manzanas y Peras. En sumas de diferentes “talentos”. Postula:

- Que en cada ser humano yace un objeto llamado “calidad académica”, el cual no tiene el mismo tamaño en todas las personas.
- Que dicho tamaño se expresa con un número único (el “índice de calidad”), el cual no se puede determinar en forma directa sino a través de un conjunto de “medidas”.
- Que la “calidad académica” es un objeto compacto de 5 diámetros o talentos (“lógico”, “lingüístico”, “biológico”, “humanista” y “matemático”). En los tests de admisión de la Universidad Nacional de Colombia y en los tests Saber se definen otros.
- Que no es necesario medir todos los 5 “diámetros”, sino sólo 3 que dependen de la carrera elegida por el aspirante.
- Que el tamaño de la “calidad académica” se puede definir como el promedio ponderado de los “talentos” escogidos.
- Que estas “mediciones” no requieren someterse a protocolos de metrología.
- Que los números obtenidos mediante las “mediciones” y los cálculos de los promedios no tienen errores, lo cual permite que todas las cifras enteras y decimales obtenidas a partir de los “índices de calidad” sean significativas.
- Que a las carreras universitarias sólo deben ingresar personas cuyo “índice de calidad” sea suficientemente alto.
- Que dicha suficiencia varía de una carrera a otra.

El puntaje del último aspirante admitido se conoce como Puntaje de Corte. En el pregrado de psicología (cohorte 2016-1) el puntaje de corte fue 73.245 (Puntaje Estandarizado obtenido a partir del promedio de los 3 diámetros,

“lógico”, “lingüístico”, “humanista”), eso significa que los admitidos, 24, obtuvieron más de 73.245 en calificación estandarizada⁵. Admitieron sólo 1 de cada 109. Similarmente en medicina, 77.099, instrumentación quirúrgica, 74.039, ingeniería civil, 71.471, microbiología, 72. En medicina admitieron 1 de cada 70 y en ingeniería civil 1 de cada 31.

En un programa de Ingeniería de Telecomunicaciones, por ejemplo, el puntaje de corte fue de sólo 53.356, 24 puntos por debajo del de medicina, localizado entre los más bajos de todas las cohortes 2016-1. Allí se inscribieron 159 aspirantes en primera opción y 365 en segunda opción, fueron admitidos 31 y 6, respectivamente; por primera opción, 1 de cada 5. Si el puntaje guardara relación con la calidad, eso significaría que todos los alumnos de medicina tendrían “calidad” superior a 77 y todos los de Ingeniería de Telecomunicaciones en la Sede Medellín solamente superior a 53 (según los autores del modelo, estos admitidos serían personas de “pésima calidad” comparados con los de medicina). Posiblemente la mayoría de los admitidos en Ingeniería de Telecomunicaciones lo fueron con un puntaje menor al del “más malo” de Medicina. Esto permitiría afirmar que si el puntaje realmente mide el talento, ese 1% de admitidos a medicina estaría formado por personas “**geniales**”. Similarmente, Zootecnia tuvo la misma prueba específica que Medicina, de 288 aspirantes por primera opción fueron admitidos 7, y 47 por segunda opción entre 1409, el puntaje de corte fue de 62.82. Lo anterior permite la observación de que los programas en los cuales la relación entre admitidos y aspirantes es más baja se están “quedando” con los aspirantes “buenos” y los programas menos demandados con los “malos”; esto pone en entredicho el mito de que el puntaje guarda relación con la “calidad”, incluso la existencia del objeto “calidad académica”.

Los llamados exámenes de admisión y exámenes del ICFES presuponen la existencia del objeto llamado “calidad académica”. En el sistema de admisión a la Universidad de Antioquia se realiza un **promedio ponderado** de los puntajes para obtener la calificación global de la “prueba”. Cada “diámetro” toma 26 valores enteros, entre 0 y 25, por lo tanto se tienen 17576 conjuntos diferentes de 3 “diámetros”, los cuales al promediarse darán un número total de “calidades académicas” posibles diferentes que ciertamente será menor que 17576. En la prueba “perfecta” los 3 “diámetros” valen 25, la “esfera”, la más imperfecta tiene tamaño cero, el “punto”, el resto serán volúmenes tridimensionales de formas muy variadas. El “tamaño” del objeto lo definen como el promedio ponderado de los 3 “diámetros”. La determinación del tamaño de un objeto tridimensional por medio de la suma ponderada de los diámetros tiene limitaciones inevitables, por cuanto no se conoce la forma del objeto cuyo tamaño se quiere determinar. Ciertamente tres mediciones permitirían calcular el volumen de un paralelepípedo (el producto de los tres lados), incluso de un elipsoide (el producto de los tres semiejes); ¿pero qué decir de un objeto que tenga la forma de una botella, o de un bumerán? Es inevitable admitir que las medidas de los 3 “diámetros” tienen diferencias respecto al promedio que merecen caracterizarse

⁵Valores obtenidos en Datos_Históricos_Admisión_Examen_20131_20161.xlsx, Suministrado por el jefe de Admisiones de la Universidad de Antioquia el 4 de diciembre de 2015 en respuesta a un derecho de petición.

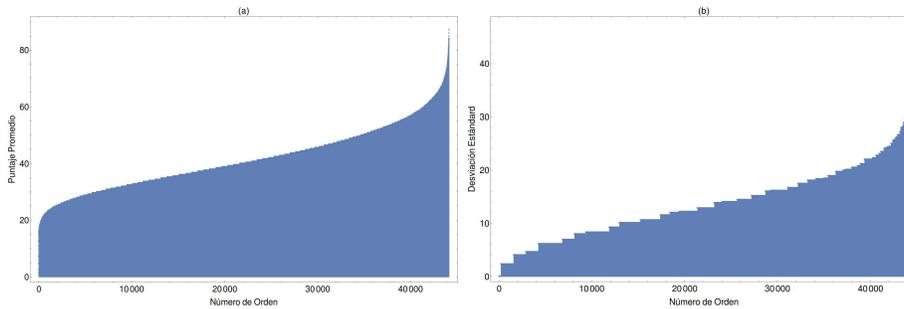


Figura 1: Resultados en el proceso de admisión de la cohorte 2016-1, que implementa el “modelo” AA 480. (a) Los puntajes promedios de los 3 tests de cada uno de los aspirantes. (b) Las desviaciones estándar de los puntajes de los tres tests de cada individuo respecto a su promedio ponderado.

mediante la **desviación estándar ponderada**⁶. Se reconoce aquí el facilismo de aceptar la estadística de manera amañada: Se adopta el promedio pero se ignoran las fluctuaciones alrededor del promedio.

5 Reconocer lo microscópico a partir de lo macroscópico

El problema de indagar por las capacidades tiene alguna analogía con los estudios de Mendel, a mediados del siglo XIX, acerca de herencia e hibridación en vegetales. Mediante observaciones macroscópicas llegó a inferir la existencia de elementos microscópicos de los cuales se derivaron las observaciones. Son los genes, elementos microscópicos ocultos en las plantas que estudió. 100 años más tarde, Crick, Franklin y Watson descubrieron la estructura del ADN de las células que permitió el soporte microscópico de las observaciones de Mendel. Los genes corresponden a la ubicación de las bases nucleicas presentes en la doble hélice de moléculas de ADN.

El objeto llamado “calidad académica” sería microscópico y de naturaleza desconocida. El “índice de calidad” sería su contraparte macroscópica.

Mendel trabajó con alverjas, no con “promedios de Peras y Manzanas”.

La figura 1 muestra la totalidad de los resultados promedios de los tests de admisión de todos los 44145 aspirantes⁷. Se basa en los datos contenidos en el archivo `Matriz-respuestas-correctas.xlsx`⁸. Las entradas del archivo son: 1 Número de orden, 2 Correctas Competencia Lectora, 3 Correctas Razonamiento Lógico, 4 Correctas Prueba Específica, 5 Nombre de la Prueba Específica. “Correctas” es un número entero entre 0 y 25.

⁶Incluso con la entropía de Shannon. Véase el Apéndice.

⁷Mendel no llegó a trabajar con números tan grandes de muestras.

⁸Suministrado por el jefe de Admisiones de la Universidad de Antioquia el 16 de diciembre de 2015 en respuesta a un derecho de petición.

Se observa que unos 4000 aspirantes obtuvieron un puntaje superior a 55% y unos 10000 inferior a 35%. Por lo tanto, unos 30000 obtuvieron puntajes entre 35% y 55%. En cuanto a la desviación estándar de los puntajes de cada individuo en los tres tests, se observa que cubre un rango muy amplio, entre 2 y 50; la de unos 1000, vale 2; la de unos 1000, es superior a 25; la de más de 30000 tests está entre 8 y 24. En todos los casos es muy superior a la distancia mínima entre dos promedios, que vale 0.4. Esta distancia mínima es muy inferior a la mínima de las desviaciones estándar.

La respuesta a cada pregunta posee incertidumbre, la cual no es fácil de determinar. En consecuencia, la calificación global estará afectada por dos incertidumbres: Las inherentes a cada “prueba” (“lógica”, “lenguaje” y “específica” según el AA 480), y la proveniente de la mencionada desviación estándar.

El factor estadístico de corrección de Student para muestras pequeñas, amplifica la desviación estándar con un factor que se aumenta si se quiere un mayor nivel de confianza (por ejemplo 95% o mayor) en la determinación de la cantidad de “talento” de cada candidato.

No es válido comparar las medias sin tomar en cuenta la desviación estándar de las notas de las diferentes pruebas y la incertidumbre de las notas de cada prueba. Un “criterio de resolución” para decidir si una media está después de otra consecutiva, es que la distancia entre las dos medias sea mayor que la suma de sus correspondientes desviaciones estándar (multiplicadas con el factor estadístico de Student y divididas por la raíz cuadrada de 2 o de 3 según el caso). Véanse las figuras 2 y 7. Similarmente, un criterio para decidir si una calificación es aprobatoria, es que la distancia de la correspondiente media al umbral sea mayor que la incertidumbre. Por lo tanto, se puede concluir que las desviaciones estándar de los puntajes de los aspirantes son altas, lo cual hace que el modelo de ponderación no sea apropiado para el fin de establecer el *ranking* de los aspirantes según su grado de “calidad”.

El azar tiene un papel preponderante en los tests de escogencia múltiple, por lo tanto contribuye a aumentar la, ya alta, incertidumbre del *ranking*.

6 Estrategias del juego de azar llamado “escogencia múltiple”

Consideremos preguntas típicas de escogencia múltiple con 4 opciones y lo que hará el aspirante. De las N preguntas, responder las x cuya respuesta pueda obtener con probabilidad cercana a 1. Para las $N-x$ restantes, usar plausibilidad. La simple escogencia al azar da una probabilidad de $1/4$ de acertar una respuesta. Por lo tanto el número de respuestas que acierta el aspirante será $x + (N-x)/4$. Si $N=12$ y $x=4$, logrará a lo sumo 6 respuestas.

Si es posible con probabilidad cercana a 1 descartar un “distractor” y con probabilidad p mayor que 0 descartar un segundo “distractor”, se pueden escoger los dos restantes con probabilidad $1/2$ sin tener idea de cual es el correcto.

La probabilidad de acertar una pregunta con este procedimiento varía entre $1/3$ (el peor caso, $p=0$) y $1/2$ (si $p=1$), que ciertamente es mayor que $1/4$, lo cual permite subir el número de respuestas correctas de 6 a 8. ¿No está mal, verdad? Sin tener seguridad en ninguna respuesta ($x=0$), el azar completo permite acertar 3 y el “azar inteligente” entre 4 y 6, en un test de 12 preguntas.

Como al calificar no se sabe si una respuesta correcta provino del “conocimiento” o del azar, puede concluirse que de y respuestas correctas hay x que provinieron del “conocimiento” y las $N-x$ restantes están afectadas por el azar. $y = x + q \times (N-x)$, donde q puede valer entre $1/4$ y $1/2$, digamos que $1/4$ y N es el número total de preguntas. Se tiene una contribución proveniente del azar dada por $(N-x)/4$, la cual vale 0 si $x=N$ y $N/4$ si $x=0$. Un estimativo se logra tomando $x=N/2$, en cuyo caso el número de respuestas provenientes del azar es $N/8$. En un test con 25 preguntas, esto significa que típicamente 3 de las respuestas correctas provienen del azar. Si $q=3/8$ este número se elevaría a 4 o 5, y si $q=1/2$ a 6.

Todos los aspirantes tienen igual oportunidad de beneficiarse del azar. Pero no se diga que el procedimiento se dirige a seleccionar según la “calidad”. Ni que logra establecer un *ranking* válido. Si la diferencia mínima posible entre dos puntajes en la escala 0-25 del AA 480 es de 0.1, y en la escala 0-40 del AA 236 de 0.5, y los errores debidos al azar (en cada test y, por lo tanto, en el promedio) son mayores que 5, la contribución del azar es entre 10 y 50 veces mayor que la diferencia mínima entre puntajes, sólo por eso es imposible usar esos puntajes para ordenar según la “calidad” y decidir quién ingresa y quién no a la universidad.

El procedimiento anterior tiene problemas metodológicos adicionales. El más importante, es la atrevida afirmación de que la “calidad” se puede identificar con el puntaje. Además el índice de “calidad” se halla promediando Peras con Manzanas: Los indicadores de la calidad en “lógica”, “lenguaje” y “conocimiento específico” (el AA 480 agrega Yucas). El uso de la “curva” es irrelevante para la conclusión, lo mismo que los factores de peso de 30%, 30% y 40% asignados caprichosamente a las pruebas (exaltados como un gran logro por los autores).

7 La mayor fuente de incertidumbre

La Suma de Peras y Manzanas está inmersa en cada uno de los 3 tests. Por lo tanto existe incertidumbre por la diversidad de las “preguntas”, el error de las mediciones de los “talentos” se incrementa.

El test perfecto de 25 preguntas se obtiene de 1 única manera: 1,1,...1, o sea 25 respuestas correctas. La totalmente imperfecta se obtiene de 1 única manera: 0,0,...0, o sea 25 respuestas incorrectas. La complejidad se manifiesta con solo 24 respuestas correctas. 1 incorrecta, ¿cuál? No es lo mismo ignorar la respuesta acerca de integrales, que la de aritmética modular, o la de propiedades del triángulo equilátero,... Hay 25 posibilidades. La situación se torna muy complicada al aumentar las correctas. El número de maneras diferentes de

responder 13 correctas es

$$\binom{25}{13} = \frac{25!}{13!12!} = 5'200.300,$$

¿Esas 5 millones de maneras son equivalentes? Reemplazarlas por el promedio implica la penalidad de aceptar que el promedio de respuestas correctas, $m = 13/25 = 0.52$ está afectado por una gran incertidumbre.

La incertidumbre se puede caracterizar por la varianza:

$$\Delta m = \sqrt{\frac{13(1-m)^2 + 12(0-m)^2}{25}} = 0.5.$$

Esto es grave: La calificación por el único número, 0.52, debe reemplazarse por un intervalo. Está entre 0.0 y 1.0 (redondeo con 1 cifra decimal). Conceptualmente, dado que el diseñador de los tests escogió un conjunto de 25 temas que los aspirantes deberían conocer (obviando que se esté erigiendo en supremo juez calificador-descalificador del éxito de los procesos formativos de personas muy diversas durante 11 años), al reemplazar las respuestas a las preguntas diferentes por un único número, el “puntaje”, está anulando su altruista intención de verificar el aprendizaje de los 25 temas. La venganza de la estadística es la incertidumbre completa, porque dada la desviación estándar, no será capaz, con validez metrológica, de asignarle un puntaje al alumno.

Y no debería enojarse por el resultado, porque su intención inicial fue la de fijar los “contenidos mínimos” que cada bachiller debería conocer, y saber sólo 13 de los 25 posibles sería, supuestamente, una grave limitación. Si acepta la diversidad, el patrón de respuestas correctas sería indicador de la misma: Cada uno de los 50000 aspirantes tendría un patrón propio y, además, ¡válido!, porque no se necesita y es imposible que todos sean iguales. Por el contrario, es muy bueno reconocer que todos son diferentes.

El método de las encuestas de cada conjunto es inherentemente probabilístico. No es posible diferenciar entre las siguientes razones para obtener una respuesta “correcta”: (1) Usar conocimientos y capacidad de análisis. (2) Haberse sometido a programas de entrenamientos dirigidos a desarrollar la capacidad de responder en algunos minutos cada pregunta. (3) Al azar (nótese que el instructivo entregado recomienda no dejar respuestas sin marcar). Por lo tanto cada puntaje posee incertidumbre.

En el presente caso, cada calificación se obtiene mediante un promedio ponderado de los puntajes obtenidos en 3 tests. En Saber 11 el número de tests es mayor, lo cual aumenta la desviación estándar. No es válido comparar las medias sin tomar en cuenta la desviación estándar de las notas de los diferentes tests y la incertidumbre de las notas de cada test.

El test específico en “biología” no cubre un cuerpo cerrado de saberes, sino áreas diversas. Lo mismo puede decirse de cada uno de los otros tests. La totalización de las respuestas que conforman un test conlleva las dificultades de la Suma de Peras y Manzanas. No hay diferencia esencial entre totalizar las preguntas de “biología”, y promediar las calificaciones obtenidas en “lógica”,

“lenguaje” y “conocimientos específicos”. En esencia, un test de admisión formado por 120 preguntas, es un muestreo de todos los saberes a través de 120 temas (se supone que el diseñador de los tests no hará dos preguntas similares). Es irrelevante clasificar las 120 preguntas en 25 de “matemáticas”, 25 de “ciencias naturales”, 25 de “ciencias sociales”, 25 de “lectura” y 20 de “análisis de imagen”. Porque cada pregunta se refiere a una área diferente de las de las 119 restantes. Si es válido definir “cercanía” entre las preguntas, puede formularse una pregunta en “matemáticas” y una en “ciencias naturales” que sean más cercanas entre sí que dos preguntas en “sociales”.

Los saberes no están terminados, se entrecruzan de maneras impredecibles. Su compartimentalización es subjetiva, obedece a conveniencia administrativa. No tiene sentido compartimentalizar y estandarizar los saberes.

La definición de aquello que toda persona debería “saber” podría violentar la libertad de aprendizaje y el derecho a no ser censurado ni discriminado. *Todos somos muy ignorantes. Lo que ocurre es que no todos ignoramos las mismas cosas*, observó Einstein.

¿Qué es “biología”? Ecología, genética, herencia, virus, bacterias, hongos, botánica, zoología, fisiología animal y humana, anatomía animal y humana, astrobiología, ...

A su vez, “zoología” es un campo complejo, que incluye: Taxonomía, anatomía y fisiología comparadas, mastozoología, entomología, malacología, ictiología, ornitología, herpetología, parasitología, veterinaria, zootecnia, paleontología, ... Las subdivisiones siguen hasta el infinito: La “entomología” consta de ...

Un test específico estandarizado en “biología” no tiene sentido: El número de preguntas posible es infinito, necesariamente dependerá de la selección caprichosa del diseñador de los tests, quien escogerá los temas que “sabe” y hará preguntas cuya respuesta conoce⁹.

8 La imposibilidad de “ranquear” los aspirantes

Si se toma el conjunto de aspirantes que presenten las pruebas, no es posible cumplir el declarado propósito de “ordenar” en forma descendente los puntajes para asignar los cupos. Véanse las figuras ilustrativas 6 y 7, en las cuales, por simplicidad, puede suponerse que el máximo de la “campana” asociada a cada aspirante coincide con el puntaje promedio. Se obtuvieron a partir de los datos de la Tabla citada, correspondiente a la cohorte 2016-1.

Aún ignorando las fluctuaciones alrededor de la media, el ordenamiento de las calificaciones no es creíble. Ni con el AA 236 ni con el AA 480. Con el primero se tenían 2 “pruebas” de 40 “preguntas” cada una. En cada una hay 41 resultados, en total $41 \times 41 = 1681$ posibles tests, desde el imperfecto, con 0 aciertos en cada prueba, hasta el perfecto con 40 aciertos en las dos pruebas, pasando por muchos casos intermedios. El puntaje de cada aspirante se basa en el promedio aritmético de los dos puntajes. El número de promedios diferentes es mucho menor, sólo 81. Hay muchos promedios “degenerados”, que

⁹Y por lo tanto será capaz de responder en 144 segundos.

proviene de diferentes puntajes de las pruebas individuales; similar al promedio 2.5 de Manzanas y Peras, que puede obtenerse igualmente con 5 Manzanas y 0 Peras, con 0 Manzanas y 5 Peras, etc. Así, por ejemplo, el promedio 20 entre “razonamiento lógico” y “capacidad de lectura” puede obtenerse de 41 maneras: (40R,0C), (30R,1C), ... (0R,40C). Sólo hay 2 resultados “no degenerados”: (0R,0C) y (20R,20C), cuando todas las respuestas son correctas o cuando todas están mal, en todos los demás hay muchas maneras de obtener un mismo promedio. ¡Una clara sustentación del uso de la varianza! Eso explica que de los 1681 tests, sólo haya 81 diferentes. Pero lo más grave es la diferencia entre dos promedios consecutivos: Da 0.5. Esto exige asegurar que las incertidumbres son mucho menores que 0.5. De hecho, **el modelo supone que los puntajes no tienen incertidumbres.**

Con el AA 480 se tienen 3 “pruebas” de 25 “preguntas” cada una. En cada una hay 26 resultados, en total $26 \times 26 \times 26 = 17576$ posibles resultados, desde el imperfecto, con 0 aciertos en cada prueba, hasta el perfecto con 25 aciertos en las tres pruebas, pasando por muchos casos intermedios. El puntaje de cada aspirante se basa en el promedio ponderado de los tres puntajes. El número de promedios diferentes es mucho menor que 17578, sólo 245. Hay muchos promedios “degenerados”, como en el caso del AA 236. Así, por ejemplo, el promedio 13.0 entre “razonamiento lógico”, “capacidad de lectura” y “conocimiento específico” puede obtenerse de 153 maneras. La diferencia entre dos promedios consecutivos no es fija en este caso, sin embargo en la inmensa mayoría de las veces da 0.1. Esto exige asegurar que las incertidumbres son mucho menores que 0.1. Las desviaciones estándar de los 153 promedios de 13.0 varían entre 4.51 y 12.53, ciertamente mucho mayores que 0.1. Véase la figura 2.

La “degeneración” de los promedios es máxima en 12.7, allí vale 154, su gráfica en función del promedio tiene forma de “campana”, que tiende a 1 en 0 y en 25.

El puntaje promedio 12.9 puede obtenerse de 152 maneras, por ejemplo: a=(25,14,3); b=(14,25,3); c=(4,7,24); d=(15,12,12). El siguiente más alto, 13.0, se puede obtener de 153 maneras, por ejemplo: A=(25,17,1); B=(1,25,13); C=(1,9,25); D=(13,13,13). El aspirante cuyo test sea “a”, es de la misma “calidad” que quien logró “d”, puesto que ambos tienen promedio 12.9. Pero el de “D”, cuyo promedio vale 13, es más “calidoso” que el de “a”, incluso que el que hizo (6,5,24), cuyo promedio también es 12.9; porque, 13.0 es más grande que 12.9.

En algunos datos históricos disponibles se observa que la distribución de los Puntajes Estandarizados totales de los aspirantes, se ajusta muy bien a la función de distribución asimétrica descrita en

https://en.wikipedia.org/wiki/Skew_normal_distribution.

El comportamiento general consiste en un ascenso rápido de la gráfica seguido de una caída más lenta y en puntajes muy altos con mayor probabilidad que los muy bajos.

La figura 3 muestra los resultados de una simulación de la totalidad de los tests en la admisión de la cohorte 2016-1. Se tomó el parámetro de asimetría $\alpha = 2.79412$, el promedio $\mu = 27.6315$ y la desviación estándar $\sigma = 17.3849$.

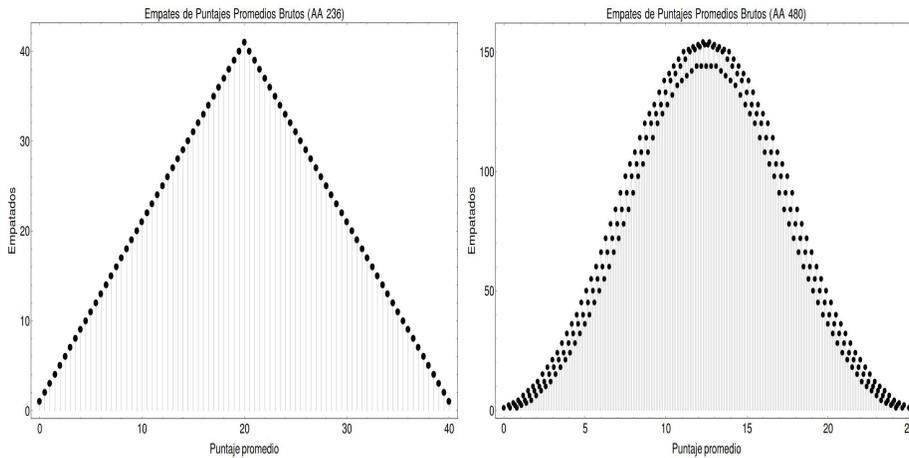


Figura 2: “Degeneración” de los puntajes promedios, según los “modelos” AA 236 (“prueba” de admisión para “medir” dos “talentos”) y AA 480 (“prueba” de admisión para “medir” tres “talentos”). El eje vertical representa el número de maneras diferentes de obtener el mismo puntaje promedio. El test de un aspirante necesariamente está en alguna de esas líneas verticales. Nótese que si las líneas verticales fueran ligeramente más gruesas, no se podría percibir que los posibles puntajes forman un conjunto discreto y por lo tanto no sería posible distinguir dos puntajes consecutivos.

El comportamiento “natural” es el de la figura 2 de la derecha, la cual muestra una distribución simétrica.

Una lista de 23 calificaciones con sus desviaciones estándar:

[81.6, 8.32666]	[81.6, 14.4222]	[82.0, 15.1438]	[82.0, 8.32666]	[82.4, 10.0664]
[82.4, 10.0664]	[82.4, 10.0664]	[82.8, 14.0475]	[82.8, 12.8582]	[83.2, 16.1658]
[83.2, 4.6188]	[83.6, 10.583]	[83.6, 8.32666]	[83.6, 8.32666]	[83.6, 4.0]
[83.6, 6.9282]	[84.0, 12.2202]	[84.0, 4.0]	[84.0, 12.2202]	[84.8, 9.2376]
[86.0, 11.547]	[87.2, 6.1101]	[87.2, 12.8582]		

correspondiente a los 23 promedios más altos en la cohorte 2016-1¹⁰. La figura 4 representa los 23 promedios. Nótese que la repetición de los promedios 81.6, 82.0, 82.4, 82.8, 83.2, 83.6, 84.0, 87.2 da lugar a sólo 10 puntos visibles. Son los puntajes del extremo de la “cola” de la figura 3, donde hay muy pocas repeticiones. Los puntajes de corte se encuentran al inicio de la “cola”, donde es muy probable encontrar puntajes con más de 100 repeticiones. ¿Cómo logran los operarios del modelo resolver los empates en carreras que tienen 100 cupos?

¹⁰En el presente texto se busca simplemente estudiar el modelo como tal. Aquí no se pretende analizar la implementación del mismo, tarea que realizan las comisiones del Consejo Académico de la Universidad de Antioquia.

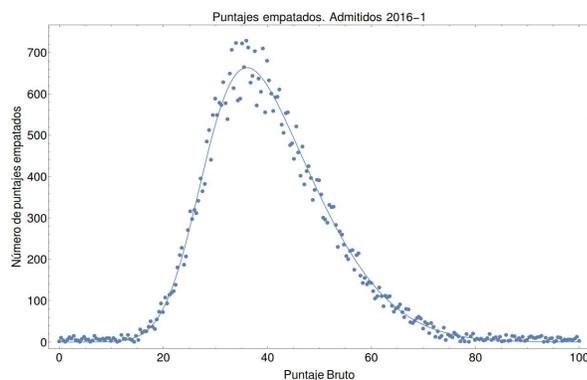


Figura 3: Los valores en el eje horizontal son los puntajes promedios, en pasos de 0.4. El eje vertical muestra el número de puntajes empatados. Nótese la asimetría: Tiene una pequeña “cabeza” en los puntajes entre 0 y 20 y una larga “cola” en los puntajes entre 60 y 100.

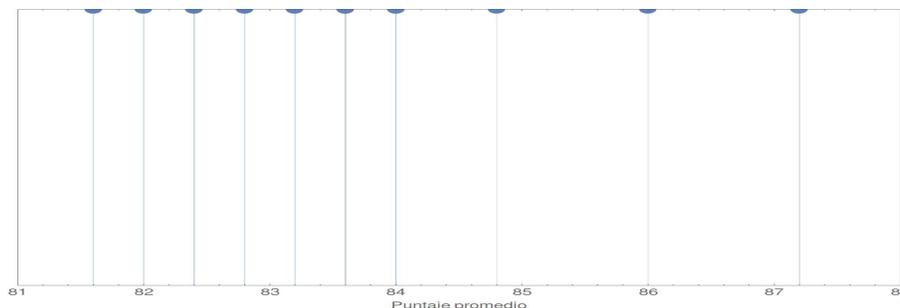


Figura 4: Los 23 promedios ordenados más altos. Sólo son visibles 10 a causa de los empates.

9 Discriminaciones odiosas y antitécnicas

El fin del sistema **no es la calidad sino la calificación**, la cual se confunde eufemísticamente con calidad. Si la admisión se relacionara con la calidad, debería concluirse que el sistema está discriminando diferentes áreas (posible violación del derecho a no ser censurado y de la autonomía universitaria). En efecto, en los programas en los cuales el número de aspirantes es muy grande, los puntajes de corte son muy altos; y en los programas en los cuales el número de aspirantes es pequeño, los puntajes de corte son bajos, esto último es especialmente cierto cuando el número de aspirantes es comparable al número de cupos. El sistema, supuestamente, entregaría a algunas carreras un número muy grande de alumnos casi “genios” (o “pilos” que llaman), el resto de carreras deberían conformarse con un número grande de alumnos “de baja calidad”. Una forma de **discriminar los saberes**, en caso de que los autores realmente creyeran que

los “más pilos” tienen “más méritos” o son de “mejor calidad”.

Un análisis similar se puede realizar para los diferentes *rankings*: Admisión a la Universidad Nacional de Colombia, MIDE, Pruebas Saber, Convocatorias de Colciencias, etc.

Las discriminaciones cometidas en la primera mitad del siglo XX motivaron la Declaración Universal de Derechos Humanos (promulgada el 10 de diciembre de 1948, poco después de la caída del régimen nazi). Colombia, como país signatario, incluyó en su constitución: “**Artículo 13.** *Todas las personas nacen libres e iguales ante la ley, recibirán la misma protección y trato de las autoridades y gozarán de los mismos derechos, libertades y oportunidades sin ninguna discriminación por razones de sexo, raza, origen nacional o familiar, lengua, religión, opinión política o filosófica.*” Es verdad que no menciona la posición en el *ranking* de Shanghai de la universidad en la cual obtuvo un título. Tampoco menciona la destreza para responder tests de escogencia múltiple y con ello obtener un puntaje en las llamadas “pruebas” del ICFES. Pero eso no legaliza la discriminación por esos conceptos y otros similares.

El problema de discriminación que se pretende resolver mediante las pruebas de admisión consiste en “leer” improntas grabadas en la mente de las personas con la finalidad de “predecir la calidad académica ... para el logro de sus objetivos formativos”, “evaluar los conocimientos y la capacidad de aplicarlos en el análisis y resolución de problemas”, “fomentar la probabilidad de éxito escolar, ... perfilar ... la vocación profesional”, entre otros. Independientemente del aspecto jurídico, nos hemos atrevido a dilucidar el problema técnico de “medición” de la “calidad académica” para decidir quien merece ser admitido y quien no a cursar los estudios superiores.

El artículo 110 de las normas académicas de la Universidad de Antioquia establece que 3.0 es la mínima nota aprobatoria. Cuando un examen de un curso lo perdía un gran número de estudiantes, la “curva” fue un método usado para burlar dicho artículo. Como corolario del presente análisis, se deduce que las Normas Académicas merecen reformarse en su artículo 110: Las calificaciones numéricas por si solas no permiten definir un umbral de aprobación-desaprobación, ni establecer *rankings* para otorgar estímulos académicos (Matrícula de Honor y otros).

10 Algunas conclusiones y recomendaciones

- Las principales conclusiones son: (1) Los puntajes individuales de cada uno de los 2 o 3 tests tienen un error intrínseco debido a la Suma de Peras y Manzanas. (2) El error de Suma de Peras y Manzanas (suma de valores matemáticamente incompatibles por ser de unidades diferentes) en los 2 o 3 tests se aumenta al considerar que a su vez, por ejemplo, el puntaje de “biología” resulta de 25 áreas diversas, 25 Sumas de Peras y Manzanas. (3) Dada la enorme incertidumbre de los resultados, no es posible llegar a ninguna conclusión, honestamente no se puede afirmar que el puntaje sea una medida de los logros académicos.

- Tanto el modelo de admisión del AA 236 como del AA 480 poseen naturaleza discreta. Con el primero la diferencia mínima entre dos puntajes promedios, escalados entre 0 y 100, es de 1.25, con el segundo es de 0.4. No hay variaciones grandes de dichas distancias cuando el resultado se describe con Puntajes Estandarizados. Por lo tanto, es imposible que existan dos puntajes cuya diferencia sea inferior a dichas distancias mínimas.
- Los números fabricados por el modelo son de naturaleza inherentemente probabilística, no guardan relación con el alegado propósito de clasificar las personas según su grado de “calidad”.
- El modelo es probabilístico sesgado. Tiene relación con la capacidad económica y disponibilidad personal para someterse a los intensos entrenamientos específicos requeridos para responder los tests de escogencia múltiple.
- El modelo se aplica en medio de publicidad engañosa: Identificar el puntaje con la calidad. Afirmar que el sistema permite distinguir quien tiene calidad suficiente para cursar determinada carrera y quien no. Afirmar que el sistema permite distinguir entre quien tiene vocación para cursar determinada carrera universitaria y quien no. Afirmar que el sistema no discrimina y es “transparente”, al supuestamente facilitar el ingreso a los aspirantes de mayor calidad.
- La publicidad dada a la supuesta innovación introducida mediante el AA 480 es tan engañosa como la que se le daba al sistema anterior (basado en los tests de “Competencia Lectora” y de “Razonamiento Lógico”). Se decía que era un patrimonio de la universidad, una garantía de transparencia y equidad. Algunos profesores comparten esa afirmación: “*En principio el examen de admisión actual, con sus dos pruebas ha posibilitado el ejercicio de la equidad...*”, se lee en un comunicado de la Asociación de Profesores del 31 de agosto de 2015¹¹. La larga “cola” de la distribución de aspirantes mostrada en las figuras 3 y 5, que se logra solo con costosas ayudas artificiales, permite afirmar que la tal “equidad” no existe.
- Al observar la evolución de los Puntajes de Corte se puede concluir que con y sin el AA 480 resultan los mismos patrones del método anterior. Por esta razón, ambas variantes del mismo modelo son inadecuadas para medir la “calidad”.
- Derogar los Acuerdos 480 y 236. Porque son “más de lo mismo”, simples variantes de un mismo sistema que se ha venido usando sin cambios esenciales desde hace 50 años.

¹¹Incluso muchos estudiantes: “... *la confiabilidad, transparencia, credibilidad, solidez, seriedad y reconocimiento a nivel nacional que caracterizaban a su anterior proceso de admisión ... se garantice un proceso de admisión serio y confiable tal como lo asegura el regulado por el Acuerdo 236 sin las modificaciones dictadas por el Acuerdo 480.*”, se lee en un comunicado de la Asamblea Estudiantil del 25 de noviembre de 2015.

- Los Puntajes Estandarizados merecen, entre otras, las siguiente crítica. Si lo que se busca con las preguntas tipo test es descubrir los conocimientos, habilidades, vocaciones y actitudes de los aspirantes, expresables en el número de “respuestas correctas” y “respuestas incorrectas”, ¿por qué no usar esos números en vez de los Puntajes Estandarizados? ¿La tal “curva” podría ser un artilugio dependiente del “modelo”?¹²
- No volver a realizar admisiones usando los tests de opinión por cuanto requieren entrenamientos *ad hoc*, nada tienen que ver con la calidad ni con los diversos procesos formativos durante 11 años y son abiertamente discriminatorios.
- Promover la creación de diplomados a realizarse en los años 11 y 12 de la escuela primaria y media, con enfoque similar al del SENA y el INEM, en temas diversos. La finalización habilitaría para cursar las carreras universitarias¹³.
- Si no se quiere implementar la sugerencia anterior, entonces, de manera transitoria, escoger los admitidos a las carreras de la Universidad de Antioquia mediante una rifa entre los aspirantes a cada carrera. Desde ya se deberían **crear** las condiciones para admitir a todos los aspirantes.
- Los sistemas actuales de medición analizados no son adecuados para representar la complejidad y diversidad del ser humano en general, en especial durante su proceso de aprendizaje.

11 Apéndice: Acerca de la pérdida de información

La suma $0P + 5M$ da $5M$ y $5P + 0M$ son $5P$. En tanto que $xP + (5-x)M$, con $x=0,1,2,3,4,5$ no se sabe qué es cuando $x \neq 0$ o $x \neq 5$. Las respectivas desviaciones estándar respecto al promedio “2.5” son: $[(0,5),3.53553]$, $[(1,4),2.12132]$, $[(2,3),0.707107]$, $[(3,2),0.707107]$, $[(4,1),2.12132]$, $[(5,0),3.53553]$.

Dado que no es lo mismo 3 Manzanas y 2 Peras que 2 Manzanas y 3 Peras, ni que 0 Manzanas y 5 Peras, ni que 5 Manzanas y 0 Peras, ¿Cómo es posible que en todos los casos el promedio de “2.5”? Es evidente que el precio de Sumar Peras con Manzanas para obtener 1 número en vez de los 2 números originales implica una pérdida de información. Si se supone que las 6 posibilidades son

¹²Al hacerle ingeniería inversa a las tablas en Excel para hallar la regla de conversión entre puntaje bruto y estandarizado, se halla que es una transformación lineal: $P_E = aP_B + b$, donde a y b son diferentes para los tests biología, matemáticas y humanidades (véase la figura 8). Dado que la diferencia mínima entre dos Puntajes Brutos (en la escala 0-100) es de 0.4, se sigue que la diferencia mínima entre dos Puntajes Estandarizados es de 0.421 o de 0.339. Pero esto **no significa** la aparición de puntajes estandarizados con cualquier valor en la tercera cifra decimal, dos puntajes brutos en empate dan lugar a dos puntajes estandarizados en empate.

¹³Véase una sustentación en la p. 7 de <http://goo.gl/G3Wr6N>. *El resultado de estudiar durante doce años*, Jorge Mahecha G., El Santuariano, año XCIV, No. 500, febrero de 2015.

equiprobables, la obtención de los 6 resultados es equivalente a lanzar un dado. Los dos resultados posibles al lanzar una moneda se representan con 1 bit de información. Los 4 posibles resultados de lanzar 2 monedas se representan con 2 bits. Los 8 posibles resultados de lanzar 3 monedas se representan con 3 bits. Entonces los 6 posibles resultados de lanzar un dado contienen una información intermedia entre 2 y 3 bits. En efecto, la entropía de Shannon vale en este caso $-6 \times (1/6) \times \log_2(1/6) = 2.58496$ bits. Por su parte, el promedio “2.5” conlleva $\log_2 2.5 = 1.32193$ bits. Por lo tanto, reemplazar los 6 posibles resultados originales por un único número, el promedio, conlleva una pérdida de información de $2.58496 - 1.32193 = 1.26303$ bits.

El promedio de xP, xM y xU daría “x”. La entropía, igual que la desviación estándar vale 0. Éste es el único caso en el cual el promedio contiene la misma información que los datos que lo originaron.

En el test de 25 preguntas, los posibles resultados se expresan mediante tiras de 25 ceros y unos (un número binario de 25 bits), hay 2^{25} posibles. Cada resultado tiene x “1” y 25-x “0”. El número de maneras diferentes de responder x correctas es $25!/[x!(25-x)!]$. Si se suponen equiprobables, su contenido de información vale el logaritmo en base 2 de este número.

El “puntaje”, por su parte, vale “x”, un número entre 0 y 25, para cuya representación se requieren entre 1 y $\log_2 25 = 4.64386$ bits.

Por lo tanto, al reemplazar una tira de 25 ceros y unos por la suma de sus dígitos¹⁴ se produce una enorme pérdida de información, representada en la figura 9. Alcanza su máximo para $x = 12$ y $x = 13$, cuando vale 22.3102 bits.

El número de maneras diferentes de obtener 12 o 13 respuestas correctas vale $5 \cdot 200.300$. ¿Esas 5 millones de maneras son equivalentes? Reemplazarlas por la suma, 12 o 13, implica la penalidad de aceptar que “el puntaje” está afectado por una gran incertidumbre. En efecto, 12 y 13 puntos conllevan 3.58 y 3.7 bits. La pérdida de información vale 18.725 y 18.609 bits, respectivamente. La pérdida de información es unas 5 veces la información aportada por el “puntaje”.

Con la parafernalia del test de admisión se busca “resolver” el siguiente Problema de Toma Decisiones: ¿Un aspirante dado merece o no ingresar? La solución se expresa con un número binario (“bit admisión”): 0 no ingresa, 1 si ingresa. Dado que la entropía es aditiva y que cada test tiene una entropía de hasta 18 bits, la entropía del promedio ponderado de los 3 tests puede alcanzar hasta $54 + 7 = 61$ bits (incluyendo los 7 debidos a la figura 2). La solución ideal de un Problema de Toma Decisiones no debe implicar una pérdida de información significativa. Si bien la compresión de información no se logra, en general, sin pérdidas, al menos debe dar lugar a una representación del objeto que permita una reconstrucción del mismo, así sea aproximada. En el presente caso, es impensable reconstruir la llamada “calidad” a partir del “bit admisión”, ni siquiera con los 4 o 5 bits del “puntaje ponderado”. Se concluye que el Problema de Toma de Decisiones que se quiere resolver con el “modelo” de admisión descrito, es un problema mal planteado.

¹⁴Es el *checksum* o *hash sum* usado en telecomunicaciones e informática.

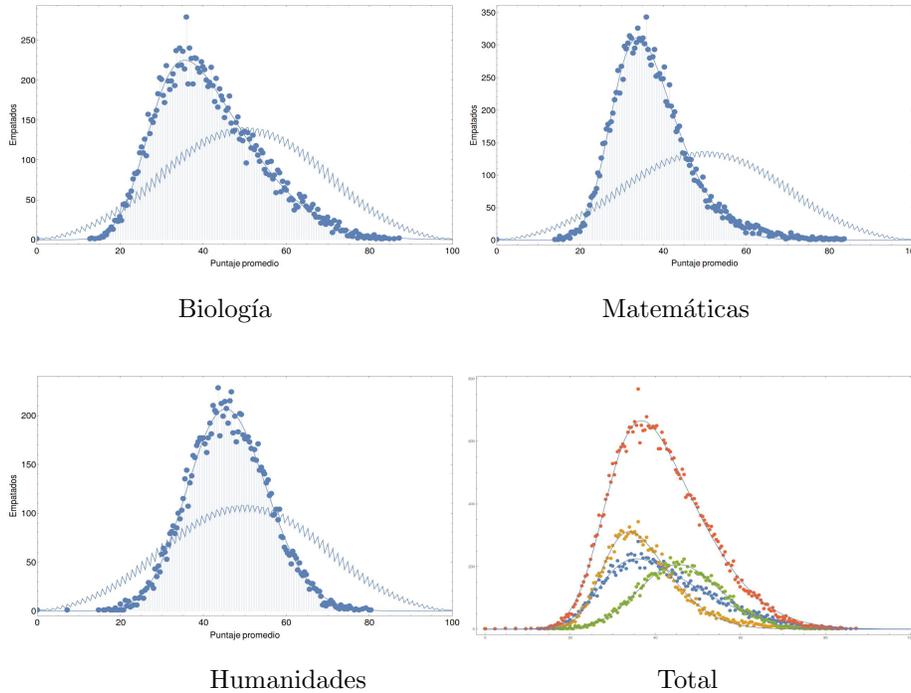


Figura 5: 44144 personas aspiraron a ingresar en la cohorte 2016-1. De ellas, 16093 realizaron el test específico sobre biología, 15608 el de matemáticas, 12377 el de humanidades. Las figuras muestran los números de personas que obtuvieron cada uno de los puntajes ponderados posibles. Quienes presentaron el test específico de biología obtuvieron 180 puntajes diferentes en el promedio ponderado de los tres tests, los de matemáticas 174, los de humanidades 159. Las gráficas muestran los correspondientes resultados. Se muestra la gráfica simétrica de los resultados que se obtendrían a partir de las probabilidades de los puntajes del “modelo”, representados en la figura 2. Nótese que la distribución más simétrica es la de “humanidades” y que “biología”, “matemáticas” y los puntajes de la totalidad de aspirantes tienen larga “cola” en los puntajes altos. También se pusieron los ajustes a distribuciones normales asimétricas con parámetros:

Test específico	α	μ	σ
Biología	2.86	28.39	17.42
Matemáticas	3.30	26.82	19.37
Humanidades	1.02	39.37	11.81
Total	2.48	27.70	12.36

α es el parámetro de asimetría, μ la media, σ la desviación estándar.

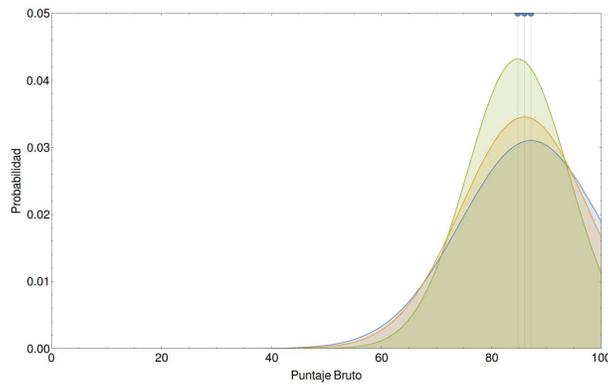


Figura 6: Los tres puntajes más altos con las correspondientes “campanas”. En el eje horizontal están los posibles puntajes y en el vertical la probabilidad de cada puntaje. Los puntajes y sus correspondientes desviaciones estándar son $[87.2, 12.8582]$, $[86.0, 11.547]$, $[84.8, 9.2376]$, obtenidos a partir de la tabla citada.

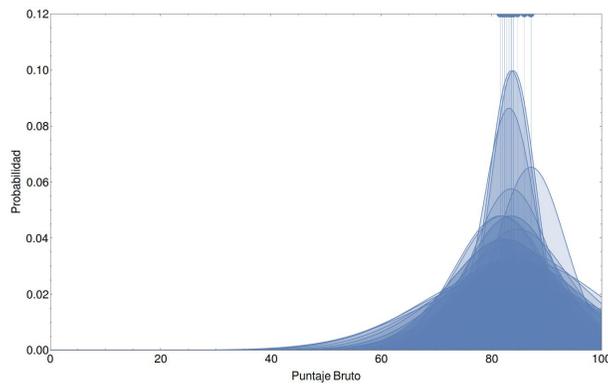


Figura 7: Las 23 calificaciones más altas con sus correspondientes “campanas”. En el eje horizontal están los posibles puntajes y en el vertical la probabilidad de cada puntaje. Nótese que hay 10 puntajes diferentes y 23 “campanas”. Esencialmente los 23 están en “**empate técnico**”. El tal ordenamiento no existe. **La validez del ordenamiento usando los promedios es similar a la del obtenido con una rifa.**

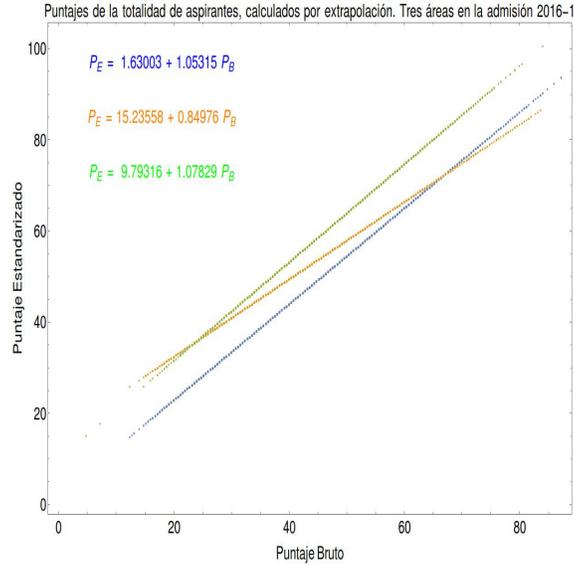


Figura 8: La tabla `Admitidos_2013_2016.xlsx` contiene los puntajes Brutos de todos los admitidos en los tres tests, en escala 0-100 y los correspondientes “Puntajes Estandarizados”. A partir de los mismos se obtuvo la regla de transformación entre las dos escalas. La tabla `Matriz-respuestas-correctas.xlsx` contiene los Puntajes Brutos de los aspirantes admitidos y no admitidos en los tres tests, en escala 0-25. Mediante extrapolación de dicha transformación a los no admitidos se obtuvo esta gráfica, en la cual cada recta corresponde a los aspirantes catalogados por determinado test específico.

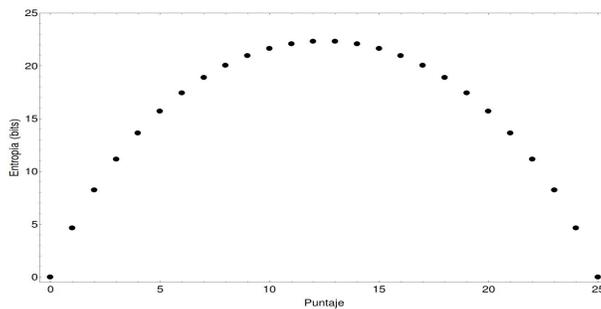


Figura 9: $\log_2\{25!/[x!(25-x)!\}]$ en función de x .