**Notas sobre el análisis descriptivo\***

\*Este documento fue redactado con la finalidad de esclarecer los pasos que se llevaron a cabo para la entrega de este primer análisis descriptivo. Si bien puede recuperarse con total libertad cualquier información citada en este documento, debe mantenerse en mente que se trata de un documento informalmente redactado y cuyo único propósito es dar sentido al análisis de datos y ayudar a su presentación a un tercero.

Índice

1. Descripción del contenido de la Carpeta en Drive
2. Tratamiento de las bases de datos
3. Exploración GENERAL de los puntajes obtenidos en la prueba
4. Exploración por Grado
5. Exploración por Grupo
6. Exploración por “Sexo” (¡Cuidado!)
7. ANEXO
   1. Indices de dificultad

**0.- Descripción de la Carpeta en Drive**

Se incluyen:

* 2 bases de datos a manera de insumo:
  + El archivo “Respuestas.csv”, que es el insumo principal Se trata de la misma base que me fue entregada con las respuestas registradas por los estudiantes para cada item, guardada en formato .csv (comma separated values, “fácil” de leer en R) y con un par de columnas adicionales (que se pueden ver en el extremo derecho de la hoja de calculo, si se abre con Excel) donde se clasificó a los estudiantes por Grado (según su grupo de procedencia) y sexo [[1]](#endnote-1)(según su nombre)
  + El archivo “Claves.csv” que contiene las respuestas correctas por ítem. Las columnas Item\_San y RCorr\_San presentan las respuestas correctas correspondientes a los ítems presentados en el Examen que me fue proporcionado en un PowerPoint, en el orden y con la numeración ahí empleados. Las columnas Item\_Ram y RCorr\_Ram señalan la numeración que corresponde a estos mismos ítems en la versión de PLANEA en línea que corresponde con la clave de respuestas presentada por Ramsés.
* 2 base de datos construida
  + El archivo “MatrizAciertos.csv” que se construye con uno de los códigos en R anexos, haciendo la traducción de las respuestas registradas en el archivo de Respuestas en valores 0 y 1, dependiendo su correspondencia con la Clave de Respuestas.
  + El archivo “Calificaciones.xls” que creé yo, sin código, a partir del archivo previamente mencionado. Este archivo contiene la misma información que el archivo csv previamente descrito, sólo que con un formato y presentación más “amigable”, teniendo como único objetivo presentar los puntajes computados para cada estudiante (para las formas A, B y total) de una forma que sea fácil de manipular a conveniencia.
* 2 presentaciones en Power Point
  + El “Repertorio de Gráficas” con todas las gráficas generadas en R, almacenadas en diapositivas independientes para que se puedan recuperar y utilizar a voluntad.
  + La “Presentación Tentativa” que pretende fungir como un boceto inicial de las gráficas que considero podrían ser más interesantes para presentar a la profesora, acompañada de algunas notas explicativas y comentarios.
* El presente reporte
  + Este reporte es más bien informal y pretende servir únicamente para que las personas involucradas en el proyecto (de momento sólo tengo a Sandy en mente :P), estén familiarizadas con el análisis descriptivo que se hizo y con cómo presentarlo a un tercero (a.k.a la profesora interesada)

**1.- Tratamiento de las bases de datos**

Parte 1: Notando discrepancias entre las clave de Ram y nuestro examen

El día miércoles 19 de diciembre me dispuse a trabajar con la clave de respuestas que Ramsés envió a Sandy. Por comodidad, dado que el formato en que fue enviado el archivo Excel tenía celdas combinadas (cuya lectura es complicada en R), intenté hacer manualmente mi propio archivo .csv a dos columnas donde sencillamente se registrara la relación ítem – respuesta correcta. Al hacer esto, lo primero que noté es que la base de Ramsés tenía 46 items, mientras que el examen contenido en los cuadernillos del Powerpoint tenía 50.

En el mejor de los casos, pensé, los últimos cuatro ítems contenidos en la base de Ramsés serían los últimos cuatro presentados por Sandy (esto implicaría que las numeraciones de los primeros 46 items coincidirían). Sin embargo, pronto me di cuenta que echándole un ojo a los ítems contenidos en el cuadernillo hecho por Sandy, las respuestas correctas no coincidían con lo que Ram tenía registrado (es decir, la respuesta correcta que Ramsés tenía marcada para el “Item 1”, por ejemplo, no tenía sentido cuando se miraba al Item 1 contenido en el cuadernillo del PP).

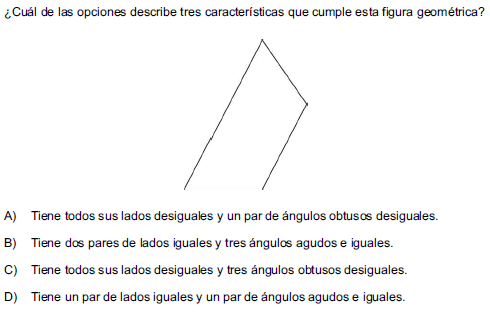
Así que la primera tarea fue resolver el examen yo misma para poder empezar con el análisis y tener algo que entregar a la profesora.

Parte 2: Resolviendo las diferencias encontradas

Sin embargo, la gran amenaza detrás del hecho de que la clave de Ramsés no coincidiera con nuestros ítems persistía: no podía descartarse la posibilidad de que de hecho se tratara de conjuntos totalmente distintos de ítems, lo cual haría completamente imposible utilizar la Matriz Q del Dr. Guaner et al para hacer la evaluación diagnóstica cognitiva.

En fin, me encontraba resolviendo el examen por mí misma cuando llegué al reactivo 22 del cuadernillo de respuestas:

22.



¡Ninguna de las opciones de respuesta es correcta! Los ángulos obtusos son mayores a 90° pero menores a 180° y, dado que la figura se presenta incompleta (no están cerrados todos sus lados), ninguna de las opciones de respuesta presentadas podía ser correcta.

Sabiendo que el examen PLANEA ha sido super validado, etc., me di a la tarea de buscar en línea el reactivo liberado original, para cerciorarme de que efectivamente la figura se presentara incompleta. Así fue como llegué a esta liga: <http://planea.sep.gob.mx/ba/prueba_en_linea/> donde encontré una versión aplicable en línea del examen PLANEA… ¡con 46 items!

Encontrar esta versión en línea del examen me permitió resolver dos problemas: primero, comprobar que la figura sí debía presentarse cerrada (por lo que debe invalidarse el ítem 22 del cuadernillo de preguntas en Power Point, ya que el ítem viene mal “de fábrica”) y segundo, tras verificar que las respuestas a los ítems contenidos en esta versión con 46 items correspondía con la clave de respuestas de Ramsés, nos permitió tener seguridad en que dado que las versiones son equivalentes, la Matriz Q para hacer la evaluación diagnóstica servirá para al menos 45 de los reactivos aplicados.

Parte 3: Ahora sí, “calificando” el examen

En fin, una vez habiendo definido las respuestas correctas para la versión del examen aplicada en Lancaster, se procedió a hacer el código en R que permitiera traducir las respuestas registradas por los estudiantes (Respuestas.csv) en una nueva matriz con valores 0 y 1 que permitiera identificar los aciertos y errores (MatrizAciertos.csv)

**2.- Exploración GENERAL de los puntajes obtenidos en la prueba**

Una vez habiendo computado las puntuaciones obtenidas por los estudiantes, se procedió a hacer el análisis descriptivo prometido. Empezamos por el análisis global (el que considera la respuesta de todos los 127 estudiantes) de los puntajes obtenidos en las Formas A y B, del puntaje total obtenido en el y los índices de dificultad.

|  |
| --- |
| **NOTA IMPORTANTE**  **¡Antes de continuar con la lectura!** |
| A partir de este punto, comienzan a presentarse las principales conclusiones que pueden abstraerse a partir de los análisis realizados. De manera general, las comparaciones que podrían evaluarse (por ejemplo, ¿es más fácil para los alumnos de Sexto responder la prueba que para los alumnos de quinto? ¿Quiénes obtienen mejor puntuación, las niñas o los niños?, etc) suelen presentare acompañadas de algún gráfico (anexos todos en RepertorioGraficas.pptx) que facilite su lectura, sin embargo, lo realmente correcto sería comprobar que las diferencias encontradas sean **estadísticamente significativas.**  Esta nota pretende explicar brevemente este concepto, de manera que haga sentido a la profesora.  La idea es muy simple: Supongamos que quiero evaluar si los niños en mi salón de clases son mejores que las niñas en un examen de matemáticas. Una primera aproximación para resolver esta pregunta sería observar los datos crudos: por ejemplo, 1) si graficamos un histograma en el que aparezca el número de veces que los niños y las niñas obtuvieron cada calificación posible en el examen y en general se observa que la distribución de los niños está más cargada hacia la derecha (sobre los valores-calificaciones más altas), que la de las niñas, o 2) si sacamos el promedio de las calificaciones obtenidas por todos los niños y todas las niñas del salón y las comparamos (imaginemos que los niños sacan de promedio 9 y las niñas 8.7)… en ambos casos estaríamos tentados a decir que a los niños les fue mejor… ¿pero, es realmente el caso?.  La idea de conducir una prueba estadística para comparar los datos obtenidos en dos muestras (para el presente análisis se usaron **“pruebas t”**) tiene que ver con la pregunta, ¿qué tan grande tiene que ser la diferencia entre dos muestras para que yo pueda asumir que los datos observados están de hecho influídos por un efecto real (intrínsecamente relacionado con el las características de la muestra) y no al mero azar?  El problema con las técnicas de “a ojo de buen cubero” descritas previamente es que la única forma en que yo podría llegar a la conclusión “no, no hay diferencias entre niños y niñas” sería si yo observara exactamente las mismas calificaciones en el conjunto de los niños y en el conjunto de las niñas. Este escenario resulta, por supuesto, inverosímil y altamente improbable. Una forma de pensar en la interpretación de las diferencias como algo “estadísticamente significativo” sería, ¿qué tan diferentes tienen que ser los datos obtenidos de mis dos muestras, para que yo pueda decir que son consistentemente diferentes en función a mi variable de interés? (Por ejemplo, esperaría encontrar que consistentemente los niños puntúen por encima de las niñas, y no sólo “en promedio”). |

Las gráficas resultantes del análisis global pueden consultarse en cualquiera de los Powerpoint. Para fines de este reporte se enfatiza únicamente en las posibles conclusiones, que para este primer análisis global serían:

* La Forma B es más difícil que la Forma A (este hallazgo se reafirma constantemente a lo largo del resto de los análisis, donde consistentemente se encuentran puntuaciones más altas en la Forma A que en la Forma B)
* La Forma A parece ser tan fácil, que no hace mucha diferencia si los estudiantes van en Quinto o Sexto (la prueba estadística señaló que se trata de diferencias **no significativas**).
  + Esto no ocurre en la Forma B, o en la Puntuación Total, donde de hecho se observan diferencias en función al grado que cursa el estudiante.
  + A continuación se presentan un par de tablas que a los estudiantes con las mejores puntuaciones en las Formas A, B y en el examen total (el tamaño de la tabla varía porque en todos los casos se incluyó a todos los estudiantes que tuvieran la misma puntuación que el 10° mejor estudiante)

|  |  |  |
| --- | --- | --- |
| Los estudiantes con los 3 puntajes más altos en la Forma A | | |
| Puntaje | Nombre | Grupo |
| 24 | Medina Peláez Diego | 6-A |
| 23 | Ancona Rodríguez Camila | 5-A |
| 23 | Guzmán Bravo Carlos Iram | 5-A |
| 23 | Nuñez Rodríguez Jeannine Andrea | 5-A |
| 23 | Rebollar Trejo María Fernada | 5-A |
| 23 | Jiménez Castellanos José Daniel | 6-B |
| 23 | San Juan Jiménez Iñaki | 6-B |
| 22 | Castillo Jiménez Victor | 5-A |
| 22 | Ramírez Sánchez Dulce Shari | 5-A |
| 22 | Montes de Oca Valencia Marianna X. | 5-B |
| 22 | Zavala Hidalgo Ángel Daniel | 5-B |
| 22 | López Hernández Sofía | 6-A |
| 22 | Delgado Soledad Mariana | 6-B |
| 22 | Miranda Yañez Roberto Gael | 6-B |

**Sobre un total de 25 puntos**

|  |  |  |
| --- | --- | --- |
| Los estudiantes con el mejor puntaje en la Forma B | | |
| Puntaje | Nombre | Grupo |
| 23 | Medina Peláez Diego | 6-A |
| 22 | Luna Hernández Sarai | 6-B |
| 21 | Galván Cruz Rodrigo | 6-B |
| 20 | Zavala Hidalgo Ángel Daniel | 5-B |
| 18 | Miranda Yañez Roberto Gael | 6-B |
| 18 | Soto Resendiz Angel Said | 6-A |
| 18 | Perez Zamora Alexis | 5-A |
| 17 | Guzmán Bravo Carlos Iram | 5-A |
| 17 | González Velázquez Diego | 5-A |
| 17 | Orta Hernández Saraí | 6-B |
| 17 | Miranda Yáñez José Daniel | 6-A |
| 17 | Ramos Ramos Leonardo Adriel | 6-A |
| 17 | Castro Rodríguez Karime | 6-A |

**Sobre un total de 25 puntos**

|  |  |  |
| --- | --- | --- |
| Los estudiantes con el mejor puntaje TOTAL | | |
| Puntaje | Nombre | Grupo |
| 47 | Medina Peláez Diego | 6-A |
| 43 | Luna Hernández Sarai | 6-B |
| 42 | Galván Cruz Rodrigo | 6-B |
| 42 | Zavala Hidalgo Ángel Daniel | 5-B |
| 40 | Miranda Yañez Roberto Gael | 6-B |
| 40 | Guzmán Bravo Carlos Iram | 5-A |
| 39 | San Juan Jiménez Iñaki | 6-A |
| 38 | Soto Resendiz Angel Said | 6-A |
| 38 | González Velázquez Diego | 5-A |
| 38 | Orta Hernández Saraí | 6-B |
| 38 | López Hernández Sofía | 6-A |

**Sobre un total de 50 puntos**

Las tablas anteriores fueron construidas filtrando y reordenando los puntajes registrados en el archivo Calificaciones.xls construido manualmente con el propósito explícito de presentar las puntuaciones computadas por cada participante. Este último archivo presenta exactamente la misma información que el archivo construido en R MatrizAciertos.csv de forma más amigable. Para filtrar y manipular los puntajes presentados en este último archivo, hay que contemplar los siguientes códigos:

|  |  |
| --- | --- |
| Variable | Código |
| Grado | 1 – Quinto ; 2 – Sexto |
| Grupo | 1 – 5 A; 2 – 6 B; 3 – 6 A; 4 – 6 B |
| Sexo | 1 – Hombre; 2 – Indefinido; 3 – Mujer |
| Nombre | Lugar que ocupa dicho estudiante en la base inicial Respuestas.csv |

Es importante señalar que en este, y todos los casos posteriores en que se hable de las Puntuaciones Totales (la suma de los puntajes obtenidos por cada estudiante en las Formas A y B), se toman en cuenta únicamente los estudiantes que asistieron y presentaron tanto la Forma A como la Forma B del examen. La decisión de excluir a los participantes que hubieran faltado a cualquiera de las dos Formas se hizo pensando que estos estudiantes, por default, no podrían tener una puntuación mayor a los 25 puntos correspondientes a la versión del examen que sí respondieron (penalizándolos en automático sin los 25 puntos restantes), por lo que agregarlos al análisis habría “metido ruido” al mismo..

En cuanto a los índices de dificultad computados, destacan las siguientes conclusiones:

* No hay ningún ítem que sea tan difícil como para que menos del 5% de los estudiantes lo acertara (como se puede apreciar en el huequito del primer rango)
* No hay ningún ítem tan fácil como para que más del 90% de los estudiantes lo acierten.
* La mayoría de los ítems se concentran entre el 30 y el 70%, lo cual quiere decir que la mayoría de los ítems fueron acertados por al menos el 30% de los estudiantes, pero no más del 70%.

Al final de este documento se anexa una tabla con la relación entre los ítems que fueron respondidos por menos del 40% del total de los participantes y los ítems que fueron respondidos por menos del 30% de los estudiantes de cada subgrupo (Quinto, sexto, 6-A, 6-B, 5-A, 5-B). Los ítems puntuales son presentados al final del power point “PresentaciónTentativa”

**3.- Exploración por Grado**

En la Matriz de Aciertos construida, se utilizaron los siguientes códigos para referir a cada grado

|  |  |
| --- | --- |
| Valor | Significado |
| 1 | Quinto |
| 2 | Sexto |

En general, se observa que:

* Para la Forma A **no hay diferencias significativas**  entre los puntajes obtenidos por los estudiantes de quinto y sexto año.
* Para la Forma B **sí hay una diferencia significativa** que sugiere que los estudiantes de sexto son significativamente mejores que los estudiantes de quinto año.
* En cuanto a las puntuaciones totales, también se encuentran **diferencias significativas** entre los puntajes obtenidos por los estudiantes de quinto y sexto año. Este hallazgo suena lógico considerando que las diferencias encontradas para la Forma B ya habían mostrado ser significativas, así que puede pensarse que estas se arrastraron hacia los puntajes totales.

En cuanto a los índices de dificultad computados de manera separada para los estudiantes de quinto y sexto año, tomando como referencia los extremos inferior (<30%) y superior (>70%), se observa que:

* Para los estudiantes de Quinto Año se observan muchos más ítems que fueron acertados por menos del 30% de sus estudiantes (9 ítems) que en el caso de Sexto año (6 ítems).
* Para los estudiantes de Sexto año se observan más ítems que fueron acertados por más del 70% de los estudiantes (16 ítems) que en Quinto año (9 ítems).
  + Es decir que en general, se observa que para los estudiantes de Sexto año hay más ítems que resultaron ser fáciles (aciertos >70%) y menos ítems que resultaron ser difíciles (aciertos<30%). .

**5.- Exploración por “Sexo” (¡Cuidado!)**

**NOTA IMPORTANTE:** El análisis de las diferencias por sexo se añadió como un posible “hallazgo interesante”, sin embargo, es importantísimo mencionar que dado que no se solicitó explícitamente a los estudiantes que señalaran su sexo antes de responder el examen, sería erróneo incluir este análisis en cualquier reporte formal o publicación que se derivara de su análisis. Es de mi particular creencia que no hay problema si se presenta a la profesora en cuestión, como un “plus” al análisis presentado, pero hay que tener cuidado en mantener en mente que la clasificación de los estudiantes en las categorías “Hombre” y “Mujer” fue hecha por una servidora, así que no hay garantía de que sea correcta (pero puedes consultarla tú misma en una de las últimas columnas del archivo Respuestas.csv)

Cinco estudiantes fueron omitidos en este análisis porque su puro nombre no se consideró suficiente para descifrar, con total certidumbre, el sexo del participante. Dichos estudiantes son:

* Carrasco Villanueva Keren
* Hernánrez Alexa Ariel
* Santos Zapato Yolotl Tlacaelel
* Soto Gómez Yehua
* Rodríguez Pérez Joshua

Del análisis realizado con los participantes que sí fueron clasificados, destacan las siguientes conclusiones:

* **No hay diferencias** significativas entre los puntajes obtenidos por los hombres y las mujeres en la Forma A del examen.
* Hay una **diferencia significativa** en los puntajes observados en la Forma B que sugiere que a los hombres les fue mejor en el examen.
* De igual forma, los hombres obtienen puntuaciones **significativamente mayores** que las mujeres en el total del examen.

Finalmente, dado que las clases de matemáticas no son diferidas por sexo, no se consideró pertinente agregar información respecto a los índices de dificultad (aunque están listas y disponibles si me dices que las quisieras añadir al reporte).

6.- Anexos

6.a Tabla de resumen de Indices de Dificultad

Global < .4 19 22 25 34 36 37 38 39 41 42 44 45 46 47

Global < .3 22 36 37 39 41 42

Quinto < .3 22 34 36 37 39 41 42 44 46

Sexto < .3 22 36 37 39 40 41

"5A" < .3 22 36 37 39 41 42 46 47 48

"5B" < .3 19 22 25 27 31 34 36 37 39 41 42 44

"6A" < .3 22 36 37 39 40 41 42 44 46

"6B" < .3 22 25 36 37 39 41

En la tabla anterior se presentan los ítems que fueron consistentemente identificados como “difíciles” para todos los estudiantes, y por cada subgrupo de los mismos.

En la presentación tentativa se incluyen todos los ítems que se presentan en color verde, que son aquellos que presentaron un porcentaje de aciertos global menor al 40%. Por cada uno de estos ítems, se incluye en la presentación el porcentaje de estudiantes de cada uno de los subgrupos enlistados que acertó el ítem.

En la tabla anterior se marcan en amarillo los ítems que resultaron particularmente difíciles para un grupo específico. Es decir, se marcan aquellos ítems que **no** presentaron un porcentaje de aciertos global menor al 40%, pero que para alguno de los grupos enlistados debajo, fueron acertados por menos del 30% de los estudiantes.

En general, se puede apreciar que la mayoría de los ítems que fueron identificados como “difíciles” para la gran parte de los estudiantes son aquellos que tienen que ver con problemas geométricos y operaciones con fracciones.

1. La clasificación por sexos se hizo con la finalidad de hacer el análisis más “interesante” para la profesora. Sin embargo, es importante enfatizar que es poco recomendable presentarlo, en tanto que se trata de datos construidos a posteriori, y que no fueron recolectados durante la aplicación del examen (es decir, en sentido estricto, “está mal” que yo, o alguien que no sea cada participante, agregue información que no les fue explícitamente solicitada) [↑](#endnote-ref-1)