

Modelización de un test incorporando los tiempos de respuesta al ítem

Laura Aspirot, Mario Luzardo, Leonardo Moreno*, Darío Padula

Núcleo de Investigación, Medición y Evaluación Educativa (NIMEE)

* Iesta, Departamento de Métodos Cuantitativos,
FCEA, Udelar

4 de mayo del 2023

Plan de la Presentación

- 1 Introducción
- 2 Asociación entre precisión y velocidad
- 3 Un ejemplito en una prueba múltiple opción
- 4 Algunos enfoques metodológicos.
 - La curva característica de un ítem
 - Modelos no paramétricos
 - Modelos paramétricos
- 5 Trabajo a futuro
- 6 Referencias

Preguntas

- ¿Qué vínculo existe entre la habilidad del estudiante y el tiempo ?

Preguntas

- ¿Qué vínculo existe entre la habilidad del estudiante y el tiempo ?
- ¿Items mas difíciles consumen más tiempo?

Preguntas

- ¿Qué vínculo existe entre la habilidad del estudiante y el tiempo ?
- ¿Items mas difíciles consumen más tiempo?
- ¿De que nos sirve incorporar en la modelización el tiempo?

Introducción

- En general en las pruebas de evaluación educativa estamos interesados en medir el conocimiento del alumno de manera independiente al tiempo requerido para su ejecución.
- En esta dirección existe un supuesto implícito, la prueba esta diseñada de forma que el tiempo no es una variable que afecte el rendimiento.
- Por ello no son frecuentes pruebas donde el tiempo de respuesta en cada pregunta este acotado, e incluso se piensa una duración total de la prueba que no influya en el desempeño del estudiante.
- Algunos autores las clasifican en pruebas desaceleradas (si se da mas tiempo al alumno su puntaje no cambiaría) y aceleradas (si al dar mas tiempo cambia el orden de habilidades en el grupo).

Introducción

- Sin embargo, muchos autores piensan (aún en el caso anterior) que estos supuestos son una idealización y que no son veraces, ver [Van der Linden and Hambleton, 1997]
- Es claro la relación entre el proceso cognitivo y la velocidad con que este se realiza. Aunque muchas veces la relación de asociación depende del tipo del test (positiva, negativa, lineal, no lineal).
- Si bien la medición del tiempo de respuesta como una forma de medir la actividad mental es de larga data en psicología, [Schnipke and Scrams, 2002], no así en pruebas educativas.

Introducción

- La dificultad de la recolección de esta información hace además que las investigaciones tuvieran un avance sustantivo en referencia a la precisión.
- Las pruebas informatizadas tienen la ventaja de proporcionarnos el tiempo que cada alumno destina a la prueba y en particular el que dedica a cada ítem.
- El objetivo es por tanto modelar la actividad una combinación entre velocidad y precisión que proporcionan información sobre un mismo constructo.

Introducción

Algunas utilidades de modelar los tiempos de respuesta de un item:

- Detección de tiempos no plausibles y proporcionar alguna medida temporal de falseamiento de la prueba.
- Tener una medida de la dificultad temporal del item que nos permita por ejemplo al confeccionar la prueba a partir de un banco de items poder estimar el tiempo de la misma.
- Conocer las características temporales de los alumnos, por ejemplo para la distribución de tareas o la conformación de grupos.

Introducción

- La motivación de estos estudios no sólo radica en saber cuanto va a demorar cada ítem, la prueba, o modelar el vínculo entre velocidad y precisión, ver [Lee and Chen, 2011], sino en brindar soluciones a los problemas de adivinación (por "apuro", por "falta de motivación" o fraude).
- En este sentido la adivinación puede causar sesgos severos en los parámetros por TRI, y la detección de estos individuos es de suma relevancia, ver [Hadadi and Luecht, 1998].
- Otros tipos de procesos se vinculan con el tiempo: "Efecto de calentamiento" y "Efecto de velocidad" (a veces por cansancio), ver [Shao, 2017].

Como se asocian?

- Una primera tradición en pruebas educacionales es suponer que existe una asociación predominantemente negativa (Cuanto más sabe el examinado contesta afirmativamente en menos tiempo y cuanto menos se esfuerza mas para contestar con mayor error)
- Una segunda corriente es suponer un trade-off entre velocidad y precisión (correlación positiva entre la proporción de respuestas correctas y el tiempo promedio). En general provenientes de pruebas donde para tareas del mismo orden de dificultad se cambia el tiempo asignado para su ejecución. Los sujetos que trabajan más rápido lo hacen a expensas de su precisión y, en consecuencia, producen menos respuestas correctas.
- Por último si se examinan los tiempos de respuesta cada ítem de dos estudiantes se aprecia un correlación positiva entre los tiempos de repuesta. Puesto que simplemente algunos ítems requieren menos tiempo que otros.

Habilidad vs tiempo

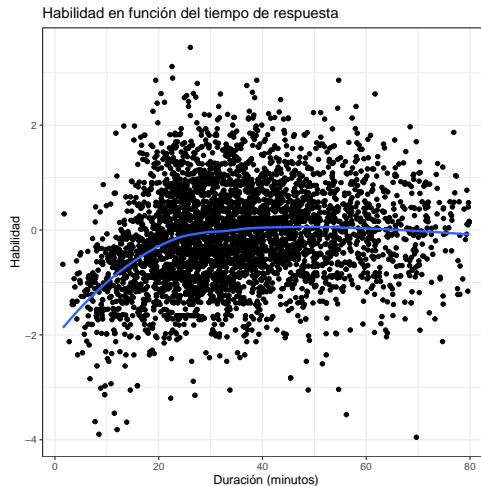


Figure: Dependencia entre la habilidad del estudiante y el tiempo total destinado a la prueba

Dificultad vs tiempo

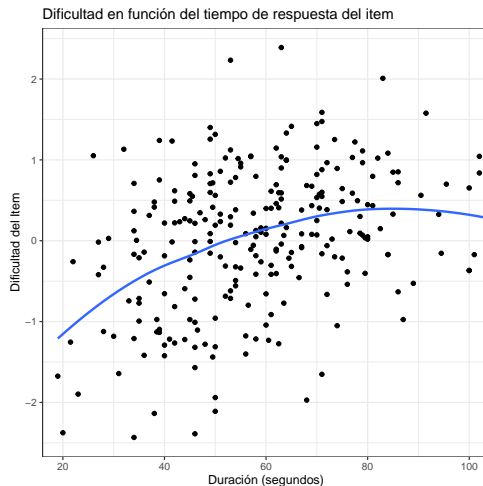


Figure: Dependencia entre la dificultad del ítem y el tiempo promedio de respuesta

Histogramas de los tiempos de respuesta por ítem

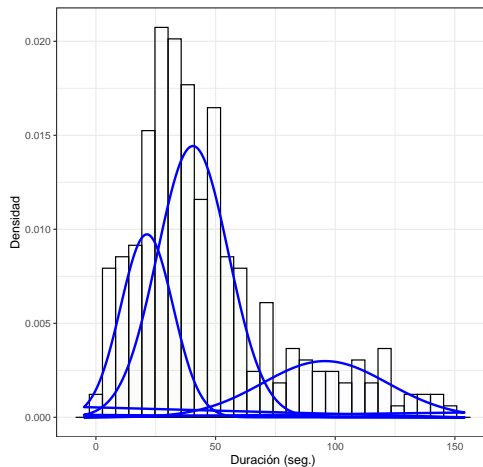


Figure: Dificultad alta y tiempo promedio bajo

Histogramas de los tiempos de respuesta por ítem

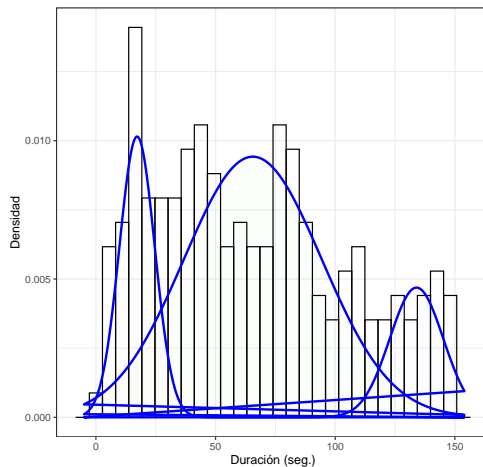


Figure: Dificultad alta y tiempo promedio alto

Histogramas de los tiempos de respuesta por ítem

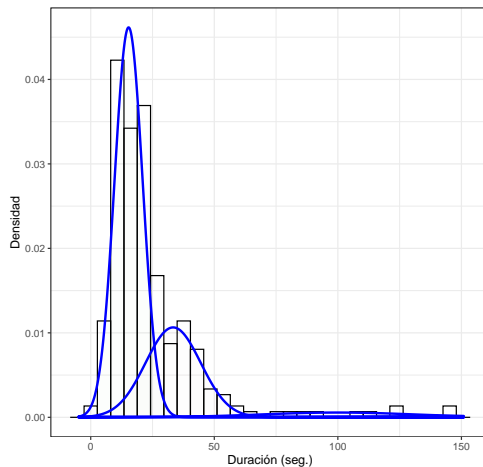


Figure: Dificultad baja y tiempo promedio bajo

Histogramas de los tiempos de respuesta por ítem

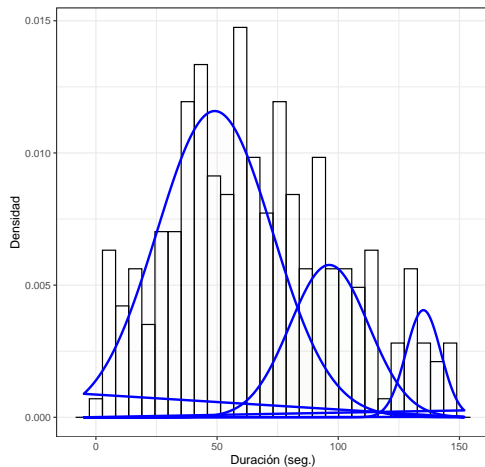


Figure: Dificultad baja y tiempo promedio alto

Tiempos de respuesta no plausibles

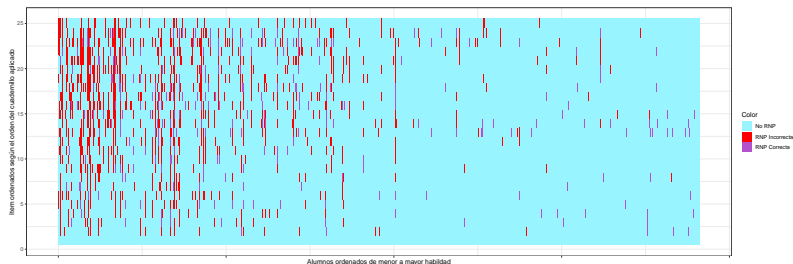


Figure: Diagrama donde se observa que al aumentar la habilidad el estudiante contesta menos items en tiempos de respuestas no plausibles. Vemos además que este comportamiento se acrecienta en los items finales de la prueba

Otra posible visualización descriptiva: Máxima racha de respuestas no plausibles según su habilidad.

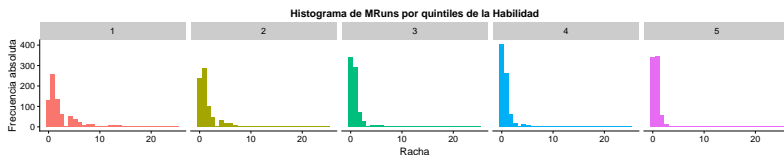


Figure: Histogramas de máximas rachas de respuestas no plausibles según su habilidad.

Algunos enfoques metodológicos. La curva característica de un ítem (CCI)

$$P_j(\theta) = P(U_j = 1|\theta)$$

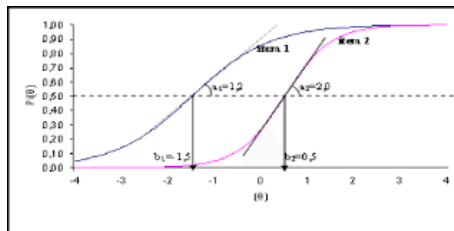


Figure: CCI de dos items.

Algunos enfoques metodológicos para la modelización (incorporando el tiempo): Modelos no paramétricos

- Dado un item, τ Característica temporal.
- $\hat{\theta}_i$ es la estimación de la habilidad de individuo i .
- $\hat{\tau}_i$ es la estimación de la característica temporal de individuo i .
- K es una función de núcleos.

$$m_{(\theta, \tau)}(t) = P(T \leq t | \theta, \tau, Y = 1) = E(1(T \leq t) | \theta, \tau, Y = 1).$$

El estimador es dado por,

$$\hat{m}_{(\theta, \tau)}(t) = \frac{\sum_{i=1}^N 1(T_i \leq t) Y_i K\left(\frac{\hat{\theta}_i - \theta}{h_1}\right) K\left(\frac{\hat{\tau}_i - \tau}{h_2}\right)}{\sum_{i=1}^N Y_i K\left(\frac{\hat{\theta}_i - \theta}{h_1}\right) K\left(\frac{\hat{\tau}_i - \tau}{h_2}\right)}$$

Modelos no paramétrico

$$P(T \leq t, Y = 1 | \theta, \tau) = m_{(\theta, \tau)}(t) P(\theta, \tau)$$

con $P(\theta, \tau) = P(Y = 1 | \theta, \tau)$.

Este término puede ser estimado como

$$\widehat{P(\theta, \tau)} = \frac{\sum_{i=1}^N Y_i K\left(\frac{\hat{\theta}_i - \theta}{h_1}\right) K\left(\frac{\hat{\tau}_i - \tau}{h_2}\right)}{\sum_{i=1}^N K\left(\frac{\hat{\theta}_i - \theta}{h_1}\right) K\left(\frac{\hat{\tau}_i - \tau}{h_2}\right)}$$

Modelos no paramétrico

Si suponemos que (θ, τ) tiene distribuciones marginales uniformes podemos determinar la distribución de la función de cópula C que puede ser estimada a través de las observaciones $(\hat{\theta}_1, \hat{\tau}_1), \dots, (\hat{\theta}_N, \hat{\tau}_N)$.

Modelos no paramétrico

Recordemos que

$$E(T|\theta, \tau, Y = 1) = \int_0^{+\infty} (1 - m_{(\theta, \tau)}(t)) dt,$$

Lo que permite estimar dicho valor de forma sencilla,

$$E(T|\widehat{\theta}, \widehat{\tau}, Y = 1) = \int_0^{+\infty} (1 - \hat{m}_{(\theta, \tau)}(t)) dt.$$

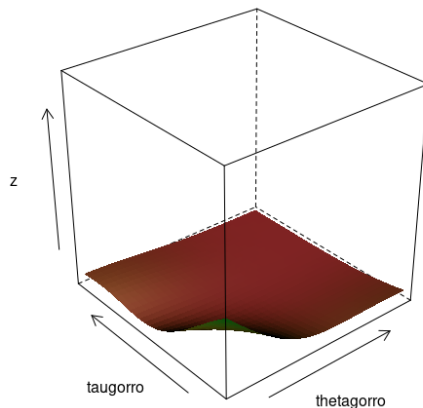
Modelos no paramétrico

Estamos ahora las probabilidades del tiempo esperado de T , independientemente de las habilidades de los sujetos que contestan la prueba.

$$E(T|Y = 1) =$$
$$E[E(T|\theta, \tau, Y = 1)] = \int_0^1 \int_0^1 E(T|\theta, \tau, Y = 1) \frac{C(\theta, \tau)}{\partial\theta\partial\tau} d\theta d\tau$$

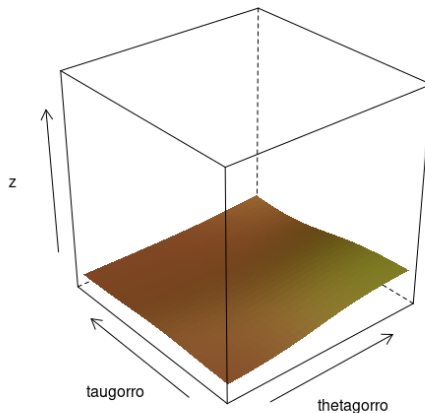
Estimación de $P(T \leq t | \theta, \tau, Y = 1)$

- t chico en un item de dificultad media



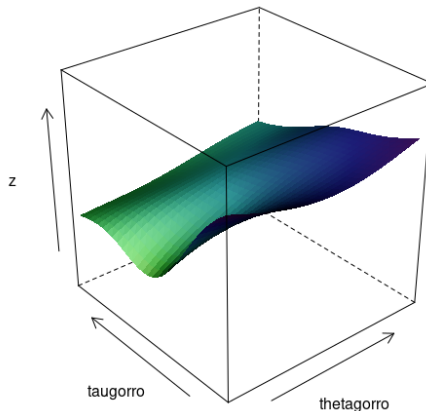
Estimación de $P(T \leq t, Y = 1 | \theta, \tau)$

- t chico en un item de dificultad media



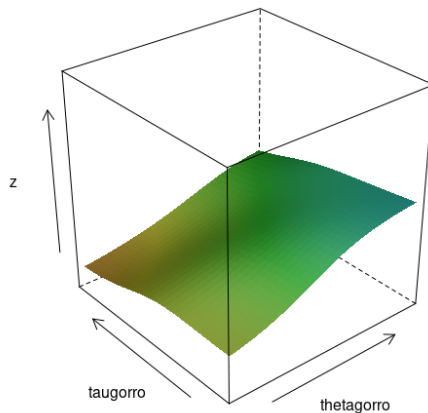
Estimación de $P(T \leq t | \theta, \tau, Y = 1)$

- t medio en un ítem de dificultad media



Estimación de $P(T \leq t, Y = 1 | \theta, \tau)$

- t medio en un ítem de dificultad media



Algunos enfoques metodológicos para la modelización:

Modelos paramétricos

- Probabilidad que el individuo p conteste correctamente el ítem i -ésimo dada la habilidad θ_p

$$P(U_{pi} = 1 | \theta_p) = \frac{\exp\{\theta_p + \log(t_{pi}) - b_i\}}{1 + \exp\{\theta_p + \log(t_{pi}) - b_i\}}$$

- Introducido por [Roskam, 1987].
- La idea es un trade-off entre velocidad y tiempo, se sustituye θ por $(\theta + \log(t))$ en un modelo de Rash.

Algunos enfoques metodológicos para la modelización: Las variable del modelo TRI se incorporan a uan modelización del tiempo

- Probabilidad que el individuo p conteste correctamente el item i -ésimo dada la habilidad θ_p

$$LT_{pi} = \log(T_{pi}) \sim N(\mu + \tau_p + \beta_i - \rho(a_i\theta_p - b_i), \sigma^2)$$

- Introducido por [Thissen, 1983] donde la probabilidad de respuesta es modelada con un 2PL por ejemplo.
- $\mu + \tau_p + \beta_i$ es el término que regula la “lentitud” esperada de examinado p con el item i , $(a_i\theta_p - b_i)$ es la variable asociada al modelo TRI, y ρ el parámetros de regresión de compromiso entre velocidad y precisión.

Un ejemplo con el paquete LNIRT de R

- Modelos jerárquico bayesiano.
- **(Primer nivel)**

$$P(U_{pi} = 1 | \theta_p, a_i, b_i) = \Phi[a_i(\theta_p - b_i)]$$

-

$$\log(T_{pi}) = \phi_i(\lambda_i - \psi_p) + \epsilon_{pi},$$

con $\epsilon_{pi} \sim N(0, \sigma_{\epsilon_i}^2)$.

- Asumiendo independencia entre las intensidades temporales (λ_i) y entre los residuos de dos examinados,

$$\text{cov}(\log(T_{pi}), \log(T_{pj})) = \phi_i V(\psi_p) \phi_j$$

Un ejemplo con el paquete LNIRT de R

- **(Segundo nivel)**
- Se considera una modelización bayesiana
- Distribuciones a priori,
 - $(\theta_p, \psi_p) \sim N_2(\mu_p, \Sigma_p)$
 - $(a_i, b_i, \phi_i, \lambda_i) \sim N_4(\mu_l, \Sigma_l)$
 - Se suponen distribuciones a priori para los hiperparametros Σ_p se le asocia una distribución de Wishart inversa, ver [Fox, 2010].

Un ejemplo con el paquete LNIRT de R

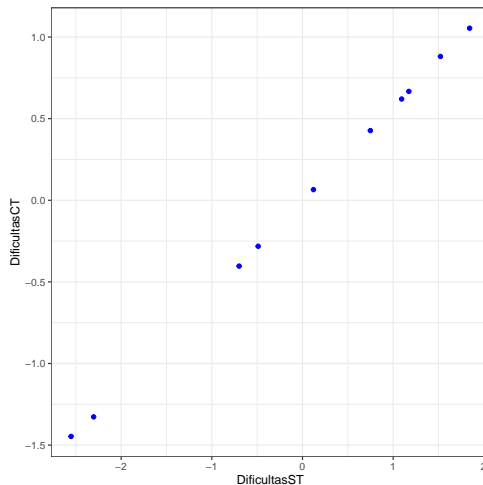


Figure: Comparación de dificultades considerando o no el tiempo de respuesta

Un ejemplo con el paquete LNIRT de R

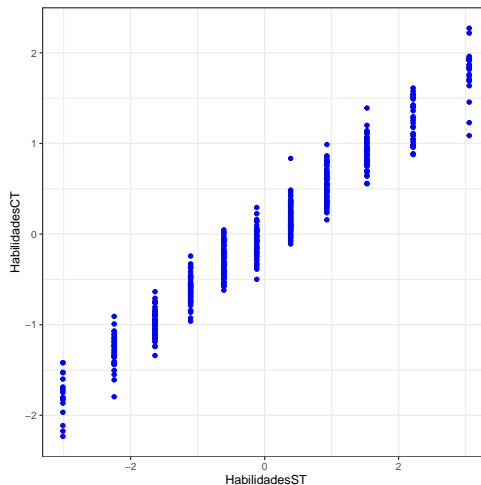


Figure: Comparación de habilidades considerando o no el tiempo de respuesta

Un ejemplo con el paquete LNIRT de R

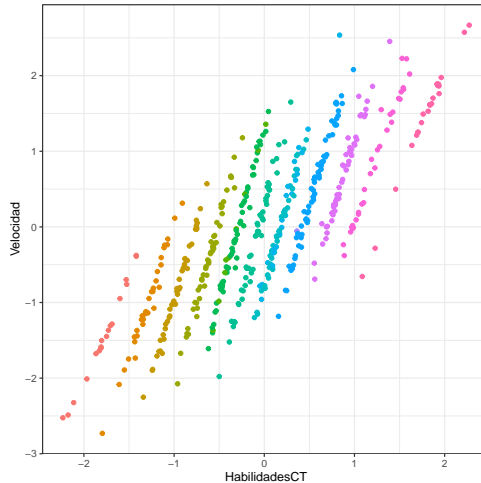
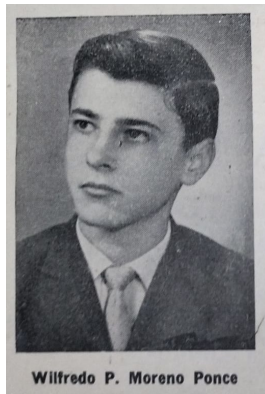


Figure: Relación entre velocidad y habilidad. A igual color igual habilidad sin considerar el tiempo

Que queremos hacer

- Comparar los ajustes paramétricos y no paramétricos.
 - Probar propiedades teóricas de los modelos no paramétricos que desarrollamos (consistencia y TLC)
 - Incorporar otras variables explicativas que den cuenta de la velocidad de respuesta del estudiante. Por ejemplo el nivel socio-económico o del área socio emocional.
 - Formulación de modelos robustos.
- * Tesis de Maestría de Juan...

GRACIAS !!



Referencias I



Fox, J.-P. (2010).

Bayesian item response modeling: Theory and applications.
Springer.



Hadadi, A. and Luecht, R. M. (1998).

Some methods for detecting and understanding test speededness on timed multiple-choice tests.

ACADEMIC MEDICINE-PHILADELPHIA-, 73:S47–S50.



Lee, Y.-H. and Chen, H. (2011).

A review of recent response-time analyses in educational testing.

Psychological Test and Assessment Modeling, 53(3):359.



Roskam, E. E. (1987).

Toward a psychometric theory of intelligence.

Referencias II

 Schnipke, D. L. and Scrams, D. J. (2002).

Exploring issues of examinee behavior: Insights gained from response-time analyses.

Computer-based testing: Building the foundation for future assessments, 34:237–266.

 Shao, C. (2017).

Aberrant response detection using change-point analysis.
University of Notre Dame.

 Thissen, D. (1983).

Timed testing: An approach using item response theory.
In *New horizons in testing*, pages 179–203. Elsevier.

 Van der Linden, W. J. and Hambleton, R. (1997).

Handbook of item response theory.
Taylor & Francis Group. Citado na pág. 1(7):8.