

Estadística para el análisis de pruebas de certificación

I Jornada Grupo de Interés en
Evaluación de Lenguas en España
(GIELE)
21 de octubre 2017

Julia Zabala
Centro de Lenguas

Universitat Politècnica de València (Spain)



Índice

Introducción

Principios de la evaluación

Análisis estadístico de pruebas

Teoría clásica del test y teoría de respuesta al ítem

Análisis a nivel de la prueba

Cálculo e interpretación de media, mediana, moda, rango, desviación estándar - Representación gráfica: histogramas

Cálculo de Alpha de Cronbach

Análisis a nivel de ítems

Cálculo e interpretación de índices de facilidad, índices de discriminación y análisis de descriptores

Revisión de ítems problemáticos

Conclusiones



Principios de la evaluación



Practicidad

Fiabilidad

Validez

Autenticidad

Efecto colateral



Entonces ...

Como profesionales de la enseñanza de lenguas ¿Para qué nos sirve la estadística?

Practicidad – si un ítem funciona bien, obtendremos más información de la capacidad del candidato

Fiabilidad – si un ítem funciona bien de manera estable, medirá lo que queremos

Validez – una prueba fiable no siempre es válida, pero una prueba válida siempre tiene que ser fiable

Autenticidad – para calcular el difícil equilibrio entre la fiabilidad y la autenticidad

Efecto colateral – si nuestro examen mide lo que queremos, impactará positivamente en el aula y fuera de ella

... sin olvidar



PERO
...¡somos
lingüistas!



TEORIA CLÁSICA DEL TEST

VENTAJAS

Descriptiva
Accesible
Requiere un menor número de candidatos
>40 (idealmente >100)

DESVENTAJAS

Dependiente de la muestra
No permite generalizar para poblaciones
diversas

SPSS, TiaPlus, TAP, Excel, CITAS

TEORIA DE RESPUESTA AL ÍTEM

VENTAJAS

Independiente de la muestra
Nos permite generalizar los resultados
Imprescindible para anclar ítems

DESVENTAJAS

Menos accesible
Requiere un mayor número de candidatos
(>200)

Winsteps, Facets, R



TEORIA CLÁSICA DEL TEST

Excel

Análisis a nivel de prueba

- Media
- Mediana
- Moda
- Desviación estándar

Alpha de
Cronbach

Histogramas

Análisis estadístico de pruebas

Análisis a nivel de ítems
o tareas

- Índice de facilidad del ítem
- Índice de discriminación del ítem
- Funcionamiento de distractores

Análisis a nivel de prueba

Media: la media es la respuesta “media” de los candidatos a la prueba calculada dividiendo el número de respuestas correctas por el número de estudiantes

Desviación estándar: la desviación estándar es una medida de dispersión de las notas de los candidatos en una prueba y nos indica como de diversa es nuestra población.

Mediana: el centro de las puntuaciones ordenadas de menor a mayor

Moda: la puntuación que más veces se repite (en una distribución puede haber varias modas)

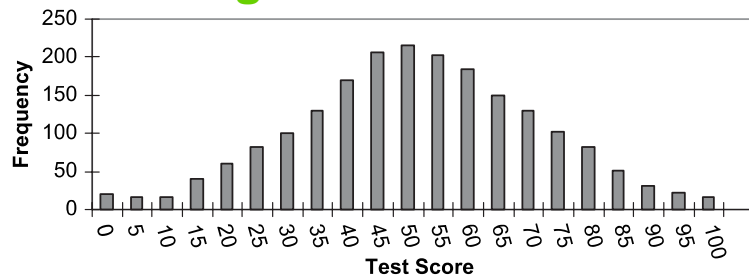
¿Cómo lo representamos gráficamente? **Histogramas**



Interpretación de un histograma

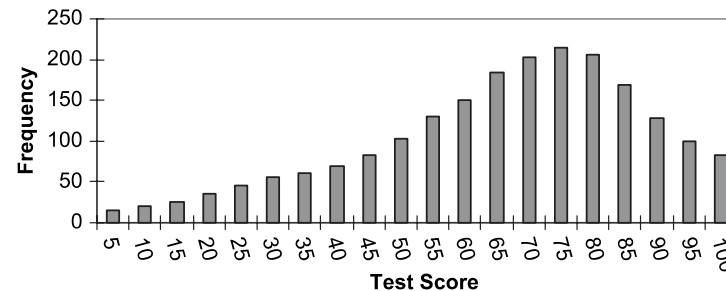
Análisis a nivel de prueba

Histograma 1



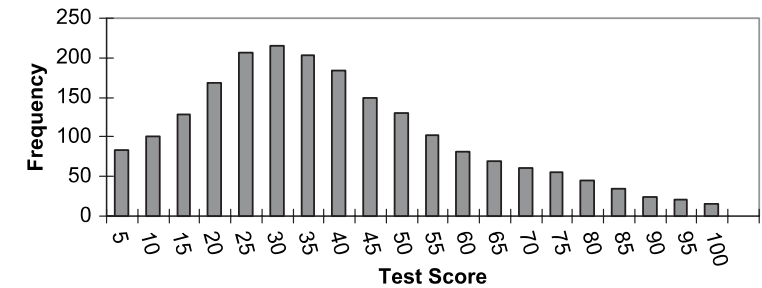
Distribución normal:
deseable para pruebas de certificación

Histograma 2



Sesgada a la izquierda: test fácil para los candidatos, deseable para una prueba de aprovechamiento

Histograma 3



Sesgada a la derecha: test difícil para los candidatos, característico de una prueba con referencia a un criterio (a principio de curso)

Ejemplos de Carr (2008)



Análisis a nivel de prueba

Fiabilidad (Alpha de Cronbach): mide la consistencia interna de la prueba en un rango de 0 a 1,00.

A partir de 0,7 se consideraría aceptable para una prueba, aunque para pruebas de certificación se prefiere >0,8.

Alpha de Cronbach	Consistencia interna
$\alpha \geq 0.9$	Excelente
$0.9 > \alpha \geq 0.8$	Buena
$0.8 > \alpha \geq 0.7$	Aceptable
$0.7 > \alpha \geq 0.6$	Pobre



Análisis a nivel de ítem

Índice de facilidad del ítem:

porcentaje de alumnos que contestó correctamente a la respuesta.

**i.e. 33 de 45 estudiantes
contestaron correctamente I.F (33/45) =0,73**

.30 -.70	Ítems óptimos
.20-.80	ítems buenos atendiendo a la discriminación (el índice de discriminación tiene que ser bueno) y a que aporten consistencia a la prueba

Índice de discriminación del ítem:

capacidad del ítem para diferenciar a los candidatos

>40	óptima
.30-.39	Razonablemente buena
.20-29	Marginal (debe mejorarse el ítem)
<19	A mejorar o descartar

Funcionamiento de

distractores: nos permite saber si el distractor funciona como tal.

i.e. 100 alumnos

6 eligieron A = 6%

34 eligieron B = 34%

39 eligieron C = 39%

21 eligieron D = 21%

<7% (no funciona)

(puede usarse 10% con ítems de 4 opciones)

Análisis a nivel de ítem
Identificación y revisión de ítems problemáticos

Consideraciones generales

Los ítems demasiado fáciles pueden no discriminar pero no necesariamente serán ítems problemáticos más allá de su facilidad.

Según el tipo de examen pueden considerarse aptos o no.

Señales de alarma

Discriminación negativa

IF adecuado pero discriminación $<.2$

.30 -.70	Ítems óptimos
.20-.80	Ítems buenos atendiendo a la discriminación (el índice de discriminación tiene que ser bueno) y a que aporten consistencia a la prueba

>40	óptima
.30-.39	Razonablemente buena
.20-29	Marginal (debe mejorarse el ítem)
<19	A mejorar o descartar



Ítem 27

	IF	ID
Ítem 27	0,0	-0,1

Ítem 87

	IF	ID
Ítem 87	0,4	0,0

Ítem 103

	IF	ID
Ítem 103	0,7	0,1



Análisis a nivel de ítem

Identificación y revisión de ítems problemáticos

27. We don't have _____ biscuits left, but I can offer you some cake.

- a) some
- b) any
- c) Many

87. He's never _____ to London

- a) gone
- b) went
- c) been

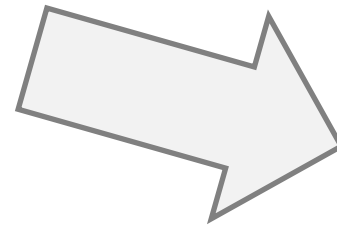
103. I haven't got my purse! I must have forgotten _____ when we were in the restaurant.

- a) to pick it up
- b) picking it up
- c) having picked it up



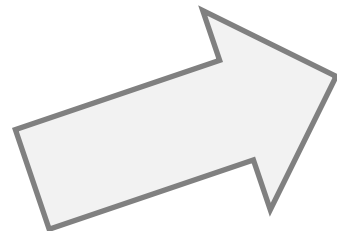
Conclusiones

Análisis de datos
estadísticos



**ATENCIÓN A LOS PRINCIPIOS DE
EVALUACIÓN**

Establecimiento de
estándares (juicio de
expertos, anclaje al
marco)



Impacto positivo

Responsabilidad social



Como
evaluador



Como
institución





Bibliografía

Libros

Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford University Press.

Green, R. (2013). *Statistical analyses for language testers*. Springer.

Bachman, L. F., & Kunnan, A. J. (2005). *Statistical Analyses for Language Assessment Workbook and CD-ROM*.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Recursos web

Fulcher, G. *Language Testing Resources* <http://languagetesting.info>

Discovering statistics: <https://www.discoveringstatistics.com/statistics-hell-p/>



Muchas Gracias

Julia Zabala juzadel@upv.es
Centro de Lenguas
UPV

