



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE PSICOLOGÍA

MODELOS ORDINALES BAYESIANOS:
UNA APLICACIÓN EN EL ESTUDIO SOBRE JUICIOS
MORALES

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Licenciado en Psicología

PRESENTA:

Alfonso Medina Velazquez

DIRECTOR DE TESIS:

Dr. Luis Rodolfo Bernal Gamboa

REVISOR:

Dra. Patricia Romero Sánchez

SINODALES:

Dr. Diego Armando Luna Bazaldua

Lic. Jose Luis Baroja Manzano

Lic. Rafael Cruz Velazquez



Ciudad de México, 2019



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

A mi madre, Leticia, a mi padre, Miguel Ángel, y a mi hermana, Marisol, quienes me lo enseñaron todo y a quienes les debo todo lo que hoy soy. Muchas gracias por todo, los amo y atesoro enormemente.

A mi familia, especialmente a mi abuela, a mi tía Chuy y a mi tía Gaby quienes siempre me ha cuidado sin importar la distancia; a Omar y Diana, quienes han sido como hermanos para mi y a quienes les debo gran parte de mi infancia.

A mis amigos más cercanos, Armando, Charly, Javier, José Juan, Rodrigo y Mauricio, quienes siempre me hacen reír y pasar un buen rato sin importar las circunstancias. Anhele pasar más tiempo con ustedes (como aquellas tardes de fútbol).

A los amigos que hice en la facultad de psicología, Alan, Atena, Alma, Chay, José Luis, Uriel, Karime, Ana, Erika, Victoria, Paulina y Ximena, los quiero y extraño enormemente. A Paola y Yuznhio quienes desde el principio de la carrera me acompañaron desde cerca y que aún forman parte importante de mi vida.

A las personas que conocí por cuestiones del azar, a Danilo por enseñarme a romper la vida, y a Gisela por su cariño, paciencia y comprensión, te quiero.

A las personas que inspiraron y forjaron el camino profesional que hoy intento seguir. A el Doctor Arturo Bouzas Riaño, una de las personas más brillantes que he tenido el placer

de conocer; y a Manuel Villarreal (mi padre académico), Jose Luis Baroja y Dario Trujano, quienes fueron los primeros en darme las herramientas que hoy me han llevado a donde estoy.

A mis asesores, el Dr. Rodolfo Bernal, la Dra. Patricia Romero, el Dr. Diego Luna, el Lic. Jose Luis Baroja y el Lic. Rafael Cruz, quienes me ayudaron a hacer de este un mejor trabajo. Muchas gracias por su apoyo, guía y comprensión.

Y especialmente a Josué Mendoza quien me acompañó en mis primeras incursiones laborales y de posgrado. Sin su ayuda, inspiración y guía este proyecto jamás habría podido ser, se tiene bien merecido el doctorado de la amistad.

Índice general

1. Introducción	7
2. Marco teórico	10
Stevens y escalas de medición	10
Juicios Morales	17
El «default» métrico	19
Problemas de los modelos métricos	22
Modelos Ordinales	25
Modelos ordinales acumulativos	25
La perspectiva bayesiana	34
Selección de Priors	41
3. Método	45
Bases de datos	45
Herramientas	48
Modelos ajustados	48
Modelo más simple	48

Modelos Lineales	51
4. Resultados	54
Exploratorios	54
Evaluación de Modelos	59
Modelo más simple	59
Modelos lineales	61
5. Conclusiones	75
A. Aspectos Técnicos	78
B. Recursos en Línea	83
Referencias	84

Índice de tablas

2.1. Escalas de medición Stevens (1946)	14
4.1. Exploración: Experimentos Pryor, 2019 [1]	55
4.2. Exploración: Experimentos Pryor, 2019 [2]	55
4.3. Exploración: Kruschke, 2015	58
4.4. Estimaciones Posteriores Métricas: Pryor, 2019	62
4.5. Estimaciones Posteriores Ordinales: Pryor, 2019 [1]	63
4.6. Estimaciones Posteriores Ordinales: Pryor, 2019 [2]	64
4.7. Estimaciones Posteriores Ordinales: Pryor, 2019 [3]	64
4.8. Estimaciones Modelos Lineales: Kruschke, 2015	67
4.9. Estimaciones Modelos Ordinales: Kruschke, 2015	69

Índice de figuras

2.1. Modelo Métrico	22
2.2. Media ordinal como función de la media y desviación estandar	23
2.3. Modelo Ordinal	26
2.4. Probabilidad Acumulada y Verosimilitud de respuestas	28
2.5. Ejemplo de datos ordinales generados por el modelo acumulativo normal . .	29
2.6. Ejemplo de regresión ordinal	32
2.7. Ejemplo de actualización Bayesiana	38
4.1. Exploración: Pryor, 2009	56
4.2. Exploración: Respuestas Personalidad	57
4.3. Exploración: Kruschke, 2015	59
4.4. Respuestas Cuestionario (Análisis Métrico)	60
4.5. Respuestas Cuestionario (Análisis Ordinal)	61
4.6. Predicciones posteriores métricas: Pryor, 2019	65
4.7. Predicciones posteriores ordinales: Pryor, 2019	66
4.8. Comparación de modelo: Kruschke, 2015	70
4.9. Predicciones posteriores: Kruschke, 2015	71

4.10. Probabilidades ajustadas 72

4.11. Predicciones posteriores 73

Resumen

La aproximación mecanicista y algorítmica hacia la estadística, tanto en su aplicación como en su enseñanza, ha sido una de las principales causas de la crisis de reproducibilidad de la que padece la psicología. Un ejemplo particularmente importante por su ubicuidad es el default métrico para el análisis de datos de escala ordinal (i.e., el trato como variables intervalares o de razón); sin embargo el uso de estos análisis genera sistemáticamente tasas altas de error Tipo I, pérdida de poder, e inversión de las inferencias. De esta forma, los objetivos que aquí se plantean son el discutir y explicar a detalle los problemas conceptuales y prácticos de la aproximación métrica, y presentar una alternativa de análisis para este tipo de datos, los modelos ordinales acumulativos probit. Además, se aporta una herramienta programada de fácil acceso para la implementación de los modelos presentados dentro del marco de la inferencia bayesiana.

Palabras clave: variables ordinales, psicología y estadística, escalas ordinales, estadística bayesiana, modelos ordinales

Capítulo 1

Introducción

Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad ¹ (Box, 1976, pag. 792).

El uso de modelos (representaciones simplificadas de la realidad que capturan ciertos aspectos importantes de esta) cuantitativos (modelos expresados matemáticamente) se ha vuelto una parte importante de la investigación en psicología (Dienes, 2008; Rodgers, 2010). Dado que la gran mayoría de fenómenos que interesan a la psicología no son deterministas los modelos que los representan y las hipótesis que estos genera tampoco pueden serlo (e.g., «toda x es y » o «siempre que ocurre x le sigue y »), sino que deben ser representados por modelos probabilísticos (e.g., que establezcan relaciones del tipo: «un aumento en x tendrá un efecto positivo sobre el valor esperado de y ») que deben ser puestos a prueba, a través de sus predicciones, por la evidencia generada con el fin de obtener inferencias en la presencia

¹Dado que todos los modelos son incorrectos, el científico debe estar alerta a lo que es importantemente incorrecto. Es inapropiado preocuparse por los ratones cuando hay tigres afuera

de la variabilidad e incertidumbre intrínseca de los fenómenos. La estadística, en este sentido, es una herramienta fundamental para esta labor.

A pesar de ello, la estadística usualmente se enseña y se utiliza como si se tratara de un libro de recetas (Abelson, 2012; Dienes2008; McElreath, 2016), una especie de colección de cajas negras entre las que se tiene elegir la más adecuada según una serie de preguntas (e.g., ¿las variables son continuas o categóricas?, ¿hay conteos menores a 5 en alguna de las categorías?, ¿cuántos grupos se tienen?, ¿se cumplen los supuestos paramétricos?, etc.) promoviendo la eliminación del juicio en puntos clave donde la teorías estadísticas lo exige (e.g., al elegir una hipótesis nula, al establecer el punto de rechazo de esta, o incluso al establecer conclusiones sobre la hipótesis de investigación). El mismo Fisher (1995), uno de los estadísticos más importantes del siglo XX, reprobaba la aproximación mecanicista o algorítmica hacia la estadística, a la vez que enfatizaba (al igual que Box, 1976) la experimentación como un proceso continuo que requería de una comunidad de mentes libres tomando sus propias decisiones dentro de un constante ir y venir entre teoría y práctica.

Gigerenzer (2018) llama a la práctica algorítmica «rituales estadísticos» y los coloca como una de las principales causas, y principal factor endógeno a la práctica, de la crisis de reproducibilidad por la que atraviesa la psicología (Baker, 2016; Pashler & Wagenmakers, 2012). Un ejemplo especialmente importante para la disciplina (dada su extendida presencia en el área) de estos rituales es el uso de modelos métricos (i.e., que asumen variables de intervalo o razón; como el cálculo de medias, desviaciones, pruebas t , regresión lineal, etc.) para el análisis de datos ordinales (Bishop & Herron, 2015; Jamieson, 2013; Knapp, 1990; Kuzon, Urbanchek, & McCabe, 1996).

Esta aproximación de análisis presenta dos grandes problemas: el primero, de carácter conceptual, proviene de pasar por alto las razones subyacentes a la categorización de las escalas de medición presentada por Stevens (1946), que explican por qué el cálculo de, por ejemplo, promedios y desviaciones es «importantemente incorrecto»; el segundo, de carácter práctico, es que este trato genera distorsiones en el tamaño del efecto (tasas altas de error Tipo I), pérdida de poder (debido al aumento de error Tipo II) y que se invierta el sentido de las inferencias (e.g., $\mu_1 < \mu_2$ cuando $\mu_1 > \mu_2$) (Liddell & Kruschke, 2018 recientemente demostraron que el análisis métrico sistemáticamente produce estos errores).

Esta tesis busca realizar aportaciones en estos dos sentidos: en lo conceptual, al discutir y explicar a detalle los problemas de la aproximación métrica, así como presentar una alternativa de análisis bayesiano para datos ordinales; y en lo práctico, al presentar herramientas (gratuitas y de fácil acceso) para analizar este tipo de datos; y es que, sumado a lo dicho hasta ahora, es importante remarcar que, aunque en los últimos años la discusión sobre el trato de las variables ordinales se ha mantenido constante, poco se ha escrito en español sobre el tema.

Capítulo 2

Marco teórico

Stevens y escalas de medición

As a matter of fact, most of the scales used widely and effectively by psychologists are ordinal scales. In the strictest propriety the ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply a knowledge of something more than the relative rank-order of data. On the other hand, for this 'illegal' statisticizing there can be invoked a kind of pragmatic sanction: In numerous instances it leads to fruitful results ¹ (Stevens, 1946, pag. 679).

En este extracto del artículo del psicólogo norteamericano S. S. Stevens (1946) «*On the Theory of Scales of Measurement*», hay tres puntos importantes que merece la pena resaltar y

¹De hecho, la mayoría de las escalas utilizadas amplia y efectivamente por psicólogos son escalas ordinales. En el sentido más estricta, las estadísticas comunes que involucran medias y desviaciones estándar no deben usarse con estas escalas, ya que estas estadísticas implican un conocimiento de algo más que el orden relativo de los datos. Por otro lado, para esta estadística ilegal se puede invocar una especie de sanción pragmática: en numerosos casos conduce a resultados fructíferos

que constituyen los motivos subyacentes a éste proyecto:

1. «*Most of the scales used widely and effectively by psychologists are ordinal scales*»² y que en realidad se puede extender al resto de las ciencias sociales así como a la investigación en diferentes disciplinas (e.g., educación, medicina, etc.).
2. «*The ordinary statistics involving means and standard deviations ought not to be used with these scales*»³ ya que, como el mismo Stevens continua, «*these statistics imply a knowledge of something more than the relative rank-order of data*»⁴.
3. «*In numerous instances it leads to fruitful results*»⁵, pero que en otras puede llevar a conclusiones erradas (e.g., caer en los errores Tipo I [rechaza la hipótesis nula cuando de hecho esta es cierta] o Tipo II [no rechaza la hipótesis nula cuando de hecho esta es falsa]).

El primer punto hace referencia a la discusión que impulsaría a Stevens a escribir su artículo de 1946, aquel en el que introduciría la clasificación de escalas de medición ampliamente usada en la actualidad. En 1932 un comité, representante de las secciones A (ciencias físico-matemáticas) y J (psicología), de la Asociación Británica para el Avance de la Ciencia se encargaba de la ardua tarea que era el considerar y reportar sobre la posibilidad de medir eventos sensoriales; pero ¿cómo medir las capacidades, actitudes, sensaciones o cualquier aspecto del ser humano que no puede pesarse o graduarse?

The task is not simply to read a meter or gauge an effect; it is to devise a procedure by which to quantify some stubborn phenomenon, to reduce it to

²La mayoría de las escalas utilizadas amplia y efectivamente por psicólogos son escalas ordinales.

³Las estadísticas comunes que involucran medias y desviaciones estándar no deben usarse con estas escalas.

⁴estas estadísticas implican un conocimiento de algo más que el orden relativo de los datos.

⁵En numerosos casos conduce a resultados fructíferos

numerical order ⁶ (Stevens, 1958, pag. 383).

Sin embargo, tras siete años de discusión, el reporte final del comité sólo acuerda el presentar un ejemplo de una escala sensorial, la medición de la magnitud subjetiva de una sensación auditiva. La discordancia giraba alrededor del concepto de medición y sus implicaciones sobre una escala sensorial (e.g., ¿qué aplicación y significado tiene la suma en una escala así?). A estas interrogantes Stevens escribe que quizás el acuerdo se alcance cuando se reconozca que la medición existe en una variedad de formas y que las escalas de medición caen dentro de ciertas clases definidas, clases que están determinadas por (1) las operaciones empíricas al momento de «medir» y (2) por las propiedades formales de la escala resultado de tal proceso (Stevens, 1946, 1958). Estos dos aspectos están implícitos en el segundo punto destacado de la cita del inicio («*The ordinary statistics involving means and standard deviations ought not to be used with these scales*»).

Detrás del primero de estos aspectos, y en relación al desacuerdo del comité de la Asociación Británica para el Avance de la Ciencia, se encuentra el concepto de medir ¿qué se debería entender por «medir»? A esto Stevens responde: «*In the broadest sense, measurement is the business of pinning numbers on things. More specifically, it is the assignment of numbers to objects or events in accordance with a rule of some sort*»⁷ (Stevens, 1958, pag. 384), reglas explícitas que establecen cierto nivel de isomorfismo entre la relación empírica de los objetos o eventos y algunas de las propiedades del sistema numérico. Algunas de éstas propiedades, y su uso en la medición son:

⁶La tarea no es simplemente leer un medidor o medir un efecto; es idear un procedimiento mediante el cual cuantificar algún fenómeno elusivo, reducirlo a un orden numérico

⁷En el sentido más amplio, la medición es la tarea de fijar números a las cosas. Más específicamente, es la asignación de números a objetos o eventos de acuerdo con una regla de algún tipo

- Identidad: los números sirven como etiquetas para identificar elementos o clases (e.g., numerar los diferentes jugadores en un equipo deportivo).
- Orden: los números reflejan el orden jerárquico de los elementos (e.g., las posiciones finales de los competidores en una carrera).
- Intervalos: los números determinan igualdad de intervalos o diferencias entre los elementos (e.g., las escalas de temperatura [Fahrenheit y Celsius]).
- Razones: Los números reflejan proporciones entre los elementos (e.g., una escala de luminosidad o volumen).

El segundo aspecto clave, y resultado del proceso de medición, es la propiedad de invarianza (Stevens, 1946, 1958, 1968) de las escalas obtenidas. Esto se refiere al tipo de transformaciones matemáticas sobre los números de la escala que no representan una pérdida de información; por ejemplo, si los números son utilizados como etiquetas para distinguir a los jugadores de un equipo de fútbol la única transformación que se puede hacer sobre estos números es la de permutarlos, cualquier otra manipulación estaría suponiendo otro tipo de información que no se tomó en cuenta al momento de medir (e.g., los números no reflejan que tan bueno es un jugador respecto a otro).

La tabla 2.1 muestra la clasificación de escalas según estos dos criterios.

La primer columna muestra el nombre de cada uno de los tipos de escala propuestos por Stevens. La segunda columna presenta las operaciones básicas (reglas) necesarias al momento de medir (e.g., los números en la escala nominal se asignan según los elementos sean iguales o diferentes), esta columna es acumulativa por lo que las operaciones listadas más arriba en la tabla también son necesarias para crear las escalas de más abajo (e.g., para determinar

Tabla 2.1: Escalas de medición Stevens (1946)

Escala	Operaciones	Transformaciones	Estadísticos
Nominal	Igualdad	$y = f(x)^*$	Frecuencias, moda
Ordinal	Mayor o menor	$y = f(x)^\dagger$	Mediana, percentiles
Intervalo	Igualdad de intervalo o diferencias	$y = ax + b$	Media, desviación estándar
Razón	Igualdad de razones	$y = ax$	Coefficiente de variación [‡]

* Con $f(x)$ cualquier sustitución uno-a-uno

† Con $f(x)$ cualquier función monotónica creciente

‡ Dispersión relativa del conjunto de datos

el ordenamiento de los elementos antes es necesario identificar que estos son diferentes). La tercer columna presenta el grupo de transformaciones que dejan invariante la escala, estas no son acumulativas (e.g., $f(x) = x^2$ mantiene el orden de los elementos, pero no la igualdad de intervalos entre los mismos). Finalmente, la cuarta columna contiene el tipo de estadísticos (principalmente descriptivos) válidos para cada escala, estos también son acumulativos hacia abajo, en el mismo sentido que la segunda columna (Allen & Yen, 1979; McDonald, 2011; Stevens, 1946, 1958, 1968).

Esta última columna es resultado de las restricciones impuestas por las dos anteriores (las operaciones empíricas necesarias para crear la escala y las propiedades formales de la misma); por ejemplo, una variable ordinal surge de la asignación de números que representen mayor o menor cantidad de un atributos (e.g., una película con 4 estrellas tiene más del atributo que la clasifica como mejor que una película con 3 estrellas, cualquiera que sea

dicho atributo), y esta escala permanecerá invariante ante transformaciones que conserven el orden relativo de las categorías (e.g., si se multiplica todos los números en la escala por una constante positiva el orden entre estos se mantiene), pero para poder ser considerada una escala de intervalo se agrega la necesidad de un cero arbitrario y una unidad de medida (aunque esta sea arbitraria) que se mantenga constante en toda la escala. Con estas nuevas características sobre la escala es posible establecer la igualdad de intervalos o la diferencia entre los mismos (e.g., la diferencia entre 40° centígrados y 50° centígrados es la misma que entre 50° centígrados y 60° centígrados), aunque aún no es posible determinar una relación de proporciones (e.g., 80° centígrados no es «el doble de caliente» o «la mitad de frío» que 40° centígrados) con ella (McDonald, 2011; Stevens, 1946, 1958).

Sobre estas restricciones es que se fundamenta la prohibición de Stevens aludida en la cita del principio de este capítulo, y es que la manipulación estadística que puede ser aplicada legítimamente a datos empíricos depende (en parte) del tipo de escala contra la cual se ordenan los datos; así, con datos medidos a nivel ordinal, donde la diferencia entre valores de la escala no necesariamente es igual, los estadísticos descriptivos se ven limitados al cálculo de medianas, modas y percentiles. Como Kuzon, Urbanchek y McCabe (1996, pag. 266) expresan, «*It is invalid to rank the results of a given surgical procedure as poor, fair, good, or excellent and state that the average result is fair-and-a-half*»⁸; incluso si se utiliza el 2 para representar *fair* (regular) y 3 para representar *good* (bueno) sería igual de inválido decir que el promedio es 2.5.

Es importante resaltar que la clasificación de Stevens, aunque de suma utilidad y de

⁸No es válido clasificar los resultados de un procedimiento quirúrgico como malo, regular, bueno o excelente, y establecer que el resultado promedio sea regular-y-medio.

uso extensivo, no es la única forma de categorizar y describir niveles de medición en una variable; por ejemplo, Crocker y Algina (2008) abordan los enfoques básicos de Torgerson para el desarrollo de escalas al implementar un procedimiento de evaluación psicológica, y en McDonald (2011) se puede encontrar una detallada explicación de la clasificación de los tipos de observación en la medición psicológica de Clyde Coombs. Además, existen diversos métodos que permiten medir (en el sentido de determinar reglas para la asignación de números) fenómenos propios de la psicología, como lo son los juicios subjetivos de propiedades físicas de los estímulos, en escalas de intervalo, a través de estimación directa, bisección y el método de juicios comparativos de Thurstone, y de razón, a partir del método de estimación directa (Allen & Yen, 1979). Sin embargo, para los fines de esta tesis la taxonomía de Stevens resulta conveniente por tres razones fundamentales:

1. Es usualmente el punto de referencia ya que introduce los niveles de medición más conocidos (Nominal, Ordinal, Intervalar y Razón), así como la idea focal detrás de estos, la de representatividad (¿qué tan bien la medición representa la propiedad o rasgo medido?), aspecto fundamental para la evaluación de algunos fenómenos propios de la psicología.
2. Sirve como introducción al debate sobre el nivel de medición que se alcanza con las pruebas psicológicas y su correcto análisis, ya que al introducir la clasificación de Stevens también se fundamenta la prohibición que se le atribuye.
3. Permite ejemplificar el primero de los problemas que motivan esta tesis (y que tampoco es ajeno a las aproximaciones alternas a la de Stevens), el problema de ignorar o desconocer el razonamiento detrás de una escala de medición.

Un campo de investigación que conjunta los tres puntos anteriores, especialmente por la complejidad y multidimensionalidad de su objeto de estudio, es el de la investigación sobre juicios morales; es decir, (1) es de interés medir un aspecto que no es directamente observable, (2) esta medición comúnmente se hace a través de reactivos tipo Likert, y (3) el análisis de los datos obtenidos es principalmente a través de modelos métricos.

Juicios Morales

El estudio sobre la intuición moral; es decir, las formas de razonamiento a través de las cuales las personas desarrollan juicios sobre la «bondad» o «maldad» moral de las acciones (Cushman, Young, & Hauser, 2006; McElreath, 2016), es de gran relevancia para todas las ciencias sociales ya que combinan diferentes aspectos concernientes a estas como lo son el razonamiento, el rol de las emociones en la toma de decisiones y el desarrollo social y moral (tanto a nivel individual como a nivel grupal). La mayor parte de la investigación en este campo se centra en dilemas morales en los que los juicios deontológicos (i.e., reglas, derechos y deberes individuales) están en conflicto con los juicios utilitarios (i.e., seguir el bien mayor definido por las consecuencias); por ejemplo, en el paradigma más común en este campo de estudio, conocido como «Trolley problem» o «Problema del Tranvía» (Bago, Aczel, Zoltan, & Kovacs, 2019; Cushman et al., 2006; McElreath, 2016):

Usted ve un tranvía fuera de control que se mueve hacia cinco personas amarradas que yacen en las vías. Se encuentra parado junto a una palanca que controla un interruptor; si tira de la palanca, el carro se redirigirá a una vía lateral salvando así las cinco personas en la vía principal. Sin embargo, hay una persona tendida en la vía lateral. Tienes dos opciones:

- . No haga nada y permita que el carro arrolle a las cinco personas en la

pista principal.

- . Tirar de la palanca, desviando el carro hacia la vía lateral donde arrollará a una persona.

Desde la perspectiva deontológica, se puede argumentar que tirar de la palanca es moralmente inaceptable, como lo sería un asesinato; desde la perspectiva utilitarista, se podría decir que es moralmente aceptable tirar de la palanca, ya que maximizaría el número de vidas salvadas. Como se puede apreciar este tipo de planteamientos pueden ser desconcertantes en el sentido de que el contenido analítico de dos escenarios puede ser idéntico (el total de vidas salvadas y perdidas), pero las personas llegan a diferentes juicios sobre qué tan moralmente aceptada es la misma acción en los diferentes escenarios (Bago et al., 2019).

Investigaciones anteriores han identificado que las variaciones de cómo evalúan las personas estas situaciones se deben en gran medida a factores psicológicos y situacionales; por ejemplo, la intención del agente, el contacto físico entre el agente y la víctima, si el daño es causado por acción directa u omisión (Bago et al., 2019; McElreath, 2016), o si se tiene información de lo que otros han hecho (Pryor, Perfors, & Howe, 2019). Respecto a este última existen tres explicaciones posibles en la literatura (Pryor et al., 2019):

1. La conducta de otros proporciona información sobre lo que probablemente sea un comportamiento fructífero.
2. Así se reduce la posibilidad de sanciones sociales.
3. La conducta es propia de un grupo con el que el individuo se identifica.

El procedimiento típico en los estudios sobre juicios morales consiste en presentar a

los participantes un dilema moral como el expuesto arriba, pedirles que elijan una acción o se imaginen un escenario en el que han tomado una acción dada (se les «impone» dicho escenario), y finalmente se les pide que indiquen, en una escala numerada tipo Likert, la medida en que piensan que la acción dada es moralmente aceptable (e.g., Bago et al., 2019; Pryor et al., 2019)

El tratamiento común (como en muchos otros campos de estudios) de los datos obtenidos es utilizar pruebas t , para la comparación de medias, ANOVAS, para los diseños factoriales, o análisis de regresión, para relacionar la respuesta con alguna variable como la edad, el sexo o la creencia religiosa.

El «default» métrico

La práctica más común para el trato de datos ordinales (e.g., escalas tipo Likert) es la de hacer caso omiso de las restricciones impuestas por las operaciones empleadas al momento de realizar las mediciones y utilizar análisis métricos (i.e., empleo de estadísticos y modelos válidos sólo para variables de intervalo o razón [variables métricas a partir de ahora]; como pruebas t , regresión lineal, análisis de varianza, correlaciones, etc.), como es el caso de las mediciones de juicios morales. Por ejemplo, Liddell y Kruschke (2018) mencionan que de 68 publicaciones de las revistas de psicología con mayor factor de impacto, **todas** emplearon modelos métricos para el análisis de datos ordinales. Pero ¿qué podría salir mal al tratar datos ordinales como métricos?

Algunos de los argumentos más comunes que se utilizan a favor de esta práctica (e.g., en Carifio & Perla, 2007, 2008; Crocker & Algina, 2008; McDonald, 2011; Norman, 2010; Solís,

2014) son:

1. Las escalas Likert (i.e., conjunto de reactivos Likert) sí son intervalares.
2. Así se ha hecho siempre.
3. Ha dado resultados fructíferos.
4. Son válidos en la medida en la que se cumplen los supuestos estadísticos.

El primer punto defiende que las escalas Likert, no así los reactivos Likert, son intervalares en naturaleza, y por lo tanto pueden ser analizados con modelos métricos. Éste es un punto interesante, ya que se hace alusión a «propiedades emergentes de la escala» (Carifio & Perla, 2007, 2008) que no existen a nivel de reactivo, es decir, a pesar de que a nivel molecular la escala es ordinal al momento de combinar la información de varios de estos elementos, al calcular su media por ejemplo, se obtiene una escala intervalar, manipulación que sigue siendo inválida ya que lo que define a una escala no es cómo se comporta (e.g., si su distribución parece continua y con forma de campana gaussiana) sino las operaciones de medición realizadas (segunda columna de la Tabla 2.1). Sumado a esto, Liddell y Kruschke (2018) muestran que el análisis sobre los promedios de variables ordinales presenta los mismos problemas que el análisis a nivel de elementos.

Además, para promediar los items individuales es necesario sumarlos ¿cómo se pueden sumar dos números cuya distancia es arbitraria, que solo tienen que garantizar el orden? por ejemplo, si las operaciones de medición realizadas sólo nos permiten saber que $a > b$, esta relación puede ser representada con $a = 2$, $b = 1$ o $a = 1/2$, $b = 100$, la escala es arbitraria; y la suma de $a + b$ podría representar $2 + 1$ o $1/2 + 100$, dependiendo de quién defina

(arbitrariamente) la escala. Es decir, para sumar dos reactivos es necesario saber algo más que una relación de orden.

El segundo punto deriva de la mala interpretación, o incluso total omisión, de lo ya discutido respecto a la relación entre los análisis válidos para cada tipo de escala, su medición (regla de asignación de números) y su propiedad de invarianza. Por ejemplo, Solis (2014) escribe «*No centramos nuestra atención en las propiedades de las escalas per se, ni en la correspondencia que, algunos **conjeturan**, guardan con diversos tipos de análisis*», pasando por alto lo ya discutido y presentándolos como si los análisis pertinentes a cada escala fueran un capricho de Stevens. Además, a la luz de la reciente crisis de reproducibilidad, donde el uso inadecuado de métodos estadísticos por parte de la mayoría de investigadores ha sido señalada como una de sus principales causas (Gigerenzer, 2018), este argumento pierde validez; de hecho parte del problema es que los investigadores siguen «rituales» estadísticos («así se ha hecho siempre») sin comprender el funcionamiento y limitaciones de los instrumentos de inferencia que utilizan.

El tercer punto se refiere a la evidencia empírica que respalda el uso de estos modelos, esta evidencia entendida como el hallazgo de una diferencia o efecto «estadísticamente significativo». Y el cuarto pone el énfasis en otro aspecto que sin duda también es clave para el empleo de los modelos métricos, justificando su uso en la medida en la que los supuestos del modelo se cumplen de una manera razonable, restando importancia a los niveles de medición. De estos dos últimos argumentos se deriva el segundo problema que motiva esta tesis.

Problemas de los modelos métricos

EL funcionamiento de los modelos métricos se describe en la figura 2.1. Los valores ordinales de la respuesta (Y) se toman como valores numéricos en el eje horizontal y la probabilidad de cada uno de estos es la densidad normal en ese punto (i.e. la altura de las barras), calculadas como:

$$P(Y = k|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(k-\mu)^2}{2\sigma^2}} \quad (2.1)$$

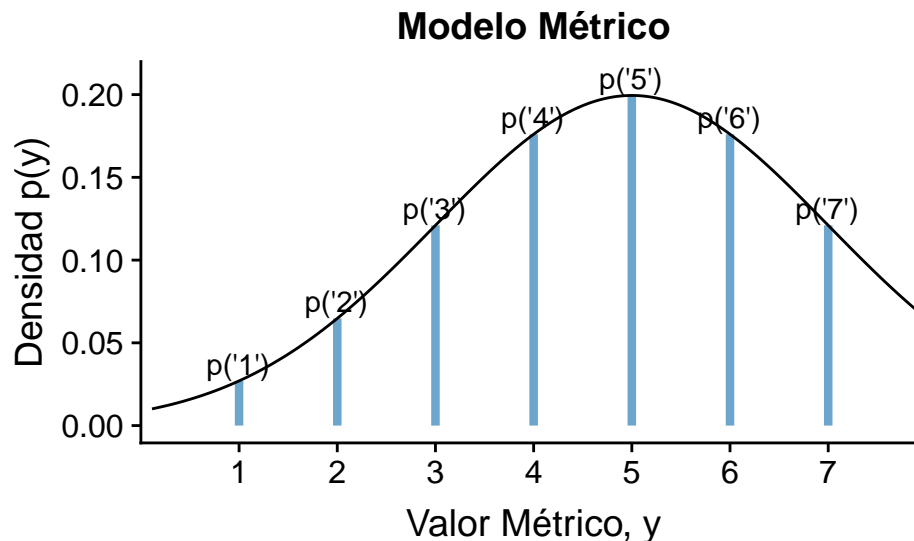


Figura 2.1: Modelo Métrico

Liddell y Kruschke (2018) resumen los errores que se pueden cometer al usar modelos métricos sobre variables ordinales en la siguiente figura; en esta se muestran curvas para dos valores de σ (desviación estándar latente). La curva para $\sigma = 1$ es más pronunciada que la de $\sigma = 4$, esto porque una desviación pequeña indica que los valores más probables serán aquellos más cercanos al valor de la media (μ), mientras que valores grandes darán paso a un rango

más amplio de valores.

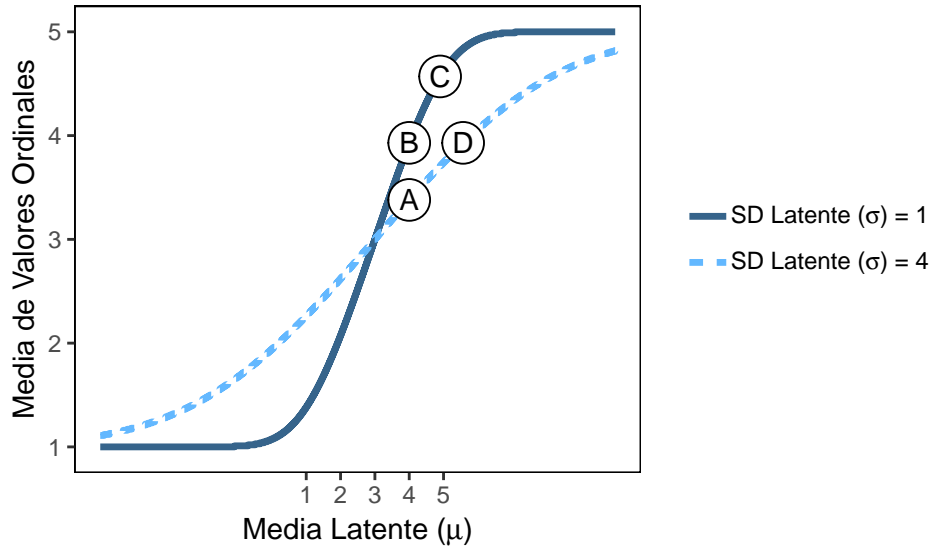


Figura 2.2: Media ordinal como función de la media y desviación estándar

En la figura 2.2 cada punto marcado con una letra es una combinación particular de μ y σ y corresponde a un patrón particular de datos ordinales que serían generados por estos parámetros. El eje vertical muestra la media de los valores ordinales, tomados como métricos, para cada uno de esos patrones (i.e. para cada combinación de μ y σ). Los puntos *A* y *B* representan dos grupos con la misma media latente (i.e. están ubicados en el mismo punto del eje horizontal), pero su medias de los valores ordinales resultantes no lo son; es decir, la estimación métrica de las medias (calculadas como la esperanza matemática) de los grupos *A* y *B* es diferente cuando en realidad ambos fueron generados por el mismo valor de μ (Error Tipo I o falsa alarma). Los puntos *B* y *D*, por otro lado, representan dos grupos con medias ordinales iguales, pero con medias latentes distintas, en este caso la estimación métrica pasaría por alto una diferencia latente real (Error Tipo II u omisión).

Dado el caso en el que los grupos tienen la misma varianza latente, como el caso de

A y D , la verdadera separación de las medias latentes en el eje horizontal es mucho más grande que la separación (comprimida por la sigmoïdal) de las medias ordinales en el eje vertical (nótese las escalas de los ejes), llevando así a una pérdida de poder para pruebas finita (Liddell & Kruschke, 2018).

Finalmente, los grupos representados por C y D muestran una inversión del efecto. La diferencia indicada por el modelo métrico es opuesta a la diferencia real entre las medias latentes ($C > D$ para la estimación métrica y $C < D$ para la diferencia latente).

En resumen, muchos problemas pueden surgir al tratar datos ordinales como métricos, incluyendo falsas alarmas (A vs B en 2.2), omisiones (B vs D) y distorsión o reversión del efecto estimado (C vs D). El concepto clave, subyacente a este proyecto, se encuentra en la figura 2.2. Nótese que esta figura sugiere que hay un número infinito de combinaciones de parámetro latentes que generan cualquiera de estos problemas.

Una alternativa para el manejo de estas escalas ordinales es ignorar el orden de la variable y tratarla como una variable nominal (usando una regresión logística multinomial), pero esto representa una gran pérdida de información (al omitir la naturaleza ordenada de la variable) y pérdida de poder para el modelo. Cualquiera de las estrategias, ver a la variable ordinal como nominal o intervalar, representa una sub o sobre-estimación (respectivamente) de la información dada por los datos. Sin embargo, existe una tercera aproximación que se ha desarrollado especialmente para modelar este tipo de datos (Agresti, 2007; Fullerton, 2009).

Modelos Ordinales

En realidad esta tercera aproximación se refiere a una vasta familia de modelos paramétricos, generalización de los modelos lineales generalizados, que toman en consideración la naturaleza ordenada y discreta de los datos; estos pueden ser clasificados en los siguientes tres grupos (Fullerton, 2009, según el tipo de comparación entre categorías que se pretende realizar):

- Acumulativos: Y se origina de la categorización de una variable subyacente \tilde{Y} a lo que se mide.
- Secuenciales: Y es resultado de un proceso secuencial (como una serie de pasos).
- Categorías adyacentes: Y se modela como la decisión entre dos categorías adyacentes.

Éste trabajo se centra en la descripción y uso de los primeros, los modelos ordinales acumulativos; ya que (1) la aproximación de las escalas ordinales desde la perspectiva de una variable latente hace mucho sentido en la investigación en psicología (e.g., al medir el grado de satisfacción o felicidad en una encuesta), y (2) son el tipo de datos más ubicuo en psicología (Bürkner & Vuorre, 2019; Stevens, 1946).

Modelos ordinales acumulativos

El supuesto principal de estos modelos es que la variable ordinal observada, Y , resulta de categorizar una variable subyacente continua, \tilde{Y} que no se observa, y que esta está aleatoriamente distribuida entre personas (o entre momentos dentro de una misma persona), de forma que hay K umbrales (τ_k , con $k \in K$) que seccionan a \tilde{Y} en $K + 1$ categorías ordenadas y observables (Bürkner & Vuorre, 2019; Fullerton, 2009; Kruschke, 2015; Liddell

& Kruschke, 2018; McElreath, 2016).

Esta idea se resume en la figura 2.3. En el panel superior se muestra la distribución hipotética de respuestas, el panel inferior muestra la escala métrica latente (\hat{Y}) dividida en intervalos por los K umbrales, marcados con líneas punteadas, en el eje horizontal.

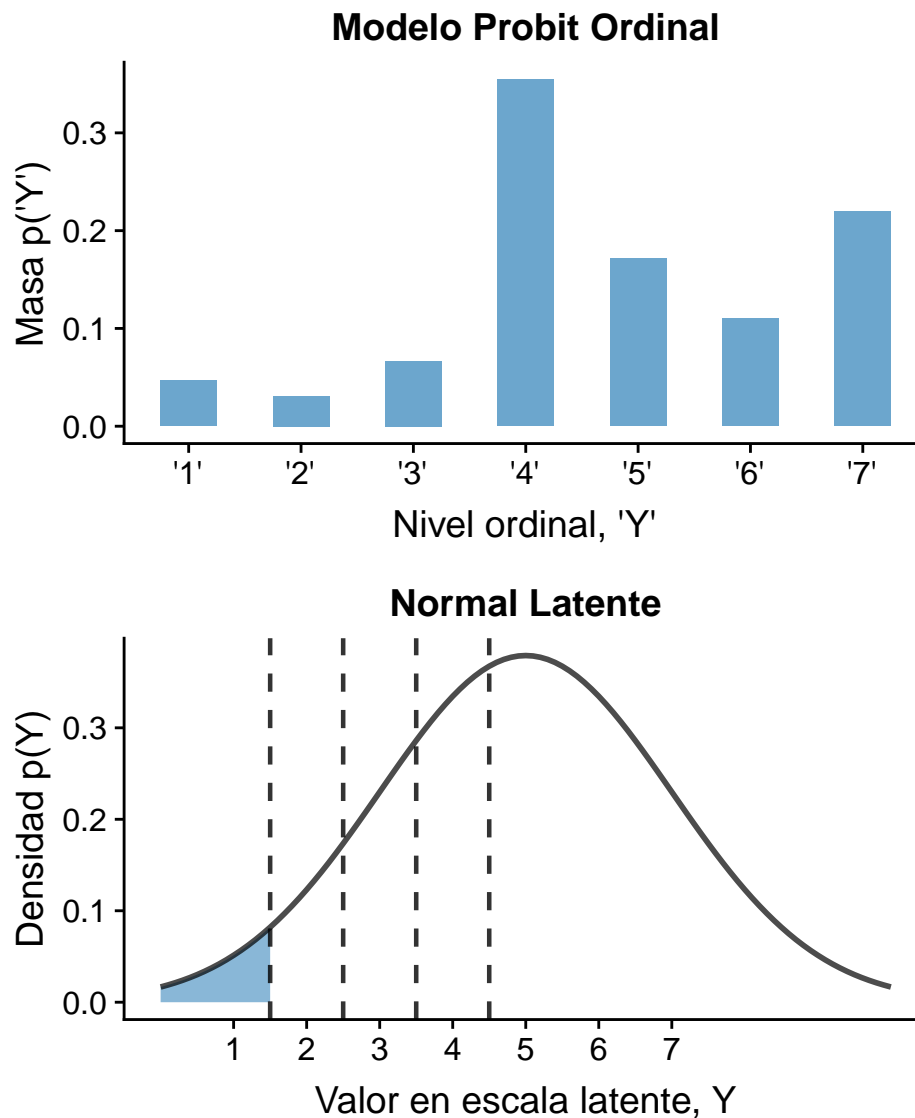


Figura 2.3: Modelo Ordinal

En el panel inferior de la figura 2.3 la probabilidad de que la respuesta observada Y

caiga en cada una de las K categorías es la probabilidad acumulada de la distribución normal subyacente en los intervalos correspondientes; por ejemplo, la probabilidad de la respuesta 1 es el área bajo la curva que se muestra coloreada del lado izquierdo del panel inferior (lo mismo para la respuesta 7 del lado derecho). En términos formales:

$$Pr(Y = k) = Pr(Y \leq \tau_k) - Pr(Y \leq \tau_{k-1}) \tag{2.2}$$

Escrito en términos de una función de distribución acumulada F de \tilde{Y} (i.e., el área bajo la función de densidad de la variable latente continua):

$$Pr(Y = k) = F(\tau_k) - F(\tau_{k-1})^9 \tag{2.3}$$

Este par de ecuaciones se pueden leer como «la probabilidad de que Y sea igual a una categoría, k en específico es igual a la probabilidad acumulada entre el umbral k y el umbral $k - 1$ ».

La figura 2.4 ilustra la idea clave detrás de este par de igualdades. El eje horizontal presenta los valores discretos observados de la respuesta Y , el eje vertical muestra la probabilidad acumulada hasta cada uno de estos valores. Las barras azules representan la probabilidad acumulada de cada valor ($Pr(Y \leq k)$), mientras que los segmentos grises muestran la verosimilitud de cada respuesta calculadas con la ecuación 2.2 (que es función de los umbrales τ_k y τ_{k-1}).

⁹En término de las integrales de su función de densidad de probabilidad: $Pr(Y = k) = \int_{-\infty}^{\tau_k} f(y)dy - \int_{-\infty}^{\tau_{k-1}} f(y)dy$

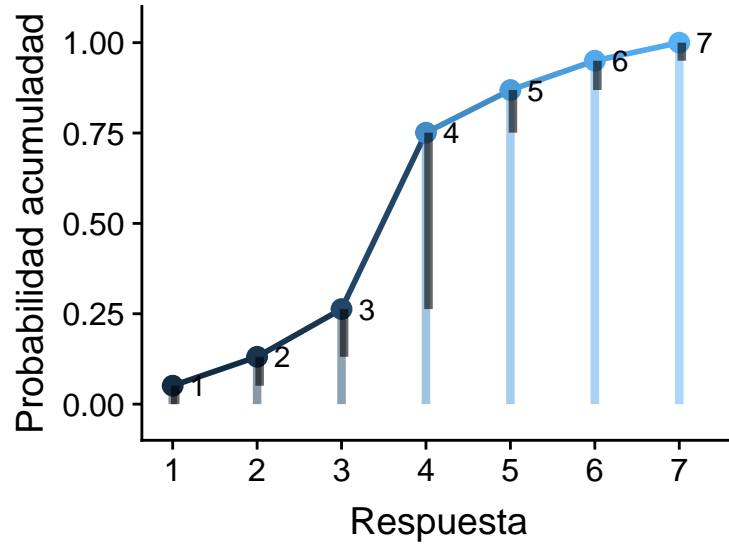


Figura 2.4: Probabilidad Acumulada y Verosimilitud de respuestas

El modelo al que principalmente se hará referencia es al modelo acumulativo con liga Probit, que asume una distribución normal sobre \tilde{Y} (es decir, la función de distribución acumulada, F , es la de una distribución normal, ϕ^{10}). Sin embargo, al observar la distribución empírica de los datos, podría resultar contra-intuitivo que estos provengan de una distribución normal; por ejemplo, se puede pensar en una distribución de respuestas cargada a la izquierda o a la derecha. La figura 2.5 muestra cuatro ejemplos de valores ordinales observados generados por una distribución normal subyacente.

¹⁰ $\phi(\tau) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\tau} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$

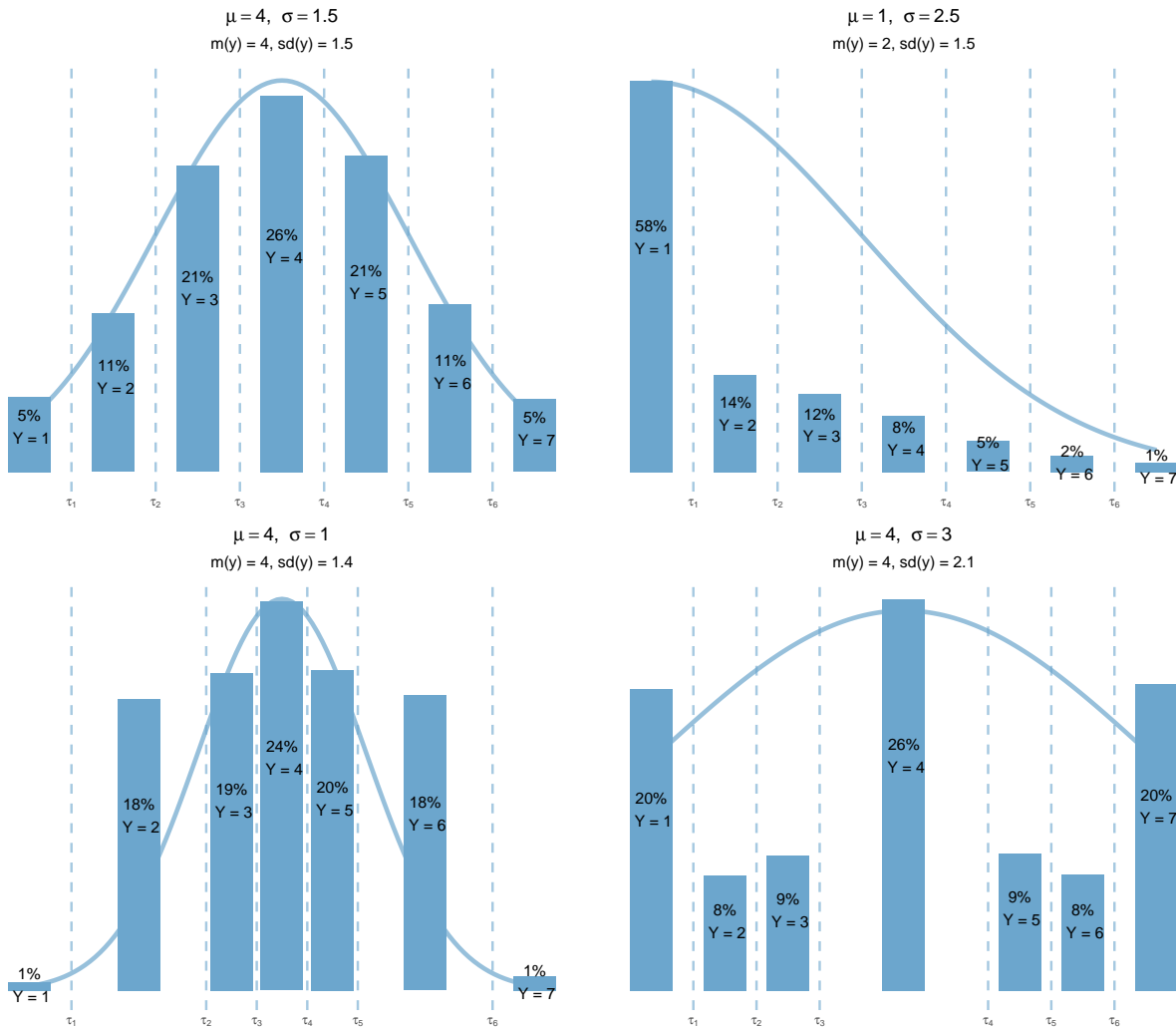


Figura 2.5: Ejemplo de datos ordinales generados por el modelo acumulativo normal

El eje de las x es el soporte de la variable métrica continua (soporte de \tilde{Y}), los umbrales (τ_k) se representan sobre este eje con líneas verticales punteadas. En todos los ejemplos, la escala ordinal tiene $K + 1 = 7$ niveles y $K = 6$ umbrales (nótese que τ_1 y τ_6 se encuentran fijos, en 1.5 y 6.5 respectivamente). Las barras en cada panel representa los porcentajes observados de cada respuesta, la curva detrás de las barras muestra la distribución normal subyacente con media μ y desviación estándar σ (estos parámetros son los que generaron los datos); $m(y)$ y $sd(y)$ muestran la media y desviación calculadas a partir de tratar lo valores

simulados como si estos fueran métricos (i.e., tratando los números como números y no como etiquetas). Las barras verticales indican la proporción de cada valor en la escala ordinal.

El panel superior izquierdo muestra un caso en el que los umbrales se encuentran igualmente separados y la distribución normal está centrada en la mitad de la escala ($\mu = 4$), con una desviación estándar pequeña ($\sigma = 1.5$), y se puede ver que las barras siguen la curva de una normal típica; el panel superior derecho muestra un caso en el que la media de la distribución subyacente se encuentra por debajo del último umbral ($\mu = 1$ y $\tau_1 = 1.5$), por lo que el valor $Y = 1$ acumula la mayoría de la probabilidad, el motivo de observar valores lejos del centro de la distribución es por la desviación estándar grande ($\sigma = 2.5$); en este caso la respuesta ordinal no parece seguir la distribución normal a pesar de haber sido generada por una, por lo que una variable subyacente distribuida normalmente puede generar una distribución no-normal de valores ordinales discretos.

Otro aspecto importante de la figura 2.5 es que se puede visualizar que la probabilidad de un valor ordinal particular es el área bajo la curva entre los umbrales de dicho valor (e.g., la probabilidad de observar el valor 4, es el área bajo la curva entre los umbrales τ_3 y τ_4), como se presenta en la ecuación 2.3. El cálculo de estas probabilidades (a partir de una distribución normal subyacente) se puede expresar como:

$$Pr(Y = k|\mu, \sigma, \tau_K) = \Phi\left(\frac{\tau_k - \mu}{\sigma}\right) - \Phi\left(\frac{\tau_{k-1} - \mu}{\sigma}\right) \quad (2.4)$$

con ϕ la función de distribución acumulativa de una normal con media (μ) cero y desviación estándar (σ) igual a uno (nótese que la función acumulativa se aplica sobre la

¹¹ $Pr(Y = k|\mu, \sigma, \tau_K) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tau_k} e^{-\frac{y^2}{2}} dy - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tau_{k-1}} e^{-\frac{y^2}{2}} dy$

estandarización, cuya distribución es $Normal(\mu = 0, \sigma = 1)$).

Para los valores en los extremos de la distribución subyacente (τ_1 y τ_K) se calcula

$$\begin{aligned}
 Pr(Y = 1|\mu, \sigma, \tau_K) &= \Phi\left(\frac{\tau_1 - \mu}{\sigma}\right) - \Phi\left(\frac{\tau_0 - \mu}{\sigma}\right) \\
 &= \Phi\left(\frac{\tau_1 - \mu}{\sigma}\right) - \Phi\left(\frac{-\infty - \mu}{\sigma}\right) \\
 &= \Phi\left(\frac{\tau_1 - \mu}{\sigma}\right) - 0
 \end{aligned} \tag{2.5}$$

y

$$\begin{aligned}
 Pr(Y = K|\mu, \sigma, \tau_K) &= \Phi\left(\frac{\tau_K - \mu}{\sigma}\right) - \Phi\left(\frac{\tau_{K-1} - \mu}{\sigma}\right) \\
 &= \Phi\left(\frac{\infty - \mu}{\sigma}\right) - \Phi\left(\frac{\tau_{K-1} - \mu}{\sigma}\right) \\
 &= 1 - \Phi\left(\frac{\tau_{K-1} - \mu}{\sigma}\right)
 \end{aligned} \tag{2.6}$$

En la ecuación 2.5 en el término $\Phi\left(\frac{-\infty - \mu}{\sigma}\right)$ se indica la probabilidad acumulada en menos infinito (o muy cercano a este punto), la cual es 0; mientras que en la ecuación 2.6 el término $\Phi\left(\frac{\infty - \mu}{\sigma}\right)$ indica la probabilidad acumulada hasta infinito, la cual es 1. En los paneles de la figura 2.5 es más claro el significado de estas expresiones, ya que estas categorías extremas contienen la probabilidad de las colas de la distribución, por ejemplo, la probabilidad acumulada hasta τ_1 es la que está contenida en el intervalo $(-\infty, \tau_1]$.

La ecuación 2.4 sólo describe la probabilidad de que la respuesta Y pertenezca a una categoría k en específico, pero puede que se quiere expresar el cambio en la variable latente

como función de una combinación lineal de ciertas variables predictoras (una regresión lineal de la forma $\tilde{Y} = \eta + \epsilon$ con $\eta = \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n$); a nivel de la escala ordinal esto se refiere al cambio en la probabilidad esperada en cada una de las categorías. La probabilidad de $Y = k$ dados los predictores lineales es

$$Pr(Y = k|\eta) = \Phi(\tau_k - \eta) - \Phi(\tau_{k-1} - \eta) \tag{2.7}$$

Es importante notar que en $\tilde{Y} = \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \epsilon$ no se incluye un intercepto (b_0) en los predictores η , ya que los umbrales τ_k reemplazan dicho término.

La figura 2.6 presenta un ejemplo de una regresión ordinal con un predictor métrico.

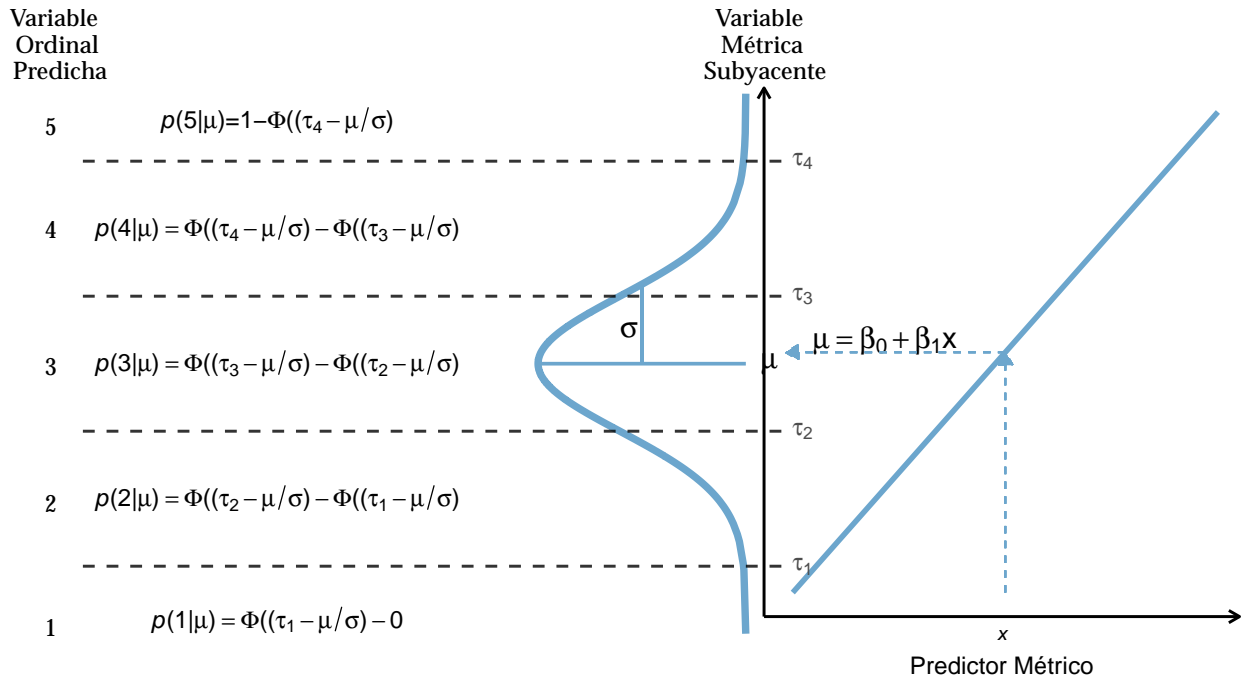


Figura 2.6: Ejemplo de regresión ordinal

Para entender cómo se modela el efecto de los predictores sobre la variable ordinal, es

importante entender por qué se resta el componente lineal η de cada intercepto τ_k . Si la probabilidad acumulada de una respuesta k en particular disminuye (por la resta de η), la masa de probabilidad incrementa hacia resultados más grandes (porque Φ integra a 1), de forma que un valor positivo de β indica que un incremento en x_i resulta en un incremento en la respuesta promedio, de esta forma se asegura el correcto ordenamiento de los valores resultantes, al modificar la verosimilitud de cada valor individual conforme el predictor x_i cambia (McElreath, 2016).

En resumen, si se comparan las figuras 2.1 y 2.3, la diferencia entre los comportamientos de los dos modelos queda clara (estas dos figuras fueron construidas a partir de los mismos datos simulados, que se presentan en el panel superior de la figura 2.3; las probabilidades en la figura 2.1 se calcularon a partir de 2.1, mientras que las probabilidades en 2.3 se calcularon a partir de 2.4). Por ejemplo, al fijarse en las probabilidades de los valores 4 y 5, el modelo métrico (figura 2.1) sugiere al 5 como el valor más probable de los dos (en realidad, el más probable de todos), mientras que el área bajo la curva entre el tercero y cuarto umbral (asociada al valor 4 en el panel inferior de la figura 2.3) es mucho más grande que el área entre el cuarto y el quinto (asociada al valor 5), por lo que en este caso 4 sería el valor más probable predicho por el modelo ordinal. Los datos en el panel superior de la figura 2.3 muestran que el valor más observado es 4. Como ya mencionó en la sección anterior, este no es el único problema que puede surgir al comparar ambos modelos (i.e., figura 2.2).

Esta tesis no sólo se centra en la argumentación a favor del uso de modelos ordinales sobre sus contrapartes métricas, sino que también presenta su implementación desde la perspectiva bayesiana, por lo que se aborda brevemente su funcionamiento así como sus ventajas sobre el

punto de vista clásico.

La perspectiva bayesiana

Bayesian inference yields intuitive conclusions. This reflects the fact that Bayesian inference is normative, stipulating how rational agents should change their opinion in the light of incoming data. Of course, it can nevertheless happen that you occasionally find a Bayesian conclusion to be surprising or counterintuitive. You are then left with one of two options—either the analysis was not carried out properly (e.g., errors in coding, errors in model specification) or your intuition is in need of schooling ¹² (Lee & Wagenmakers, 2014, pag. 12).

La aproximación bayesiana a la estadística es, fundamentalmente, sobre hacer uso de toda la información disponible al momento de hacer inferencias en la presencia de incertidumbre (Etz, 2018).

Una diferencia clave entre las perspectivas frecuentista y bayesiana para la inferencia estadística es la noción de probabilidad. Por un lado, la aproximación frecuentista entiende la probabilidad como una característica de los eventos y la sitúa en el mundo exterior (probabilidad objetiva), haciendo la tarea de la inferencia el descubrir esta probabilidad; por otro lado, la aproximación bayesiana la concibe como la cuantificación de incertidumbre o «grado de creencia» sobre algo desconocido, posicionándola como un estado subjetivo de conocimiento (probabilidad subjetiva), en este caso el propósito de la inferencia es el de reducir dicha incertidumbre (Dienes, 2008; Etz, 2018; Etz & Vandekerckhove, 2018; Lee &

¹²La inferencia bayesiana arroja conclusiones intuitivas. Esto refleja el hecho de que la inferencia bayesiana es normativa, estipulando cómo los agentes racionales deben cambiar su opinión a la luz de nuevos datos. Sin embargo, puede suceder que ocasionalmente encuentre que una conclusión bayesiana es sorprendente o contra-intuitiva. Entonces se queda con una de dos opciones: el análisis no se realizó correctamente (por ejemplo, errores en la codificación, errores en la especificación del modelo) o su intuición necesita revisión.

Wagenmakers, 2014). La base de la estadística bayesiana se puede resumir entonces como la actualización del conocimiento a priori a la luz de la evidencia; es decir, el conocimiento generado es una combinación de lo que se sabía antes de observar la evidencia y lo que se aprende una vez observado los datos (Etz & Vandekerckhove, 2018; Lee & Wagenmakers, 2014; Wagenmakers et al., 2018).

Para explicar de manera práctica el funcionamiento de la inferencia bayesiana imagine que está interesado en saber si una moneda es justa (i.e., si al ser lanzada al aire se pueden observar «Águilas» o «Soles» en la misma proporción), por lo que se propone lanzar la moneda 10 veces y contar el número de veces (k) que el resultado es «Águila», ya que supone que la probabilidad de que el resultado sea «Águila» depende de la «honradez» (representada por θ) de la moneda. Usted rápidamente se da cuenta de que este experimento es un ejemplo de un proceso binario o dicotómico ya que sólo existen dos posibles resultados, «Águila» o «Sol», y que la «honradez» de la moneda es representada por la tasa con la que esta cae «Águila» después de un lanzamiento. Dado que planea una serie de 10 lanzamientos determina que el experimento puede describirse por una distribución de probabilidad binomial, que modela el número de éxitos en una secuencia de ensayos binarios independientes entre sí, con θ la probabilidad fija de éxito (probabilidad de «Águila»), y n el número de ensayos (lanzamientos).

Antes de comenzar con dicho experimento, y como no sabe nada sobre que tan justas son las monedas (y por lo tanto ignora que tan probable es que el resultado sea «Águila»), considera que es prudente asumir que puede observar cualquier combinación de águilas y soles en los 10 lanzamientos; es decir, que la moneda puede estar ligera o fuertemente cargada hacia

cualquiera de los lados o de hecho ser justa. Igual que con el experimento, puede expresar su incertidumbre como una distribución de probabilidad, esta vez que refleje su estado de conocimiento (i.e., que asigne la misma probabilidad a todos los posibles valores de θ). Una distribución beta con parámetros $a = 1$ y $b = 1$ (equivalente a una distribución uniforme en el espacio $[0, 1]$; recuerde que θ es la «honradez» de la moneda, expresada como la probabilidad de observar «Águila» como resultado de un lanzamiento) cumple con este requisito.

En resumen, tanto la relación probabilística entre el número de «Águilas» observadas y la «honradez» de la moneda como su conocimiento *previo* sobre dicha «honradez» se pueden expresar como:

$$\theta \sim \text{Beta}(1, 1)$$

$$k \sim \text{Binomial}(\theta, n)$$

Una vez comenzado el experimento se detiene en el noveno lanzamiento habiendo observado la siguiente secuencia:

Sol, Águila, Sol, Águila, Sol, Sol, Águila, Águila, Sol

¿Qué puede decir sobre la «honradez» de su moneda? Utilizando el marco de inferencia Bayesiana usted puede determinar cuáles son los valores más probables de θ de la siguiente manera:

$$Pr(\theta|k) = \frac{Pr(k|\theta)Pr(\theta)}{Pr(k)} \quad (2.8)$$

Esta igualdad es conocida como Teorema o Regla de Bayes (Bayes, 1763), donde $Pr(\theta|k)$

(distribución posterior) especifica qué valores son los más probables para θ dado el conjunto de datos observados y representa su conocimiento actualizado sobre si la moneda es justa; $Pr(k|\theta)$ (función de verosimilitud) especifica qué tan probable es observar k éxitos bajo cada posible valor de θ ; $Pr(\theta)$ (distribución prior) especifica que tan probable es cada valor de θ antes de observar cualquier dato, esta distribución es la que refleja su grado de incertidumbre sobre la «honradez» de la moneda; y $Pr(k)$ (verosimilitud marginal) es una constante (una vez observados los datos) que asegura que el área bajo la distribución posterior iguale 1, y es resultado del Teorema de Probabilidad Total¹³ que dice que para encontrar la probabilidad sólo del evento B se suman todas las probabilidades conjuntas que involucran tanto a B como a un elemento de un conjunto independiente A (Etz & Vandekerckhove, 2018). En resumen, el Teorema de Bayes especifica cómo se combina la información obtenida por los datos ($Pr(k|\theta)$) con la información expresada en la distribución prior ($Pr(\theta)$).

¹³ $Pr(B) = Pr(A_1, B) + Pr(A_2, B) + \dots + Pr(A_J, B) = \sum_{j=1}^J Pr(A_j, B)$

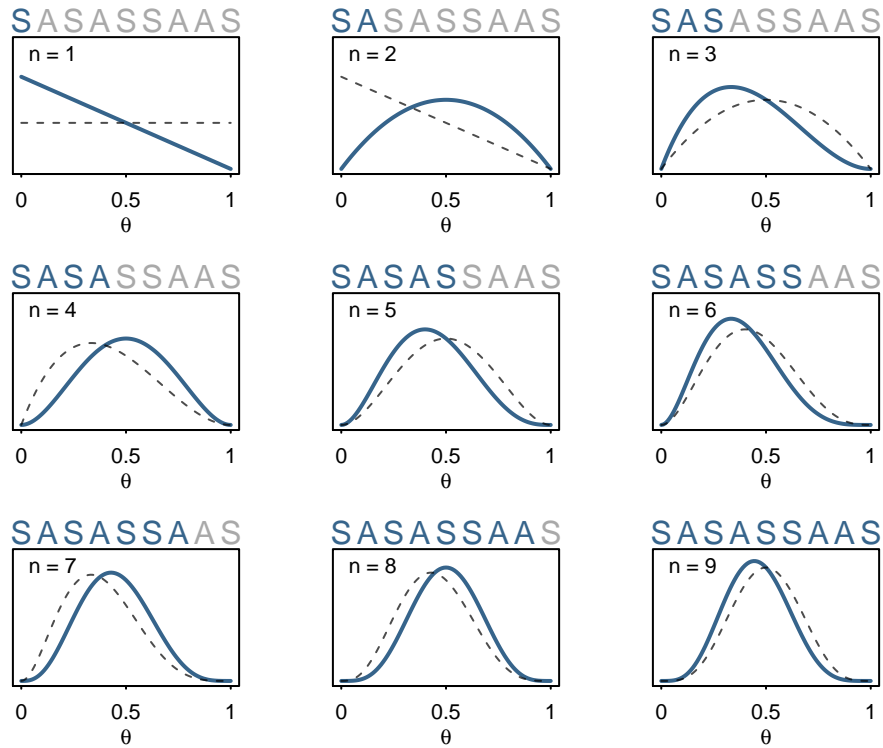


Figura 2.7: Ejemplo de actualización Bayesiana

La figura 2.7 muestra la distribución prior (línea punteada) en comparación a la distribución posterior (línea continua) para cada uno de los 9 lanzamientos que ha realizado (como si hubiera utilizado el Teorema de Bayes después de cada ensayo para actualizar su estado de conocimiento). La distribución posterior sobre θ refleja en cada paso (cada uno de los paneles) lo que sabe sobre la «honradez» de la moneda después de cada ensayo; es decir, tras cada nuevo lanzamiento se actualiza su conocimiento sobre cuáles son los valores más probables θ a partir de lo que sabía en el ensayo anterior, por ejemplo, la distribución prior (línea punteada) del segundo panel es la distribución posterior (línea continua) del primer panel y así sucesivamente (i.e., la distribución posterior de la n -ésima repetición es la distribución prior de la $n + 1$ -ésima). Note que la incertidumbre sobre la naturaleza de la

moneda se ve reducida por la nueva evidencia.

¿Qué esperaríamos observar en el décimo lanzamiento? De manera fácil e intuitiva podría utilizar la información de la distribución posterior sobre θ para responder esta pregunta. Si se fija en el último panel (inferior derecho) de la figura 2.7 podría concluir que el valor de θ es muy cercano a 0.5, por lo que esperaríamos que el resultado de su siguiente lanzamiento fuera «Águila» o «Sol» casi con la misma probabilidad.

En resumen, se puede pensar en el Teorema de Bayes como un puente entre los parámetros no observados del modelo y los datos observados, y que este puente se puede tomar de ida y de regreso (Lee & Wagenmakers, 2014):

- Distribución prior (sobre los parámetros), captura el conocimiento inicial sobre aquello que representa el o los parámetros de interés (p.e., qué tan «justa» es una moneda).
- Distribución prior predictiva (sobre los datos), que muestra qué datos esperar dado el modelo o estado actual de conocimiento (previo a cualquier observación).
- Distribución posterior (sobre los parámetros), captura lo que se sabe sobre el o los parámetros una vez actualizada la información inicial con la evidencia observada.
- Distribución posterior predictiva (sobre los datos), muestra qué datos esperar dado el modelo inicial pero en el estado actual de conocimiento (una vez observada la evidencia).

De estas cuatro distribuciones, la distribución posterior predictiva tiene un uso especialmente importante al momento de examinar qué tan descriptivamente adecuado es un modelo. Esta se puede ver como una serie de predicciones sobre qué datos espera ver el modelo basado en la distribución posterior sobre los parámetros, si estas predicciones no se parecen a los datos ya observados, entonces el modelo es descriptivamente inadecuado.

Como conclusión, Wagenmakers et. al. (2018) resumen las ventajas de la perspectiva bayesiana en los siguientes 10 puntos:

- Respecto a la estimación de parámetros
 1. La incorporación de conocimiento a priori
 2. La cuantificación de que θ esté en un intervalo específico
 3. Estimaciones condicionales a lo que se conoce (i.e., los datos)
 4. Coherencia (i.e., sin inconsistencias internas)
 5. El extenderse naturalmente a modelos con mayor complejidad
- Respecto a la prueba de hipótesis
 6. Cuantificar evidencia de los datos sobre la hipótesis nula (H_0) respecto a una hipótesis alternativa (H_1)
 7. Cuantificar evidencia a favor de H_0
 8. Permitir que el monitoreo de evidencia conforme se acumulen datos
 9. No depender de esquemas de muestreo desconocidos o ausentes
 10. No estar «fuertemente sesgada» contra H_0

Además, la aproximación bayesiana permite generalizar fácilmente los modelos tradicionales de forma que sean robustos a valores atípicos, permita varianzas diferente dentro de niveles o una estructura jerárquica (compartir información entre niveles o factores)

Por ejemplo, un problema que poco se ha mencionado en el debate del trato de datos ordinales son los errores que surgen de asumir, injustificadamente, igualdad de varianzas (supuesto que subyace al análisis de regresión y ANOVA's). Glass, Peckham y Sanders (1972) hablan de la robustez de la prueba F a violaciones del supuesto de normalidad, argumento

común en la defensa de los modelos métricos en la discusión presentada (Carifio & Perla, 2007, 2008; Norman, 2010), sin embargo el mismo Glass menciona que el resultado no es el mismo al hablar del supuesto de igualdad de varianzas. La perspectiva bayesiana permite modelar naturalmente para grupos con varianzas desiguales.

Selección de Priors

Las distribuciones prior, así como su elección, son un componente clave de la inferencia bayesiana (Gelman, 2002; Gelman, Simpson, & Betancourt, 2017), a pesar de que el análisis de los resultados se enfoca principalmente en la distribución posterior y posterior predictiva. Como ya se mencionó, esta distribución representa la información que se tiene sobre lo que representa el o los parámetros desconocido (e.g., θ en el ejemplo de la moneda). Los puntos clave en la construcción de una distribución prior son (Gelman, 2002):

1. Qué información se considera para esta.
2. Las propiedades de la distribución posterior resultante.

Estrictamente hablando las priors pueden definirse no solo antes de observar los datos, sino incluso antes de que se haya considerado el experimento; Gelman, Simpson y Betancourt (2017) argumentan a favor de la elección de prior en el contexto de todo el análisis bayesiano, desde la inferencia hasta la predicción y la evaluación del modelo.

Retomando la ecuación 2.8 podemos observar (en el denominador de la parte derecha de la igualdad, $Pr(k|\theta)Pr(\theta)$) que un aspecto importante de la distribución prior, más allá de ser una distribución de probabilidad, es que se combina con una función de verosimilitud (Gelman et al., 2017). Es decir, el formalismo bayesiano requiere que una distribución prior

sea actualizada en una distribución posterior basada en nueva evidencia (como se muestra en los diferentes paneles de la figura 2.7). De forma que la utilidad práctica de una distribución prior depende críticamente tanto de cómo interactúa con el modelo de probabilidad asumido para los datos (la función de verosimilitud) como del contexto de los datos que se observan (Gelman et al., 2017).

Por ejemplo, en el caso de la moneda ilustrado anteriormente (una binomial con $n = 9$ ensayos y una prior uniforme sobre θ), si se observa que el resultado del volado es «Águila» en 4 de los ensayos (como es el caso en la 2.7), entonces puede calcularse fácilmente la posterior y considerarse cuestiones de sensibilidad y desempeño predicativo independientemente de la elección de prior; pero, ¿qué pasa si se observa «Águila» en 9 de los ensayos? Entonces la elección de la distribución prior se vuelve más relevante, especialmente si está trabajando en un entorno en el que las distinciones entre probabilidades de 0.9, 0.99 y 0.999 son de importancia aplicada (en estos casos la elección de una prior uniforme sobre todo el espacio $[0, 1]$ no resulta muy adecuado). Ahora, como Gelman, Simpson y Betancourt (2017) mencionan, esto no implica que la distribución prior deba depender explícitamente de los datos, solo que una prior determinada puede que funciona bien en un escenario pero ser problemática en otro.

En resumen, el argumento de Gelman, Simpson y Betancourt (2017) es que se debe pensar cuidadosamente sobre cómo especificar las priors en el contexto del modelo probabilístico de los datos (la verosimilitud); en particular, pensar generativamente, considerando las mediciones potenciales consistentes con una prior, y predictivamente, al validar esas potenciales mediciones con datos recolectados. Como se mencionó en la sección anterior, una herramienta fundamental

para entender el efecto de las priors en la inferencia antes de que se hayan recopilado los datos es la distribución prior predictiva (de forma que una prior es generativa si la distribución prior predictiva genera solo datos considerados consistentes con la comprensión del problema), mientras que una herramienta fundamental para validar el modelo después de que se hayan recopilado los datos es la distribución posterior predictiva (una prior tiene un buen desempeño predictivo si la distribución predictiva posterior es consistente con el verdadero proceso de generación de datos y puede predecir nuevos datos generados a partir de experimentos similares).

La elección de las distribuciones prior no debe tomarse a la ligera, especialmente al tratar con problemas más complejos y de mayor dimensionalidad que los aquí contemplados, por lo que se recomienda encarecidamente probar cuán apropiado es una prior para un problema particular a través de las predicciones y propiedades de la distribución posterior predictiva (Gelman et al., 2017), o comprobar la influencia de las distribuciones prior sobre las inferencias a través de un análisis de sensibilidad, comparando las inferencias posteriores bajo diferentes opciones razonables de distribución prior (Gelman, 2002); sin embargo, un trato más específico del tema (i.e., que tipo de priors predeterminadas son más adecuadas, o qué procedimiento debería seguirse para construir un distribución prior «a la medida» del problema en mano) excede los objetivos de la presente tesis, por lo que se utilizarán las distribuciones predeterminadas, explícitamente para ser poco informativas e influir lo menos posible en los resultados, que utiliza el software empleado para el ajuste de los modelos bayesianos, *brms* (Bürkner, 2017).

Por defecto, los parámetros a nivel de población tienen priors planas impropias sobre los

reales (parecida a la distribución uniforme empleada en el ejemplo de la moneda); es decir, la función empleada tiene una integral infinita y no es, estrictamente hablando, una función de probabilidad (i.e, no integra a 1). Sin embargo, cuando se combina formalmente con la verosimilitud de los datos produce una distribución posterior propia coherentes (Akaike, 1980; Gelman, 2002)). Los parámetros a nivel grupo (u) se asume que vienen de una distribución normal multivariada con media cero y matriz de covarianzas desconocida Σ ($u \sim N(\mathbf{0}, \Sigma)$). Para algunas familias es necesario estimar parámetros adicionales, por ejemplo para la distribución normal el parámetro σ es necesario para dar cuenta de la varianza del error residual, este tiene una distribución media-Cauchy¹⁴ con un parámetro de escala que depende de la desviación estándar de la variable respuesta para permanecer solo débilmente informativa, independientemente de la escala de esta variable (Bürkner, 2017).

¹⁴Literalmente se refiere a la mitad de una distribución Cauchy (esta distribución es simétrica) y tiene la forma: $f(x) = \left(\frac{1}{s}\right) \frac{2}{\pi\left(1+\left(\frac{x-x_0}{s}\right)^2\right)}$ con x_0 el parámetro de localización y s el parámetro de escala, en este caso $x_0 = 0$ y s depende de la desviación estándar de Y

Capítulo 3

Método

Bases de datos

Para la elección de las bases de datos que se utilizarían se realizó una búsqueda en línea con los siguientes criterios: (1) que el estudio haya sido realizado dentro del campo de los juicios morales, (2) que la variable respuesta estuviera medida en escala ordinal, y (3) que la base de datos recolectada fuera de libre acceso.

A partir de estos criterios se seleccionaron dos bases de datos:

Base 1: Even arbitrary norms influence moral decision-making. Pryor, Perfors y Howe (2019)

El primer conjunto de datos fue tomado de los experimentos presentados en Pryor, Perfors y Howe (2019), disponibles a través del portal de Open Science Framework. En ellos se busca mostrar si los individuos siguen una norma social (i.e., copiar comportamientos

comunes entre otras personas) incluso al estar conscientes de que estas son arbitrarias y no reflejan las preferencias reales de otros. Para ello presentan 4 experimentos (de un sólo ensayo cada uno), en los que se presentaba a los participantes con un escenario que planteaba un dilema moral y en donde tenían que elegir entre diferentes cursos de acción (e.g., dentro de un rango entre no hacer nada y realizar una acción concreta, como llamar a la policía).

Las manipulaciones se hicieron sobre las normas transmitidas a los participantes, informándolos de los resultados de un supuesto estudio previo; el aspecto crucial era que a los participantes se les decía que en este anterior estudio no se les pidió elegir cómo actuarían en el dilema moral, sino que se les asignó al azar para imaginar un escenario en que realizaban una acción específica para luego evaluar qué tan bien o mal se sentirían al realizar dicha acción (es decir, las elecciones no reflejaban la conducta real de los individuos). El dato que se transmitía a los participantes era la proporción de sujetos (con rasgos como género, edad y rasgos de personalidad similares a los suyos) asignados aleatoriamente a alguna de las acciones en los extremos de la escala, por ejemplo a la mitad se les decía que esta acción había sido «Definitivamente llamar a la policía e informe del ladrón», y a la otra mitad «Definitivamente no llamar a la policía y dejar ir al ladrón». Las diferentes preparaciones fueron:

- Experimento 1a ($n = 150$): Se les decía que el 75 % de los participantes anteriores fueron asignados aleatoriamente a una de las elecciones en los extremos ($OR = 2.16$; $P = 0.009$; intervalo de confianza al 95 %: 1.22–3.85).
- Experimento 1b ($n = 151$): Exactamente igual que 1a, pero con un dilema moral distinto ($OR = 1.98$; $P = 0.019$; intervalo de confianza al 95 %: 1.12 to 3.53).
- Experimento 2 ($n = 150$): Se les decía que el 25 % de los participantes anteriores fueron

asignados aleatoriamente a una de las elecciones en los extremos, con el mismo dilema que en 1a (OR = 2.08; P = 0.0127; intervalo de confianza al 95 %: 1.17–3.72).

- Experimento 3 (n = 150): Se les decía que el 75 % de los participantes anteriores fueron asignados aleatoriamente a una de las elecciones en los extremos, pero se le había presentado un dilema diferente (OR = 0.83; P = 0.512; intervalo de confianza al 95 %: 0.47–1.46; no significativo).
- Experimento 4 (n = 180): Igual que 1a, pero se variaban los rasgos de los participantes anteriores, por ejemplo que sujetos similares favorecen la denuncia del ladrón y sujetos distintos favorece no informar del ladrón, o viceversa (OR = 0.90; P = 0.811; intervalo de confianza 95 %: 0.38–2.11; no significativo).

Los resultados de los análisis mostrados entre paréntesis son de una regresión logística, que no es equivalente al análisis que aquí nos interesa, por lo que las comparaciones con estos resultados se harán en términos cualitativos (i.e., si las inferencias favorecen las mismas conclusiones).

Adicionalmente, todos los participantes contestaron un cuestionario de personalidad compuesto por 10 reactivos (variables también ordinales) que también serán de interés para el presente trabajo. La información de estos cuestionarios sólo se empleó para modificar las instrucciones de los participantes (los rasgos que se describían de los participantes en el supuesto estudio anterior).

Base 2: Doing Bayesian Data Analysis, Kruschke (2015)

Como complemento, y para mostrar de forma sencilla y gráficamente las ventajas de los modelos ordinales al agregar predictores lineales continuos, se utilizó la base de datos

`OrdinalProbitData-LinReg-2.csv` que forma parte de los recursos proporcionados por Kruschke (2015) del libro «Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan».

Para esta segunda base también se muestran las estimaciones de los modelos en su versión clásica o frecuentista; es decir, la estimaciones obtenidas a través del método de máxima verosimilitud.

Herramientas

Para el ajuste y análisis estadístico de todos los modelos se hizo uso del software R (R Core Team, 2017). Para el ajuste de todos los modelos bayesianos se emplea el paquete `brms` (Bürkner, 2017), que implementa modelos bayesianos multinivel en R usando el lenguaje de programación Stan (Carpenter et al., 2017).

Para ajustar los modelos frecuentistas (sólo para la segunda base de datos) se empleó la función interna `lm()`, para la estimación del modelo métrico, y la función `polr()` del paquete `MASS` (Venables & Ripley, 2002), para la estimación del modelo ordinal.

Modelos ajustados

Modelo más simple

Para esta primera aproximación se puede ajustar un modelo lineal sin ningún predictor; es decir, haciendo la estimación sólo del parámetro β_0 o intercepto (equivalente a μ), varianza (σ), y los umbrales (τ_K) en el caso el modelo ordinal. Esto permitirá hacer una primera

comparación entre los modelos (métrico vs ordinal) a partir de las descripciones que hacen sobre los datos y las predicciones que pueden generar de los mismos.

Estimación métrica bayesiana

La función `brm()` del paquete `brms` permite ajustar modelos bayesiano de forma sencilla y su sintaxis es fácil de comprender. Esta función toma como principales argumentos:

- `formula`, la formula del modelo lineal.
- `data`, el conjunto de datos que se utilizarán para el ajuste.
- `family`, distribución de la respuesta.

estos no son los únicos argumentos que la función `brm()` soporta (e.g., se pueden modificar las distribuciones prior sobre los parámetros, el número de cadenas o de iteraciones), pero estos son los mínimos necesarios para hacer uso de la función.

```
brm(  
  formula = as.numeric(Y) ~ 1,  
  data = data,  
  family = "gaussian"  
)
```

Como se observa en este primer ejemplo, la fórmula se escribe de la forma $Y \sim X$, donde la variable dependiente o respuesta se escribe del lado izquierdo de la virgulilla (\sim), y la o las variables independientes o explicativas del lado derecho; dado que en esta primera aproximación se desea ajustar un modelo sin predictores lineales se incluye el 1 en la formula, haciendo explícito este punto. Todas las funciones que se utilizarán a partir de este punto

tienen como uno de sus argumentos `formula`, y siempre se utiliza de la misma forma.

La función `as.numeric()` le indica a la función que trate las categorías en la variable Y como números igualmente separados entre así (i.e. $2 - 1 = 5 - 4$), esta es precisamente la suposición que se hace al emplear modelos métricos sobre variables ordinales.

El argumento `family` cuenta con un valor por defecto ("`gaussian`", de hecho), pero se escribe para que se haga explícito que se asume que la distribución de la respuesta (o equivalente, de los errores) es normal, otra suposición implícita al ajustar modelos lineales sobre variables ordinales.

Estimación ordinal bayesiana

Para ajustar un modelo ordinal a través de la misma función `brm()` el cambio más importante que se debe realizar es sobre el argumento `formula` ya que no se incluye el `as.numeric`, lo que indica que la variable Y deberá ser tratada como ordinal (que es su naturaleza). Un segundo cambio se ejecuta sobre argumento `family` como a continuación se muestra:

```
brm(  
  formula = Y ~ 1,  
  data = data,  
  family = cumulative("probit", threshold = "flexible")  
)
```

Al declarar `family = cumulative("probit", threshold = "flexible")` se están indicando tres aspectos importantes. Primero, se pide un modelo ordinal de liga acumulativa

(`cumulative`); segundo, se establece que la variable latente se distribuye normal (`probit`); y tercero, se permite que los umbrales no sean equidistantes (`threshold = "flexible"`, e.g., los paneles inferiores en la figura 2.5). Como ya se mencionó, este proyecto se centra en modelos ordinales acumulativos de liga probit, por lo que todos los modelos ordinales utilizados compartirán esta característica.

Modelos Lineales

Como una segunda aproximación de análisis, se pueden añadir variables ayuden a explicar las variaciones en la variable respuesta de interés, para ello se añaden predictores lineales a los argumentos de las funciones descritas.

Para estos modelos se presentan las versiones frecuentista y bayesiana (Tanto de los modelos métricos como de los ordinales). Esto añade la comparación entre aproximaciones (frecuentista vs bayesiana) a partir de la interpretación que se puede hacer de las estimaciones y predicciones realizadas (esto para la segunda base de datos).

Modelos frecuentistas

Regresión Lineal (métrica)

La función interna `lm()` ajusta modelos lineales con una sintaxis similar a la descrita para la función `brm()`; su uso sólo requiere de los argumentos `formula` y `data`.

```
lm(  
  formula = as.numeric(Y) ~ 1 + X,  
  data = data
```

```
)
```

Si se desean añadir predictores se utiliza el símbolo `+` para separar las variables explicativas (e.g., $Y \sim X + Z$). En todos los modelos lineales aquí presentados sólo se utiliza un predictor lineal por simplicidad.

Regresión Ordinal Acumulativa

Para ajustar un modelo ordinal frecuentista se utiliza la función `polr()` del paquete `MASS`; los argumentos son los mismos que en la función anterior, añadiendo `method`, que establece el tipo de modelo a estimar y por consiguiente la función liga que se usa (similar al argumento `family` de la función `brm`).

```
polr(  
  formula = Y ~ 1 + X,  
  data = data,  
  method = "probit"  
)
```

Modelos Bayesianos

El ajuste de estos modelos es similar al descrito en la sección 3.3.1, solamente se agrega el predictor X a la fórmula de los modelos.

Regresión Lineal (métrica)


```
brm(  
  formula = as.numeric(Y) ~ 1 + X,  
  data = data,  
  family = "gaussian"  
)
```

Regresión Ordinal Acumulativa

```
brm(  
  formula = Y ~ 1 + X,  
  data = data,  
  family = cumulative("probit", threshold = "flexible")  
)
```

Capítulo 4

Resultados

Exploratorios

Como parte de la descripción de las bases de datos se pueden explorar tanto la frecuencia como la proporción que representa cada nivel de respuesta en la variable ordinal y que serán el foco de las siguientes secciones:

Base 1: Pryor (2019)

Para esta primera base se presenta el resumen de cada uno de los 5 experimentos realizados. Para los experimentos 1A, 2, 3 y 4 las categorías de respuesta (sobre un dilema en el que un hombre roba un banco para después dar todo el dinero a un orfanato) eran:

- (1) Definitivamente llamar a la policía y reportar al ladrón
- (2) Muy probablemente llamar a la policía y reportar al ladrón
- (3) Probablemente llamar a la policía y reportar al ladrón

- (4) Probablemente no hacer nada y dejar ir al ladrón
- (5) Muy probablemente no hacer nada y dejar ir al ladrón
- (6) Definitivamente no hacer nada y dejar ir al ladrón

Tabla 4.1: Exploración: Experimentos Pryor, 2019 [1]

Respuesta	Exp 1A		Exp 1B		Exp 2	
	Frecuencia	Proporción	Frecuencia	Proporción	Frecuencia	Proporción
1	30	0.20	22	0.15	28	0.19
2	20	0.13	30	0.20	28	0.19
3	25	0.17	37	0.25	24	0.16
4	27	0.18	32	0.21	31	0.21
5	21	0.14	19	0.13	21	0.14
6	27	0.18	11	0.07	18	0.12

Tabla 4.2: Exploración: Experimentos Pryor, 2019 [2]

Respuesta	Exp 3		Exp 4	
	Frecuencia	Proporción	Frecuencia	Proporción
1	30	0.20	46	0.26
2	18	0.12	25	0.14
3	17	0.11	31	0.17
4	40	0.27	28	0.16
5	22	0.15	32	0.18
6	23	0.15	18	0.10

Para el experimento 1B las categorías representan (sobre un dilema entre contratar a un amigo o a una persona más calificada) las respuestas eran:

- (1) Definitivamente contratar a la persona más calificada

- (2) Muy probablemente contratar a la persona más calificada
- (3) Probablemente contratar a la persona más calificada
- (4) Probablemente contratar a su amigo
- (5) Muy probablemente contratar a su amigo
- (6) Definitivamente contratar a su amigo

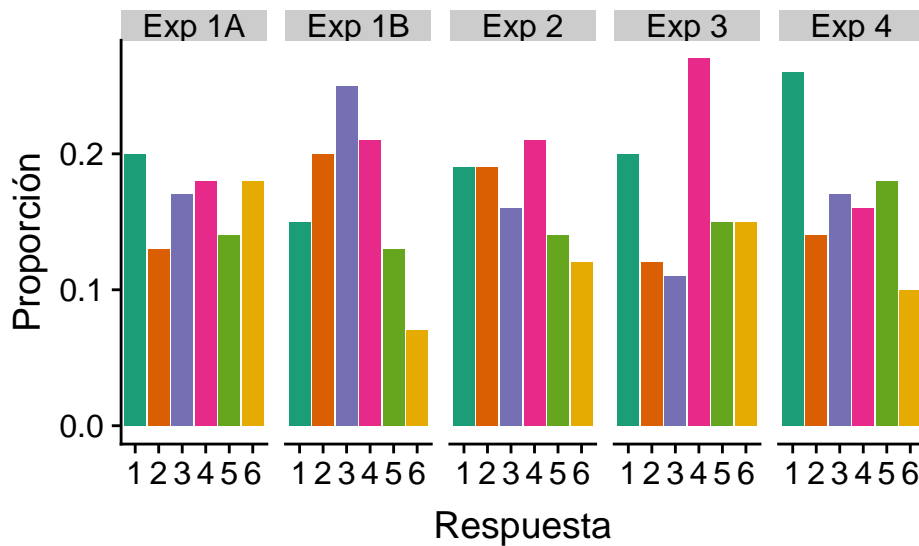


Figura 4.1: Exploración: Pryor, 2009

En las tablas 4.1 y 4.2, así como en la figura 4.1, se pueden observar tres tipos de patrones interesantes en las respuestas: (1) se distribuyen casi uniformemente entre todas las categorías (Exp 1A y Exp 2), (2) se concentran en el centro (Exp 1B), y (3) se concentran en alguno de los extremos (Exp 3 y Exp 4).

El cuestionario de personalidad empleado cuenta con de 10 ítems que miden rasgos de personalidad en las dimensiones del «Big-Five» (Extraversión, Apertura a la experiencia, Responsabilidad, Afabilidad, Estabilidad Emocional) presentado en (Gosling, Rentfrow, & Jr, 2003). Las categorías de respuesta para cada ítem son:

- (1) Fuertemente en desacuerdo
- (2) Moderadamente en desacuerdo
- (3) Un poco en desacuerdo
- (4) Ni de acuerdo ni en desacuerdo
- (5) Un poco de acuerdo
- (6) Moderadamente de acuerdo
- (7) Fuertemente de acuerdo

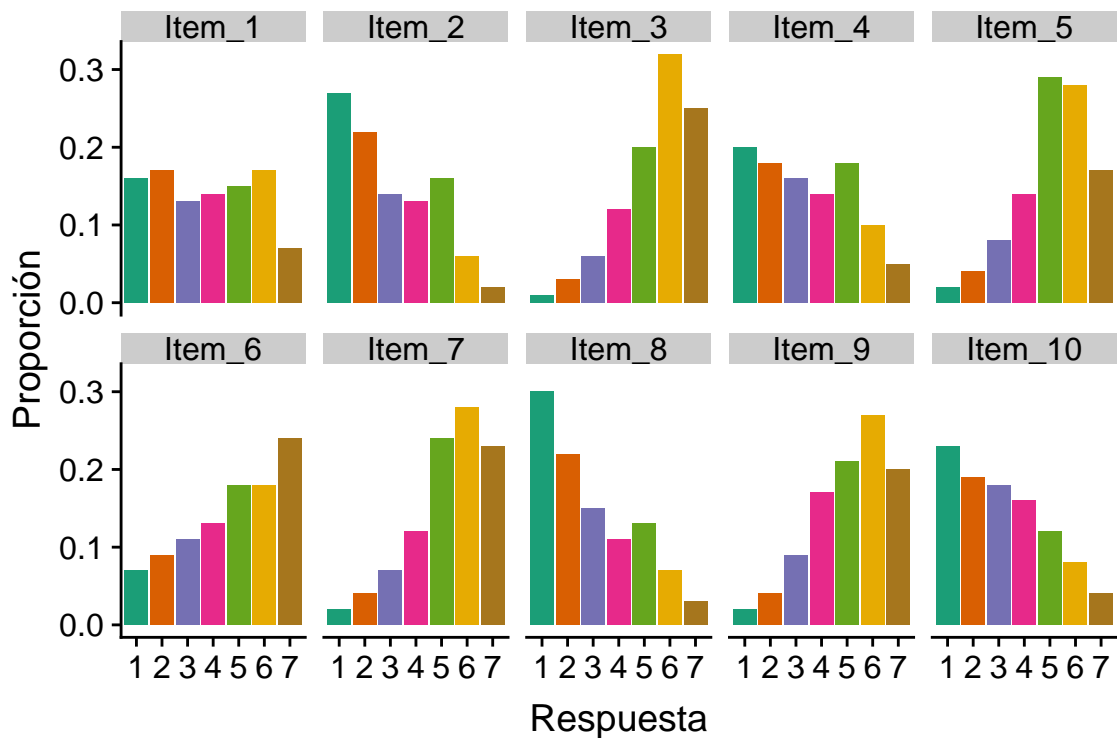


Figura 4.2: Exploración: Respuestas Personalidad

En la figura 4.2 también se observan los tres patrones de respuestas que serán importantes para las comparaciones posteriores: (1) distribución uniforme (Item 1 y 4), (2) distribución centrada (Item 5 y 7), y distribución sesgada (Item 2, 3, 6, 8, 9 y 10).

Base 2: Kruschke (2015)

Esta base cuenta con dos columnas, X variable continua y Y variable de tipo ordinal con categorías que van del 1 al 7 ($K + 1 = 7$), ambas simuladas con parámetros desconocidos, cada una con 200 observaciones. La siguiente tabla muestra un resumen de la base:

Tabla 4.3: Exploración: Kruschke, 2015

Respuesta	Frecuencia	Proporción
1	47	0.24
2	18	0.09
3	22	0.11
4	26	0.13
5	22	0.11
6	21	0.10
7	44	0.22

Una rápida inspección muestra que la mayoría de respuestas se concentran en la primera y última categoría. Se puede corroborar esta primera evaluación graficando las proporciones obtenidas.

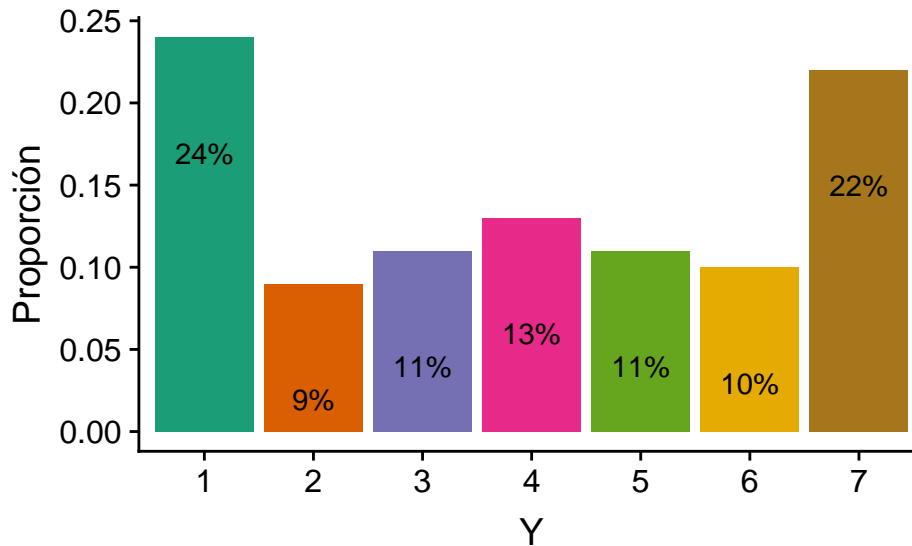


Figura 4.3: Exploración: Kruschke, 2015

La figura 4.3 muestra un patrón de respuesta distinto a los descritos antes, ya que se observa que las respuestas se concentran en las categorías de respuesta de los extremos.

Evaluación de Modelos

Modelo más simple

Cuestionario Personalidad (Pryor, 2019)

Para presentar las diferencias en las inferencias entre los análisis métrico y ordinal se presentan las distribuciones posteriores predictivas (i.e., qué predice el modelo con base en la distribución posterior obtenida al evaluar el modelo) de los modelos bayesianos empleados. Para los modelos métricos estas distribución es toman la forma de curvas normales, para los modelos ordinales toman la forma de puntos. Recuerde que en la sección 2.6 se mencionó que

estas distribuciones se utilizan para evaluar el ajuste de los modelos bayesianos.

Las figuras 4.4 y 4.5 muestra las distribuciones posteriores predictivas para cada una de las preguntas de personalidad en Pryor (2019); a simple vista se puede observar que los modelos ordinales (4.5) captura con mayor precisión el patrón observado en los datos como se muestra por los puntos superpuestos en los histogramas de datos (esto es especialmente evidente para los Items 1, 2 y 8), las barras muestran el error de estimación. Por el contrario, los modelos métricos (4.4) describen mal este aspecto de los datos y se concentra en el centro, a diferencia de los datos observados.

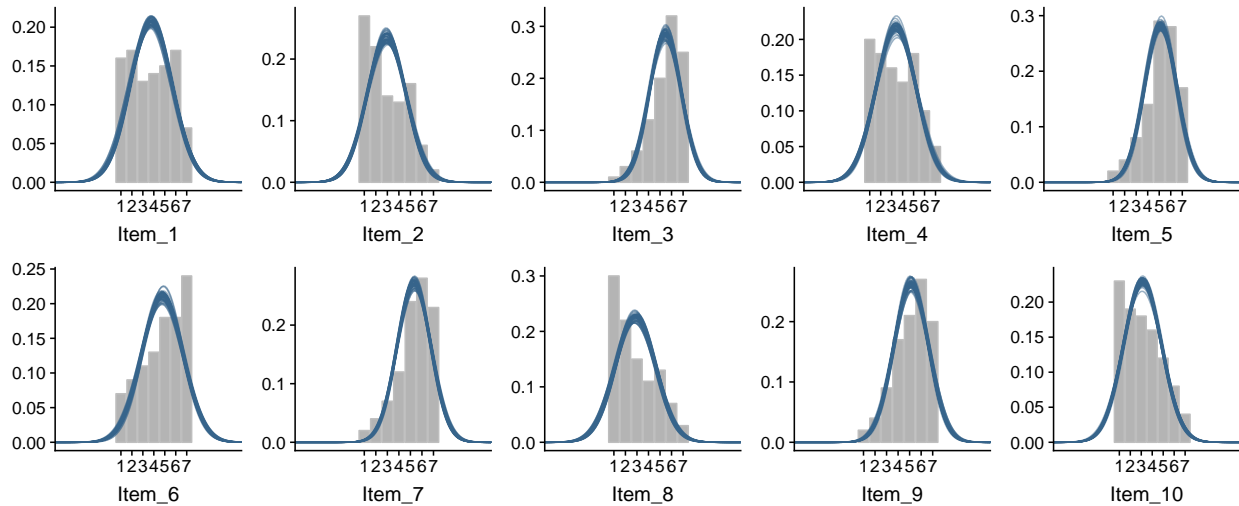


Figura 4.4: Respuestas Cuestionario (Análisis Métrico)

Además, el análisis métrico presenta un problema importante. Se puede apreciar (e.g., en los Items 2, 3, 6, 8 y 10 es particularmente notorio) que las colas de las distribuciones posteriores predictivas se extienden sobre valores que se encuentran fuera del rango de las respuestas; es decir, el modelo métrico predice que se observarían valores fuera del intervalo $[1, 7]$ (e.g., en el ítem 3 se espera observar más respuestas en valores mayores a 7 que respuestas

en 1 o 2). Este problema no se observa en las predicciones del modelo ordinal.

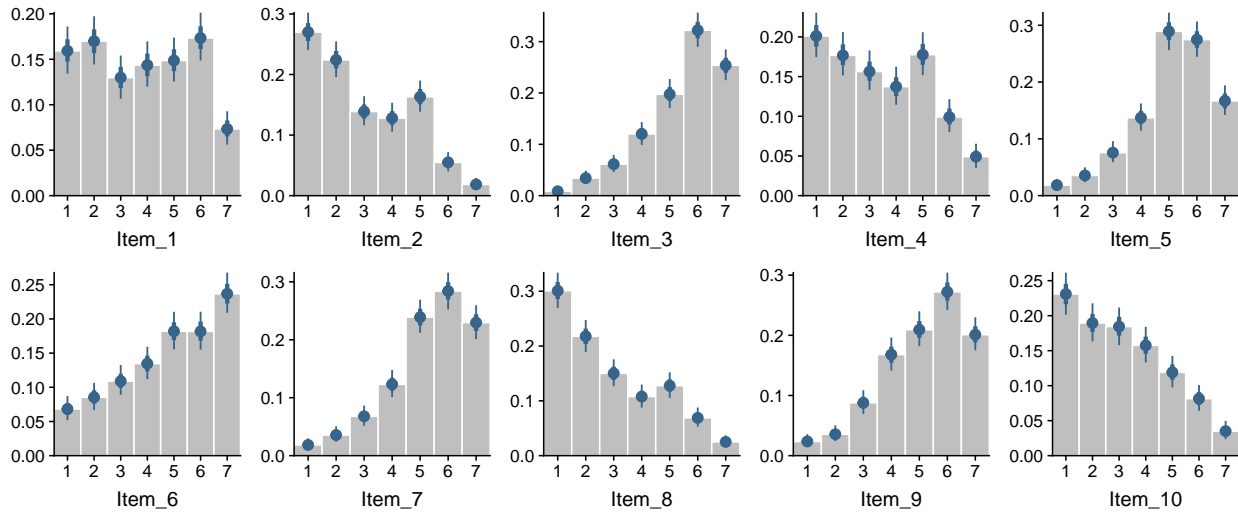


Figura 4.5: Respuestas Cuestionario (Análisis Ordinal)

En la figura 4.5 los puntos representan la mediana de las distribución predictiva posterior y las líneas muestran el intervalo de máxima densidad (HDI por sus siglas en inglés). Este resume la distribución posterior de un parámetro especificando un intervalo que cubre la mayor parte de esta (i.e. que abarca la mayoría del área bajo la curva), en este ejemplo el 95 %, de manera que cada punto dentro del intervalo tiene mayor credibilidad que cualquier otro punto fuera del mismo.

Modelos lineales

Experimentos (Pryor, 2019)

Los modelos lineales simples ajustados para cada experimento tenían como variable dependiente la respuesta de los participantes al dilema moral que se les presentaba (Respuesta en las siguientes tablas y figuras), y como variable independiente la norma arbitraria que se

les presentaba (Condición en las siguientes tablas y figuras). Recuérdese que estas normas estaban definidas como proporción de participantes, en el hipotético estudio anterior, asignados a uno de los extremos del rango de respuestas (e.g., A: «Definitivamente llamar a la policía y reportar al ladrón» y B: «Definitivamente no hacer nada y dejar ir al ladrón»). El resumen de las estimaciones métricas se presentan en la tabla 4.4.

Tabla 4.4: Estimaciones Posteriores Métricas: Pryor, 2019

Experimento	Parámetro	Media	Desviación	Límite Inferior	Límite Superior
Exp 1A	Condición	-0.7461	0.2840	-1.2104	-0.2865
	b0	3.8368	0.1999	3.5141	4.1574
	sigma	1.7406	0.1022	1.5841	1.9166
Exp 1B	Condición	-0.5483	0.2372	-0.9314	-0.1555
	b0	3.4706	0.1693	3.1982	3.7507
	sigma	1.4480	0.0862	1.3160	1.5985
Exp 2	Condición	-0.6480	0.2666	-1.0781	-0.1997
	b0	3.6095	0.1881	3.2933	3.9160
	sigma	1.6322	0.0948	1.4854	1.7914
Exp 3	Condición	0.1913	0.2844	-0.2773	0.6617
	b0	3.4109	0.2021	3.0845	3.7427
	sigma	1.7259	0.1013	1.5654	1.8984
Exp 4	Condición	-0.4123	0.2581	-0.8459	0.0042
	b0	3.3628	0.1833	3.0606	3.6629
	sigma	1.7136	0.0895	1.5741	1.8680

Dado que la variable explicativa es dicotómica el valor del parámetro se interpreta como el cambio en la respuesta por un cambio de condición, tomando como referencia la condición A (i.e., la norma arbitraria era la respuesta (1)). En la tabla 4.4 se presentan los estimados del coeficiente de la variable explicativa (Condición), intercepto (b0) y varianza de los residuales (sigma). Se observa que estos resultados son consistentes con los reportados en el artículo

original, encontrando un efecto en los experimentos 1A, 1B y 2 (los intervalos de máxima densidad no contienen el 0), y no encontrando efecto en los experimentos 3 y 4 (los intervalos de máxima densidad contienen el 0).

Las siguientes tres tablas (4.5, 4.6 y 4.7) resumen las estimaciones hechas por el modelo ordinal. La primer diferencia que se nota respecto a la tabla 4.4 es que hay cinco parámetros más, uno para cada uno de los umbrales. Nótese que, como se mencionó en la sección 2.5.1 los umbrales extremos están fijo (la desviación de la estimación es 0).

Tabla 4.5: Estimaciones Posteriores Ordinales: Pryor, 2019 [1]

Parámetro	Exp 1A				Exp 1B			
	Media	Desviación	L. Inf.	L. Sup.	Media	Desviación	L. Inf.	L. Sup.
Condición	-1.010279	0.3692686	-1.732349	-0.2878392	-0.6068863	0.2590361	-1.144476	-0.1107494
b0	3.913781	0.2785425	3.383347	4.4697242	3.4848048	0.2071166	3.080244	3.8889897
sigma	2.174095	0.1800045	1.845436	2.5401938	1.5341138	0.1163402	1.331848	1.7791228
Umbral[1]	1.500000	0.0000000	1.500000	1.5000000	1.5000000	0.0000000	1.500000	1.5000000
Umbral[2]	2.436132	0.1708457	2.116285	2.7726003	2.5373121	0.1520988	2.258413	2.8393443
Umbral[3]	3.414712	0.1898249	3.071921	3.8038361	3.5386038	0.1728897	3.202120	3.8692748
Umbral[4]	4.488087	0.1760578	4.120996	4.8085339	4.5323762	0.1694291	4.200573	4.8546612
Umbral[5]	5.500000	0.0000000	5.500000	5.5000000	5.5000000	0.0000000	5.500000	5.5000000

Una primera inspección de los resultados podría sugerir que no hay una diferencia importante entre las estimaciones de los parámetros comunes a los dos modelos (Condición, b0 y sigma), las estimaciones van en la misma dirección y las medias de las estimaciones son muy cercanas (la diferencia más grande es la estimación de Condición para el primer experimento, -0.74 para el modelo métrico y -1.01 para el modelo ordinal). Incluso se podría argumentar que las desviaciones de las estimaciones métricas son ligeramente menores que en

las estimaciones ordinales.

Tabla 4.6: Estimaciones Posteriores Ordinales: Pryor, 2019 [2]

Parámetro	Exp 2				Exp 3			
	Media	Desviación	L. Inf.	L. Sup.	Media	Desviación	L. Inf.	L. Sup.
Condición	-0.7246074	0.3208457	-1.326597	-0.0621268	0.1931869	0.3684537	-0.5080777	0.9455928
b0	3.5898111	0.2445144	3.126057	4.0750501	3.2204969	0.2696536	2.7187805	3.7724670
sigma	1.8628889	0.1420470	1.598499	2.1567524	2.0964680	0.1681177	1.8113812	2.4661488
Umbral[1]	1.5000000	0.0000000	1.500000	1.5000000	1.5000000	0.0000000	1.5000000	1.5000000
Umbral[2]	2.5764307	0.1613840	2.262968	2.8882024	2.3087594	0.1576172	2.0168650	2.6300512
Umbral[3]	3.3764640	0.1804497	3.025809	3.7448785	2.9662328	0.1826383	2.6184627	3.3306359
Umbral[4]	4.4793437	0.1732940	4.129245	4.8096180	4.4387492	0.1752068	4.0863377	4.7632272
Umbral[5]	5.5000000	0.0000000	5.500000	5.5000000	5.5000000	0.0000000	5.5000000	5.5000000

Además, se encuentra un relación entre la variable respuesta y la variable explicativa en los experimentos 1A, 1B y 2, y no se detecta un efecto en los experimentos 3 y 4.

Tabla 4.7: Estimaciones Posteriores Ordinales: Pryor, 2019 [3]

Parámetro	Exp 4			
	Media	Desviación	L. Inf.	L. Sup.
Condición	-0.4644021	0.3145152	-1.109325	0.1247252
b0	3.0807027	0.2351748	2.644897	3.5571741
sigma	2.0077214	0.1511097	1.723105	2.3112188
Umbral[1]	1.5000000	0.0000000	1.500000	1.5000000
Umbral[2]	2.2887306	0.1342365	2.037156	2.5561827
Umbral[3]	3.1953991	0.1624686	2.881403	3.5065969
Umbral[4]	4.0711536	0.1779401	3.723709	4.4044102
Umbral[5]	5.5000000	0.0000000	5.500000	5.5000000

Similar a las figuras 4.4 y 4.5, se pueden mostrar las predicciones derivadas de cada

modelo, esta vez a través de muchas (300) distribuciones simuladas (a partir de las posteriores predictivas) de la respuesta al dilema contra la distribución observada de la respuesta (líneas rojas), trazando el número de predicciones posteriores en cada simulación como un diagrama lineal (líneas grises). Los resultados de estas simulaciones se muestran en las figuras 4.6, para el modelo métrico, y en la figura 4.7, para el modelo ordinal.

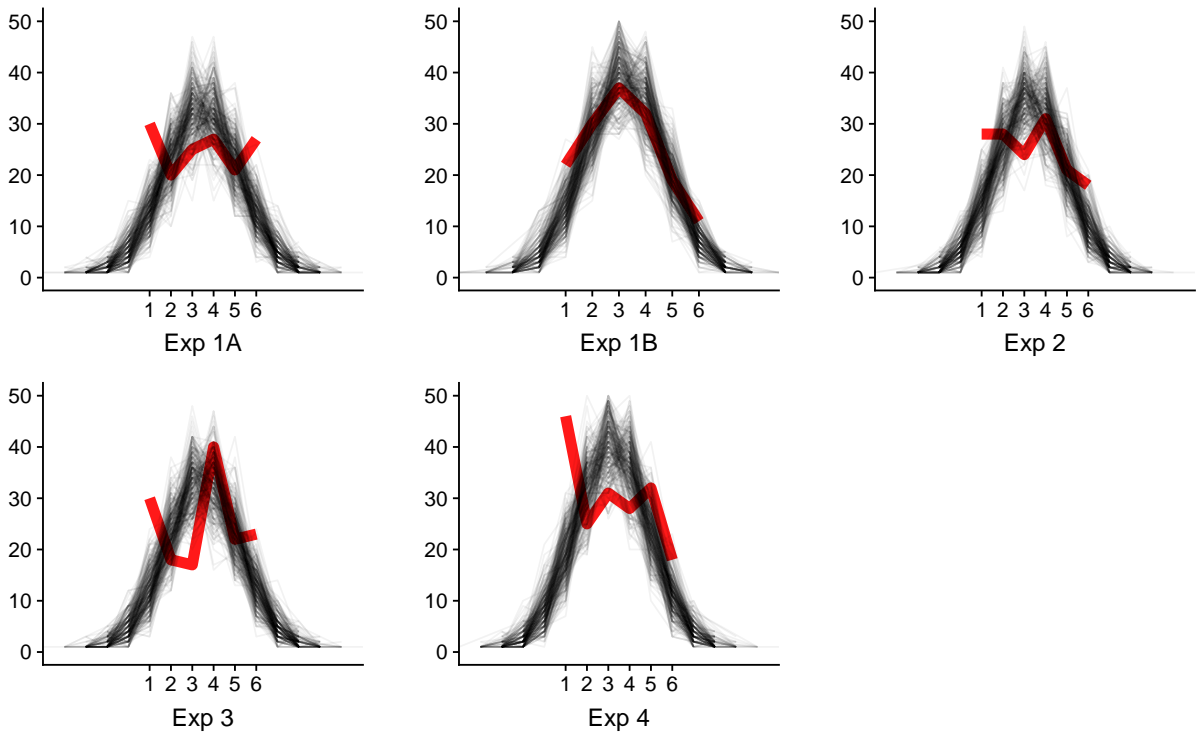


Figura 4.6: Predicciones posteriores métricas: Pryor, 2019

En la figura figuras 4.6 se puede observar fácilmente que el análisis métrico presenta los mismos problemas que se observaban en los paneles de la figura 4.4; se esperan valores fuera del rango de la respuesta ordinal, y las predicciones se concentran en el centro del rango de respuesta, a diferencia del patrón que exhiben las líneas rojas, excepto para el experimento 1B donde las líneas grises siguen correctamente el la distribución observada (que se concentra en el centro).

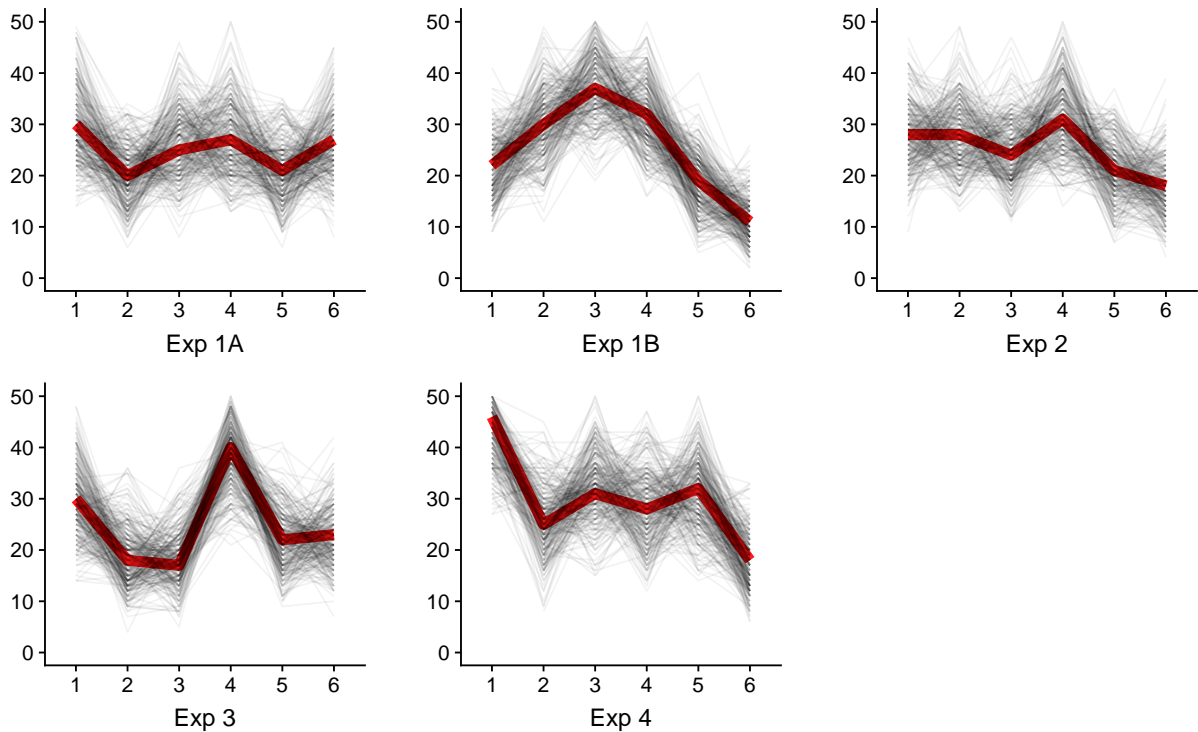


Figura 4.7: Predicciones posteriores ordinales: Pryor, 2019

En la figura 4.7 se aprecia tanto que las simulaciones se ajustan de mejor forma al patrón encontrado en los datos, pero que lo hace con mayor incertidumbre (las líneas grises no están tan concentradas como en las predicciones del modelo métrico), esto es especialmente notorio para el experimento 1A (que presenta una distribución e respuestas más o menos uniforme sobre el rango de los datos), pero para el resto de experimentos hace un buen trabajo, incluso capturando las ligeras diferencias presentadas en el comportamiento de las respuestas en el experimento 4.

Simulaciones (Kruschke, 2015)

Dado que esta base es más pequeña que la anterior se presentan, además de los modelos bayesianos, las estimaciones por máxima verosimilitud (modelos frecuentistas) con el fin

de complementar las comparaciones (añadiendo la comparación entre las aproximaciones frecuentista y bayesiana).

Las siguientes tablas resumen los valores estimados para X , b_0 y σ para los modelos métricos (4.8) y ordinales (4.9), para estos últimos se añaden las estimaciones de los umbrales.

Tabla 4.8: Estimaciones Modelos Lineales: Kruschke, 2015

Parámetro	Frecuentista				Bayesiano			
	Estimado	Desviación	Inferior	Superior	Estimado	Desviación	Inferior	Superior
X	6.7074	0.4305	5.8637	7.5512	6.7119	0.4435	5.9875	7.4456
b_0	-5.4218	0.6132	-6.6237	-4.2199	-5.4299	0.6370	-6.4930	-4.3923
σ	NA	NA	NA	NA	1.5295	0.0753	1.4132	1.6626

En la tabla 4.8 muestra la comparación entre los modelos lineales frecuentista y bayesiano; en esta comparación no se observan diferencias importantes entre las estimaciones del coeficiente de la variable explicativa X y el intercepto b_0 , principalmente porque al ajustar el modelo bayesiano se utilizaron distribuciones prior poco informativas de forma que las distribuciones posteriores se construyen sólo a partir de los datos. El hecho de no observar diferencias palpables muestra que la elección de prior no informativas, a pesar de no ser la mejor elección al no hacer uso pleno aproximación bayesiana, no ha alterado las inferencias que se pueden hacer de los datos.

Dado que las estimaciones entre ambos modelos son prácticamente iguales se podría pensar que el uso de estos modelos es indistinto, y que los intervalos pueden interpretarse de forma equivalente, esto no es correcto. Los intervalos de confianza y los intervalos de credibilidad tienen dos diferencias fundamentales: (1) la definición, los intervalos basados en la

inferencia bayesiana representa la probabilidad de contener el verdadero valor del parámetro, mientras que su contraparte clásica se construye a partir de la idea de una frecuencia límite; (2) la interpretación y la consistencia (coherencia) de su comportamiento (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2015; Nalborczyk, Bürkner, & Williams, 2018).

Respecto a este último punto, la interpretación de los intervalos de confianza suele llevar a las tres «Falacia de Confianza Fundamentales» (Morey et al., 2015):

1. Falacia de precisión: La anchura de un intervalo de confianza indica la precisión de la estimación del parámetro
2. Falacia de verosimilitud: Los valores dentro de un intervalo de confianza son al menos más verosímiles que los valores fuera del intervalo
3. Falacia de confianza: Si el intervalo de confianza fue calculada con una confianza del $X\%$, la probabilidad de que el intervalo contenga el verdadero valor del parámetro es X .

Por otro lado, los intervalos de credibilidad sí pueden interpretarse como este tipo de medida, de precisión, verosimilitud, y probabilidad de que el valor real esté contenido en el intervalo (ver Morey et al., 2015 y Nalborczyk, Bürkner, & Williams, 2018 para una discusión detallada).

Una diferencia que sí se aprecia fácilmente es la estimación de σ . Una diferencia entre las aproximaciones de análisis es que la implementación bayesiana mantiene la información de una distribución completa (la distribución posterior), mientras que la implementación frecuentista solo cuenta con las estimaciones por máxima verosimilitud sin una medida directa de incertidumbre (Kruschke, 2015).

Las mismas relaciones se encuentra entre los modelos ordinales frecuentista y bayesiano, las estimaciones de estos modelos se muestran en la siguiente tabla.

Tabla 4.9: Estimaciones Modelos Ordinales: Kruschke, 2015

Parámetro	Frecuentista				Bayesiano			
	Estimado	Desviación	Inferior	Superior	Estimado	Desviación	Inferior	Superior
X	9.5702	0.8206	7.9618	11.1787	9.4664	0.7713	8.0032	11.0259
b0	-9.5083	NA	NA	NA	-9.3562	1.0998	-11.6382	-7.3662
sigma	2.0943	NA	NA	NA	2.0443	0.1575	1.7665	2.3670
Umbral[1]	1.5000	1.0897	-0.6359	3.6359	1.5000	0.0000	1.5000	1.5000
Umbral[2]	2.3490	1.1121	0.1692	4.5288	2.3498	0.1703	2.0302	2.6846
Umbral[3]	3.3109	1.1450	1.0667	5.5551	3.3149	0.2053	2.9068	3.6995
Umbral[4]	4.4200	1.1879	2.0917	6.7484	4.4266	0.2093	3.9917	4.8100
Umbral[5]	5.4112	1.2279	3.0044	7.8180	5.4087	0.1874	5.0424	5.7757
Umbral[6]	6.5000	1.2743	4.0024	8.9976	6.5000	0.0000	6.5000	6.5000

La figura 4.8 muestra la comparación entre todos los parámetros comunes (X , b_0 y σ) de manera gráfica, de forma que se puede examinar de manera visual toda la información contenida en las dos tablas anteriores; es decir, la estimación (puntos) de cada parámetro (caja) con su intervalo de confianza o credibilidad (líneas que cruzan cada punto) para los diferentes métodos de ajuste (eje vertical), se usan los nombres de las funciones para referirse a los modelos, de forma que `lm` se refiere al métrico frecuentista, `polr` al ordinal frecuentista, `blm` al métrico bayesiano y `brm` al ordinal bayesiano.

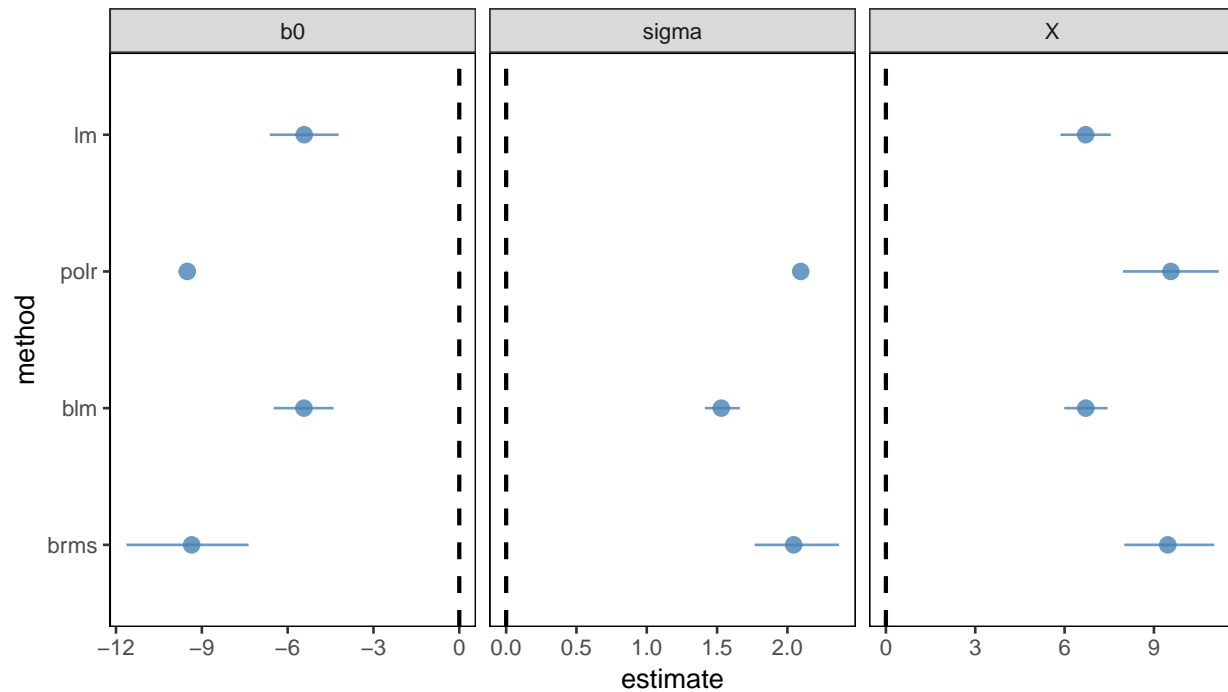


Figura 4.8: Comparación de modelo: Kruschke, 2015

Resalta el que, mientras las estimaciones son similares dentro de cada aproximación (frecuentista y la bayesiana), estas son muy distintas entre los métodos empleados (métrico y ordinal). La línea punteada que se observa en cada caja se dibuja sobre el 0, de forma que si el intervalo (de confianza o credibilidad) tocara este punto significaría que el 0 es un valor posible para el parámetro en cuestión; el color de los puntos depende de este aspecto, siendo rojo cuando cuando toca la línea punteada (que en este caso no es apreciable) y azul cuando no lo hace.

Sin embargo, y a pesar de las similitudes de las estimaciones frecuentistas y bayesianas (entre los modelos) hay unas cuantas razones por las que estas inferencias no deben tomarse como iguales (dos de ellas ya se discutieron antes): (1) los modelos bayesiana contienen la información de una distribución completa, y (2) los intervalos de confianza (frecuentistas) no

son iguales a los intervalos de credibilidad (bayesianos), (3) tanto los los intervalos de confianza como los p valores asociados con los parámetros estimados por los modelos frecuentistas dependen de supuestos asintóticos que pueden causar errores para tamaños de muestra pequeños, (4) una aproximación de un intervalo de confianza no está disponible para todos los parámetros del modelo en muchas circunstancias (este punto se puede apreciar en la figura 4.8), esta información es está disponible en la distribución posterior producida por un análisis bayesiano, y (5) los modelos frecuentistas suponen que en todos los niveles de los predictores, la varianza en el valor predicho es igual (Kruschke, 2015; Wagenmakers et al., 2018).

Similar a las figuras 4.6 y 4.7, la figura 4.9 muestra la comparación entre las predicciones hechas por el modelo métrico bayesiano contra las generadas por el ordinal bayesiano a través de la simulación múltiples distribuciones de la respuestas Y (líneas grises) comparadas con la distribución observada de los datos (línea roja).

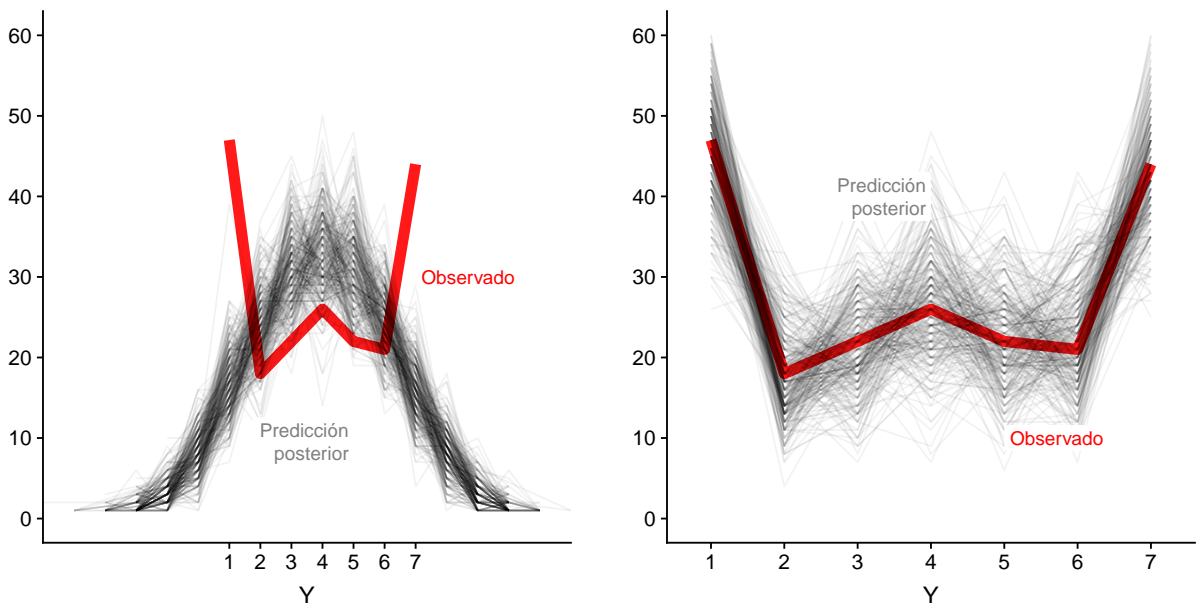


Figura 4.9: Predicciones posteriores: Kruschke, 2015

Una vez más, queda claro que el modelo ordinal bayesiano recupera de mejor forma la distribución hacia los extremos de los datos (ver figura 4.3), no así el modelo métrico que genera la mayoría de sus predicciones al rededor de los valores centrales.

Para el modelo ordinal bayesiano, y con el fin de evaluar el ajuste y las predicciones hechas por este, se pueden mostrar las probabilidades ajustadas ($P(Y|X)$) por el modelo contra la dispersión de la respuesta Y como función de X .

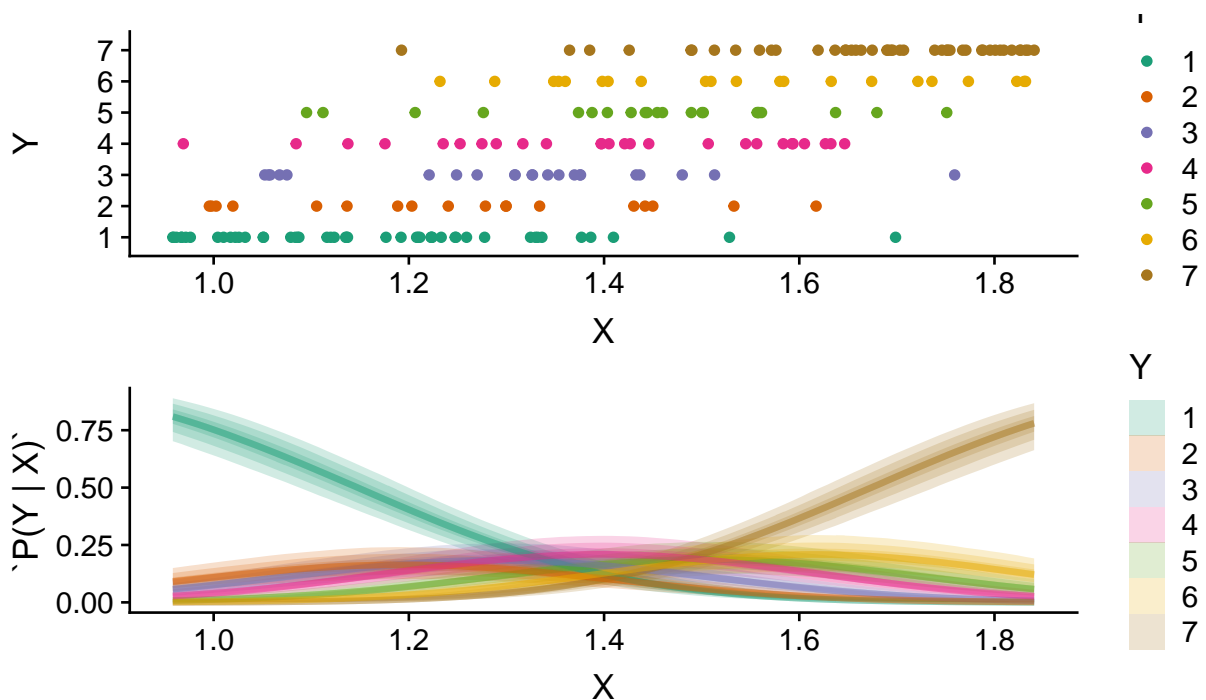


Figura 4.10: Probabilidades ajustadas

El panel superior de la figura 4.10 muestra la dispersión de los datos observados, el panel inferior las probabilidades ajustadas por el modelo para cada categoría de respuesta. En ambos paneles de esta figura se puede observar un comportamiento similar, los valores más pequeños de X están mayormente asociados con los valores más pequeños de Y y viceversa, mientras que el resto de datos tienen una ligera tendencia a concentrarse en el centro (muy

similar a lo observado en la figura 4.9).

También, para el modelo ordinal bayesiano, se puede graficar la comparación de las predicciones posteriores contra los datos observados, para ellos se generan predicciones para los valores observados de X (círculos grises) y se comparan con los datos observados (círculos de colores). Esta gráfica es otra forma de evaluar las predicciones del modelo, al igual que en el panel derecho de la figura 4.9.

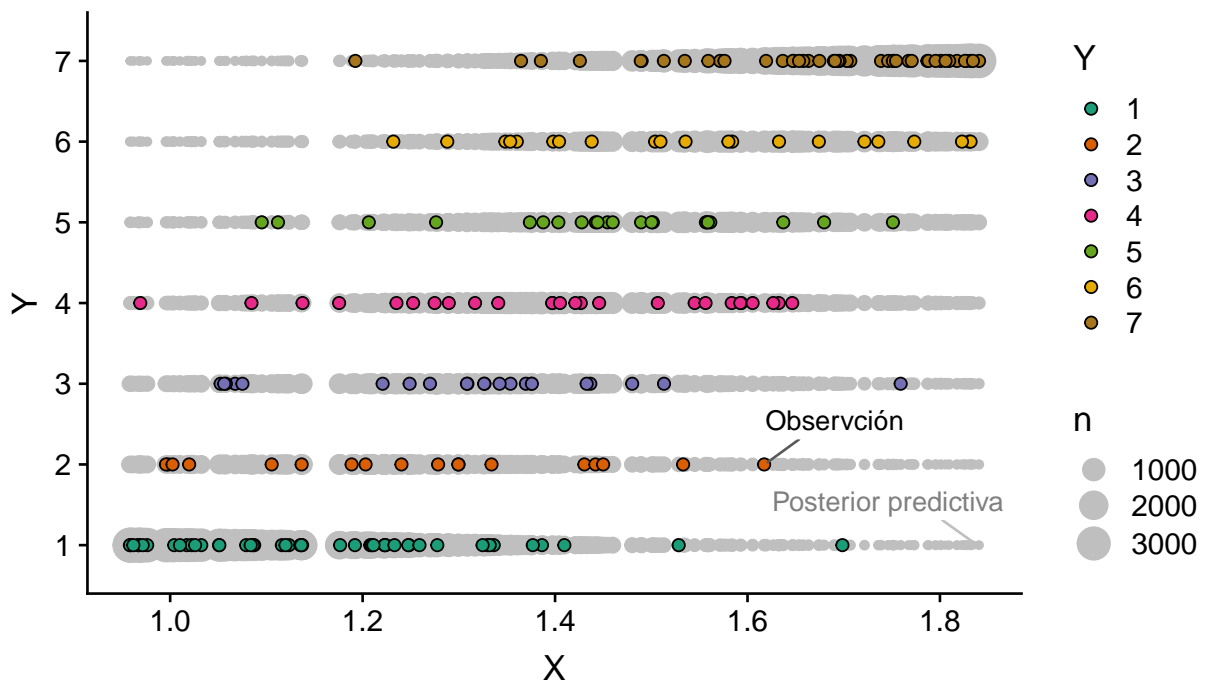


Figura 4.11: Predicciones posteriores

Se puede observar que las predicciones hechas por el modelo capturan bastante bien el comportamiento de los datos (valores más pequeños de X están mayormente asociados con valores más pequeños de Y y viceversa), esto se aprecia en que los círculos grises más grandes se encuentran en la cercanía de los círculos de colores; sin embargo, y como se había observado en las figuras anteriores, las predicciones se generan con cierto grado de incertidumbre, esto

es más evidente en las categorías más extremas de respuesta ya que se observan, aunque pocas, predicciones sobre todo el rango de X .

Capítulo 5

Conclusiones

Al principio de este trabajo se plantearon dos objetivos, uno conceptual y el otro práctico. Para alcanzar el objetivo conceptual se presentaron, de manera resumida y en español, algunos de los argumentos más recientes (de la mano de las simulaciones y la aproximación Bayesiana a la inferencia) en contra del análisis métrico de datos ordinales, apoyándose en el argumento original de Stevens al presentar la tipificación de las escalas de medición (a partir de la idea de la representatividad de la medición) y contrargumentando algunos de los puntos más comunes al momento de defender el uso de modelos paramétricos clásicos (i.e., cálculo de medias, desviaciones, regresiones lineales, etc.).

En lo que concierne al objetivo práctico, se presentó una herramienta gratuita, de fácil acceso y que promueve el modelamiento y juicio estadístico sobre el uso de rituales y algoritmos de decisión. A si mismo, se presentaron una serie de ejemplos mediante los cuales se explicó tanto el uso de el software estadístico R, los modelos ordinales acumulativos, y los conceptos más generales de la inferencia estadística y a la inferencia bayesiana, a la vez que

mostraban algunos de los problemas que (de manera sistemática) genera el trato métrico de variables ordinales.

En estos dos sentidos los modelos ordinales representan una aproximación más apropiada, incluso cuando ciertos aspectos de las inferencias son iguales (e.g., la estimación puntual del efecto de una variable predictora). Hay que recordar que, dado que en principio todos los modelos son falsos (es decir, el modelo ordinal no es «más cierto» que el métrico), se debe estar alerta de aquellos aspectos claves que se están tratando de representar a través de estos, en este caso la naturaleza ordenada de las categorías de respuesta del fenómeno de interés (e.g., el juicio sobre qué tan moralmente aceptada es una acción).

Por estas razones la conclusión final que debe considerarse es el cambio del default métrico por el tratamiento ordinal para estos conjuntos de datos, y en particular a través de la perspectiva bayesiana (por su flexibilidad, precisión y riqueza de información). En el peor de los casos, las estimaciones métricas pueden generar tasas altas de error Tipo I, pérdida de poder, e inversión de las inferencias, mientras que en el mejor de los casos las estimaciones métricas serán iguales a las estimaciones ordinales, y dado que la única forma de saber si son iguales es ajustando ambos modelos, mejor elegir el modelo ordinal como default.

Sin embargo, al considerar la aplicación de los modelos bayesianos hay que tener en cuenta un aspecto que aquí se pasó por alto, la elección de distribuciones prior. A pesar de que el uso de distribuciones prior impropias (como las que emplea el paquete `brms` para algunos parámetros) proporciona una forma razonable de utilizar la información suministrada por la función de verosimilitud (Akaike, 1980), y que con parámetros bien identificados y muestras de tamaño grande las distribuciones prior tienen efectos menores sobre las inferencias posteriores

(Gelman, 2002; Gelman et al., 2017), como muestran los análisis presentados a través de las distribuciones posteriores predictivas y la comparación con las estimaciones por máxima verosimilitud, una distribución prior impropia nunca es realmente correcta, incluso en cosas de las que literalmente no se tiene evidencia, por lo que la adecuación de las priors elegidas debe evaluarse en el contexto de los datos que se observan, tanto a través de las distribuciones posteriores predictivas (con se hizo en la presente tesis), como a través de las distribuciones priors predictivas (Gelman et al., 2017).

Apéndice A

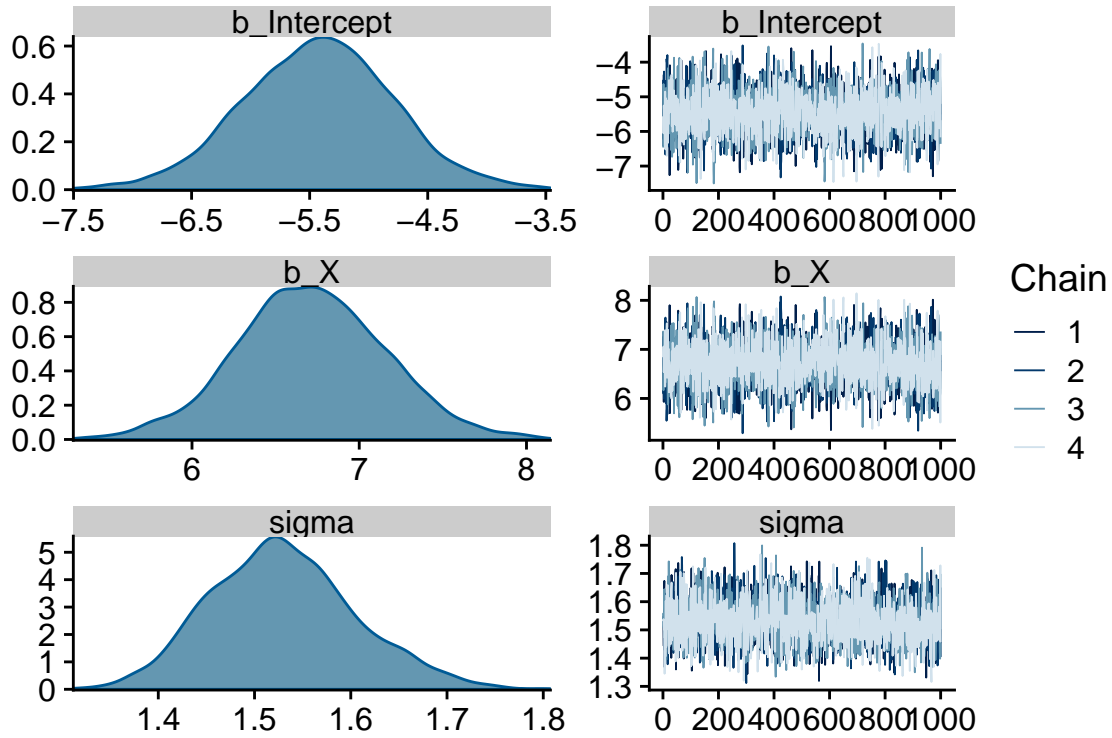
Aspectos Técnicos

En esta sección se discuten con mayor detalle algunos de los aspectos más técnicos de los modelos bayesianos empleados.

Para asegurarse de que no haya habido problemas de convergencia o «label switching» se pueden revisar los «trace plot» y el estadístico «Rhat» (debe ser cercano a 1) de cada parámetro para cada modelo. Para esta tesis se ajustaron 32 modelos bayesianos (las versiones métricas y ordinales de los 10 para cada ítem en el cuestionario de personalidad y 5 para cada experimento en Pryor et al. (2019), y 1 para la base de Kruschke (2015)), por lo que sólo se muestran estos criterios para los dos modelos usados para analizar la base proporcionada por Kruschke (2015).

Modelo métrico: Kruschke 2015

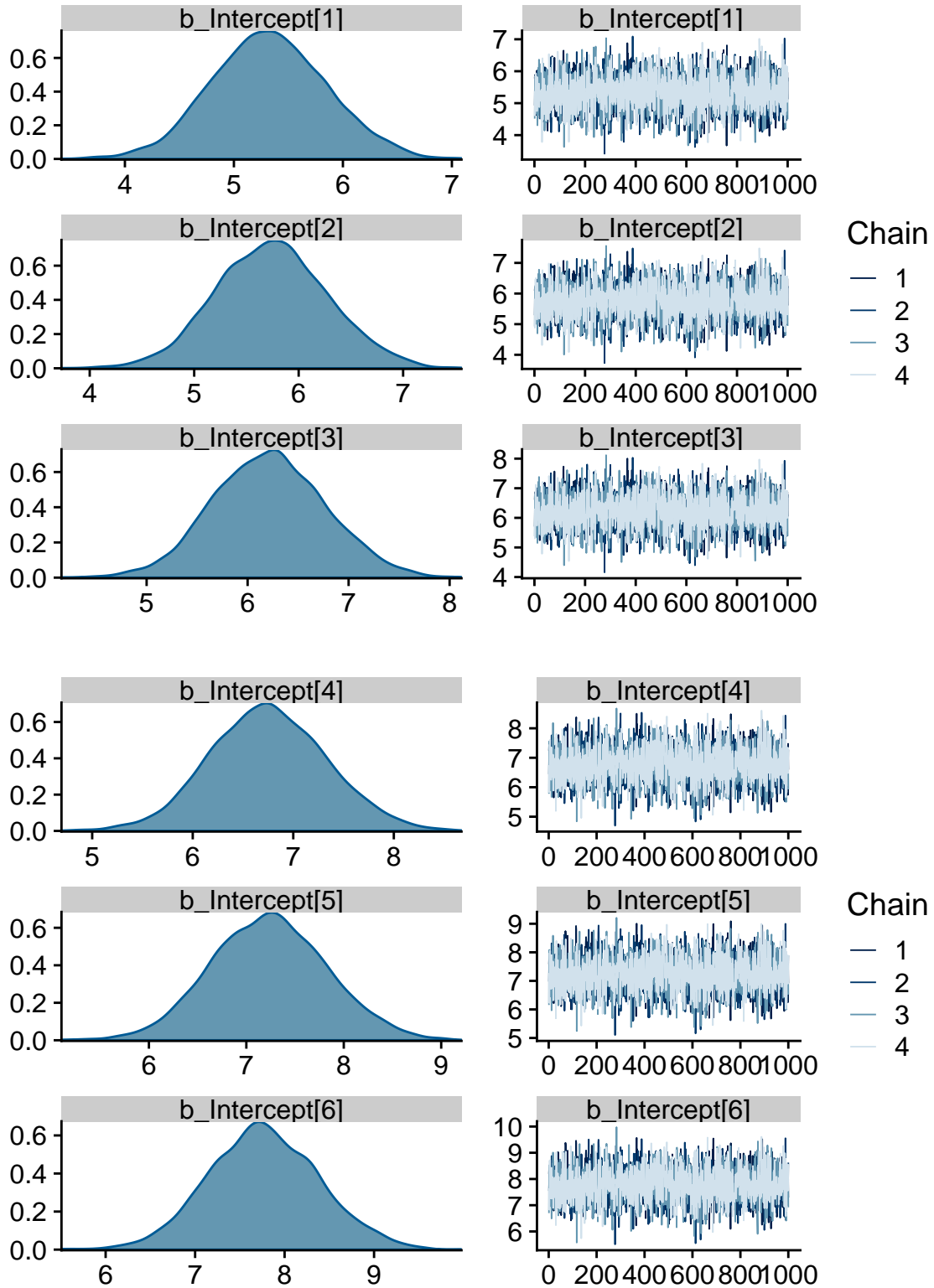
Las distribuciones posteriores de cada parámetro así como su `trace plot` se muestran en la siguiente gráfica.

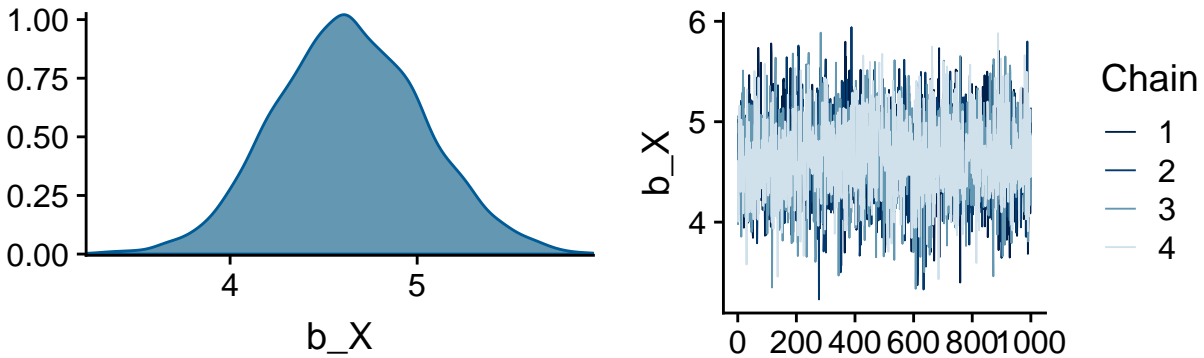


El comportamiento de las cadenas es el esperado, y el estadístico \hat{R} para todos los parámetros es igual a 1.00, por lo que no hubo problemas de convergencia.

Modelo Ordinal: Kruschke2015

Para los modelos ordinales las distribuciones posteriores y `trace plot` de los parámetros (incluidos los umbrales) se muestran en las siguientes gráficas.





Una vez más, no se observan problemas de convergencia (el estadístico \hat{R} es 1 para todos los parámetros). El comportamiento del resto de los modelos ajustados es similar al de los modelos aquí presentados; es decir, en ningún caso hubo problemas de convergencia o «label switching».

Se puede apreciar que los valores estimados de este modelo no coinciden con los reportados en la tabla 4.9 (que corresponden a este modelo ordinal), esto se debe a que tanto la función `polr()` como la función `brm()` asumen una media de 0 y una desviación estándar de 1 para la variable latente por default, esto produce estimaciones de parámetros que son difíciles de interpretar intuitivamente y que producen pendientes y umbrales que son relativos esta escala subyacente con poca relación directa con los valores ordinales en el datos. Kruschke (2015) propone que es más intuitivo fijar los umbrales de los extremos en valores que tengan sentido en la escala de datos ordinales, y estimar (a partir de los resultados arrojados por las funciones mencionadas) el resto de parámetros relativos a esos umbrales fijos. La transformación de las estimaciones a la escala de los datos se realizó de la siguiente manera:

```
sigmaMult = (K-2)/(Thresh[K-1]-Thresh[1])
inter = 1.5 - (sigmaMult*Thresh[1])
```

```
respThresh = sigmaMult*Thresh + inter  
respSigma = sigmaMult*Sigma  
respB0 = sigmaMult*Inter + inter  
respSlopes = sigmaMult*Slopes
```

Donde K es el número de categorías de respuesta de la variable ordinal, **Thresh** el vector de estimaciones de los $K - 1$ umbrales, **Slopes** es el vector de estimaciones de los coeficientes de las variables predictoras, **Inter** es la media de la variable latente (se asume que es 0), **Sigma** es la desviación estándar de la variable latente.

Apéndice B

Recursos en Línea

El primer conjunto de datos fue tomado de los experimentos presentados en Pryor, Perfors y Howe [(2019); disponibles en el repositorio de Open Science Framework: https://osf.io/n6uz5/?view_only=7dc67fcc0c1f4fdea8fe1dfe5d492480].

Los recursos proporcionados por Kruschke (2015) del libro «Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan», así como los datos empleados, se puede encontrar en la página dedicada al libro: (<https://sites.google.com/site/doingbayesiandataanalysis/>

La transformación de las estimaciones de los modelos ordinales está descrita en: <http://doingbayesiandataanalysis.blogspot.com/2014/11/ordinal-probit-regression-transforming.html>

El código de este manuscrito, así como un breve tutorial para la aplicación de los modelos ordinales aquí presentados está disponible en el repositorio de Github: <https://github.com/AlfMedina>

Referencias

- Abelson, R. P. (2012). *Statistics as Principled Argument*. Psychology Press.
- Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (Second Edi, pp. 173–203). New York: Wiley & Sons, Inc.
- Akaike, H. (1980). The Interpretation of Improper Prior Distributions as Limits of Data Dependent Proper Prior Distributions. *Journal of the Royal Statistical Society*, *42*(1), 46–52.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Belmont, CA: Wadsworth, Inc.
- Bago, B., Aczel, B., Zoltan, K., & Kovacs, M. (2019). Moral thinking across the world : Exploring the influence of personal force and intention in moral dilemma judgments. <http://doi.org/10.31234/osf.io/9uaqm>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions*, *53*, 370–418.
- Bishop, P. A., & Herron, R. L. (2015). Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *International Journal of Exercise Science*, *8*(3), 297–302.
- Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 791–799.

Bürkner, P. C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. <http://doi.org/10.18637/jss.v080.i01>

Bürkner, P. C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. <http://doi.org/10.1177/2515245918823199>

Carifio, J., & Perla, R. (2007). Ten Common Misunderstandings , Misconceptions , Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, *3*(3), 106–116.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, *42*, 1150–1152. <http://doi.org/10.1111/j.1365-2923.2008.03172.x>

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). <http://doi.org/10.18637/jss.v076.i01>

Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. Mason, Ohio: Cengage Learning.

Cushman, F., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment Testing Three Principles of Harm. *Psychological Science*, *17*(12), 1082–1089.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Macmillan International Higher Education.

Etz, A. (2018). Introduction to the Concept of Likelihood and Its Applications. *Advances in Methods and Practices in Psychological Science*, *1*(1), 60–69. <http://doi.org/10.1177/2515245917744314>

Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian Inference for Psychology. *Psychonomic Bulletin & Review*, *25*, 5–34. <http://doi.org/10.3758/s13423-017-1262-3>

Fisher, R. (1995). Statistical Methods and Scientific Induction. *Journal of the Royal*

Statistical Society: Series B (Methodological), 17(1), 69–78.

Fullerton, A. S. (2009). Framework for Ordered Logistic Regression Models. *Sociological Methods & Research*, 38(2), 306–347. <http://doi.org/10.1177/0049124109346162>

Gelman, A. (2002). Prior distribution Prior distribution. *Encyclopedia of Environmetrics*, 3, 1634–1637.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy*, 19(10), 1–13. <http://doi.org/10.3390/e19100555>

Gigerenzer, G. (2018). Statistical Rituals : The Replication Delusion and How We Got There. *Advances in Methods and Practices in Psychological Science*, 1(2), 198–218. <http://doi.org/10.1177/2515245918771329>

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42(3), 237–288.

Gosling, S. D., Rentfrow, P. J., & Jr, W. B. S. (2003). A very brief measure of the Big-Five personality domains q. *Journal of Research in Personality*, 37, 504–528. [http://doi.org/10.1016/S0092-6566\(03\)00046-1](http://doi.org/10.1016/S0092-6566(03)00046-1)

Jamieson, S. (2013). Likert scales: how to (ab)use them. *Medical Education*, 38, 1217–1218.

Knapp, T. R. (1990). Treating Ordinal Scales as Interval Scales: An Attempt To Resolve the Controversy. *Nursing Research*, 39(2), 121–123. <http://doi.org/10.1097/00006199-199003000-00019>

Kruschke, J. (2015). Ordinal Predicted Variable. In *Doing bayesian data analysis: A tutorial with r, jags, and stan* (Academic P, pp. 671–700). Boston, USA: Academic Press. <http://doi.org/10.1016/B978-0-12-405888-0.00023-4>

Kuzon, W. M. J., Urbanchek, M. G., & McCabe, S. (1996). The Seven Deadly Sins of Statistical Analysis. *Annals of Plastic Surgery*, 37, 265–272. <http://doi.org/10.1097/>

00000637-199701000-00026

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology, 79*, 328–348. <http://doi.org/10.1016/j.jesp.2018.08.009>

McDonald, R. P. (2011). *Test Theory : A Unified Treatment*. Routledge.

McElreath, R. (2016). *Statistical Rethinking A Bayesian Course with Examples in R and Stan*. Boca Raton, Florida, USA: Chapman & Hall/CRC Press.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-j. (2015). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*. <http://doi.org/10.3758/s13423-015-0947-8>

Nalborczyk, L., Bürkner, P. C., & Williams, D. R. (2018). Pragmatism should not be a substitute for statistical literacy, a commentary on Albers, Kiers, and van Ravenzwaaij (2018). *PsyArXiv Preprints*, (<https://psyarxiv.com/fmtkj/>).

Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education: Theory and Practice, 15*(5), 625–632. <http://doi.org/10.1007/s10459-010-9222-y>

Pashler, H., & Wagenmakers, E. J. (2012). Editors ’ Introduction to the Special Section on Replicability in Psychological Science : A Crisis of Confidence ? *Perspectives on Psychological Science, 7*(6), 2011–2013. <http://doi.org/10.1177/1745691612465253>

Pryor, C., Perfors, A., & Howe, P. D. L. (2019). Even arbitrary norms influence moral decision-making. *Nature Human Behaviour, 3*(January), 57–62. <http://doi.org/10.1038/s41562-018-0489-y>

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>

Rodgers, J. L. (2010). The Epistemology of Mathematical and Statistical Modeling. *American Psychological Association*, 65(1), 1–12. <http://doi.org/10.1037/a0018326>

Solís, V. M. (2014). ¿Por qué algunos aún prohíben utilizar estadística paramétrica para analizar datos ordinales? *Enseñanza E Investigación En Psicología*, 19(2).

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680.

Stevens, S. S. (1958). Measurement and Man. *Science*, 127(3295), 383–389.

Stevens, S. S. (1968). Measurement, Statistics, and the Schemapiric View. *Science*, 161(3844), 849–856.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... Morey, R. D. (2018). Bayesian inference for psychology . Part I : Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25, 35–57. <http://doi.org/10.3758/s13423-017-1343-3>