Lay Understanding of Outliers

Jennifer E. Dannals Stanford University GSB

Daniel M. Oppenheimer Carnegie Mellon University

SJDM 2017





divorce)







		lar	านส	ary	1			F	ek	oru	ar	y			March								April									
S	М	Т	W	Т	F	S	S	М	т	W	т	F	S	S	М	Т	W	т	F	S	S	М	т	w	Т	F	S					
1	2	3	4	5	6	7				1	2	3	4				1	2	3	4							1					
8	9	10	11	12	13	14	5	6	7	8	9	10	11	5	6	7	8	9	10	11	2	3	4	5	6	7	8					
15	16	17	18	19	20	21	13	13	14	15	16	17	18	1	2 1 3	14	15	16	17	18	9	10	11	12	13	14	15					
22	22	24	25	26	20	20	10	20	21	22	22	24	25	1	20	21	22	22	24	25	16	17	18	19	20	21	22					
22	23	24	25	20	21	20	1:	, 20	21	22	23	24	25	1	9 20	21	22	23	24	25	23	24	25	26	27	28	29					
29	30	31					26	5 27	28					2	5 27	28	29	30	31		30)										
May								June								July								August								
S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S					
	1	2	3	4	5	6					1	2	3							1			1	2	3	4	5					
7	8	9	10	11	12	13	4	5	6	7	8	9	10	2	3	4	5	6	7	8	6	7	8	9	10	11	12					
14	15	16	17	18	19	20	1	12	13	14	15	16	17	9	10	11	12	13	14	15	13	14	15	16	17	18	19					
21	22	22	24	25	26	27	10	10	20	21	22	22	24	1	5 17	18	19	20	21	22	20	21	22	22	24	25	26					
21	22	23	24	25	20	21	10	, 19	20	21	22	23	24	2	3 24	25	26	27	28	29	20	21	22	23	24	25	20					
28	29	30	31	_			25	26	27	28	29	30		3	31						27	28	29	30	31							
	S	ept	er	nb	er				Oc	tol	bei	-			Ν	lov	en	nbe	er			D	ec	en	nbe	er						
S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S					
					1	2	1	2	3	4	5	6	7				1	2	3	4						1	2					
3	4	5	6	7	8	9	8	9	10	11	12	13	14	5	6	7	8	9	10	11	3	4	5	6	7	8	9					
10	11	12	13	14	15	16	15	5 16	17	18	19	20	21	1	2 13	14	15	16	17	18	10	11	12	13	14	15	16					
17	10	10	20	21	22	22	2	222	24	25	26	27	20	1	20	21	22	22	24	25	17	18	19	20	21	22	23					
1/	10	19	20	21	22	23	24	23	24	20	20	21	20	1	20	21	22	23	24	25	24	25	26	27	28	29	30					
24	25	26	27	28	29	30	29	30	31					2	5 27	28	29	30			31											





																											ba b	aby orn	
	1	lar	านส	ar	/			F	Feb	oru	ar	v				М	are	ch					A	pr	il				
S	м	т	w	т	F	S	S	М	т	w	т	F	S	S	М	т	W	т	F	S	S	М	т	w	т	F	S	· /	
1	2	3	4	5	6	7				1	2	3	4				1	2	3	4							1		
8	9	10	11	12	13	14	5	6	7	8	9	10	11	5	6	7	8	9	10	11	2	3	4	5	6	7	8		
15	16	17	18	19	20	21	12	13	14	15	16	17	18	12	13	14	15	16	17	18	9	10	11	12	13	14	15		
22	22	24	25	26	20	20	10	20	21	22	22	24	25	10	20	21	22	22	24	25	16	17	18	19	20	21	22		
22	23	24	25	20	21	20	19	20	21	22	23	24	25	19	20	21	22	23	24	25	23	24	25	26	27	28	29		
29	30	31					26	27	28					26	5 27	28	29	30	31		30								
		Γ	/lay	у				June									ul	У			August								
S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S		
	1	2	3	4	5	6					1	2	3							1			1	2	3	4	5		
7	8	9	10	11	12	13	4	5	6	7	8	9	10	2	3	4	5	6	7	8	6	7	8	9	10	11	12		
14	15	16	17	18	19	20	11	12	13	14	15	16	17	9	10	11	12	13	14	15	13	14	15	16	17	18	19		
21	22	23	24	25	26	27	18	19	20	21	22	23	24	10	1/	18	19	20	21	22	20	21	22	23	24	25	26		
28	20	30	31				25	26	27	28	20	30		23	24	25	20	21	20	29	27	28	20	30	31				
20	20	50	51	-			20	20	- /	20	2.5	50		50	51		-	_	_		2,	20	20	50	51				
	S	ept	en	nb	er				Oc	tol	bei	r			N	lov	en	nbe	er			D	ec	en	nbe	er			
S	М	Ť	W	т	F	S	S	М	т	W	т	F	S	S	М	т	W	т	F	S	S	М	т	W	т	F	S		
					1	2	1	2	3	4	5	6	7				1	2	3	4						1	2		
3	4	5	6	7	8	9	8	9	10	11	12	13	14	5	6	7	8	9	10	11	3	4	5	6	7	8	9		
10	11	12	13	14	15	16	15	16	17	18	19	20	21	12	13	14	15	16	17	18	10	11	12	13	14	15	16		
17	18	10	20	21	22	23	22	23	24	25	26	27	28	10	20	21	22	23	24	25	17	18	19	20	21	22	23		
1	10	19	20	21	22	20	22	20	24	25	20	21	20	13	20	21	22	20	24	25	24	25	26	27	28	29	30		
24	25	26	27	28	29	30	29	30	31					126	27	28	29	30			31								



lanuarv							February							March								April									
S	м	т	w	т	F	S	S	М	т	w	т	F	S		S	М	т	w	т	F	S	S	М	т	w	т	F	S			
1	2	3	4	5	6	7				1	2	3	4					1	2	3	4							1			
8	9	10	11	12	13	14	5	6	7	8	9	10	11		5	6	7	8	9	10	11	2	3	4	5	6	7	8			
15	16	17	18	19	20	21	12	13	14	15	16	17	18		12	13	14	15	16	17	18	9	10	11	12	13	14	15			
22	23	24	25	26	27	28	19	20	21	22	23	24	25		19	20	21	22	23	24	25	16	17	18	19	20	21	22			
29	30	31					26	27	28						26	27	28	29	30	31		30	24	25	20	21	28	29			
						20	27	20						20	27	20	23	50	51		00										
		Ν	1ay	У				June								July								August							
S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S		S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S			
	1	2	3	4	5	6					1	2	3					-		_	1			1	2	3	4	5			
7	8	9	10	11	12	13	4	5	6	7	8	9	10		2	3	4	5	6	7	8	6	7	8	9	10	11	12			
14	15	16	17	18	19	20	11	12	13	14	15	16	17		9	10	11	12	13	14	15	13	14	15	1	17	18	19			
21	22	23	24	25	26	27	18	19	20	21	22	23	24		23	24	25	26	27	21	29	20	21	22	23	Z 4	25	26			
28	29	30	31				25	26	27	28	29	30			30	31	2.5	20		20	2.5	27	28	29	30	31					
	_																														
	Se	ept	en	nb	er			October								Ν	ov	en	nbe	er		December									
S	М	Т	w	т	F	S	S	M	T	W	T	F	S		S	М	Т	W	T	F	s	S	м	т	W	т	F	S			
_		_	-	_	1	2	1	2	3	4	5	6	/				_	1	2	3	4	3	4	5	6	7	8	2			
3	4	5	6	7	8	9	8	9	10	11	12	13	14		5	6	7	8	9	10	11	10	11	12	13	14	15	16			
10	11	12	13	14	15	16	15	16	17	18	19	20	21		12	13	14	15	16	17	18	17	18	19	20	21	22	23			
17	18	19	20	21	22	23	22	23	24	25	26	27	28		19	20	21	22	23	24	25	24	25	26	27	28	29	30			
24	25	26	27	28	29	30	29	30	31						26	27	28	29	30			31									
																											•				
							<u>_</u>	2/	1 C)	C		יכ	1		6	20	וב	-		r										
							C)~	t_{c}	2	C	IC	יב	УŸ	S	C	70	וג			71						۱.				
																											L				
															Manah																
		Ja	nu	ar	У			February								March								April							
5	M	T	W	T	F	S	S I	M	т	W	T	F	S	1	S	M	т	W	T	F	S	S	M	т	W	т	f	<u>S</u>			
1	2	3	4	5	6	/		-	-	1	2	3	4				-	1	2	3	4	2	3	4	5	6	7	8			
8	9	10	11	. 12	2 13	14	5	6	7	8	9	10) 11		5	6	7	8	9	10	11	9	10	11	12	13	14	15			
15	16	5 17	18	19	20	21	12	2 13	3 14	15	16	17	18		12	13	14	15	16	17	18	16	5 17	18	19	20	21	22			
22	23	3 24	25	26	5 27	28	19	9 20	21	22	23	24	1 25		19	20	21	. 22	23	24	25	23	3 24	25	26	27	28	29			
29	30	31					26	5 27	28	:					26	27	28	3 29	30	31		30)								
			1-								~							1						۸.		ict					
c	м		via	У	F	c	c	м	Ţ	ur	Te T	F	c		c	м		u u	У_	E	c	c	м	AL	ugu w	JSU	-	L			
5	1	2	3	4	5	6		1*1	<u> </u>		1	2	3	1	L_	1*1	T.		- '	- F	1		1*1	1	2	3	4	5			
7	-	0	10	11	12	12	1	5	6	7	0	0	10		2	3	4	5	6	7	8	6	7	•	0	10	11	12			
14	15	16	17	10	10	20	1 1 1	117	1 1 2	14	15	9	117		9	10	11	12	13	14	15	1 2	14	15	9	17	10	÷			
14	10	10	1/	18	19	20	11	1 12	13	14	10	10	, 1/		16	17	18	19	20	21	22	1.	14	13	10	1/	10	15			
21	22	23	24	25	26	27	18	\$ 19	20	21	22	23	\$ 24		23	24	25	26	27	28	29	20	21	22	23	24	25	26			
28	29	30	31				25	26	27	28	29	30)]	30	31						27	28	29	30	31					
	S	en	ter	mh	her			October								November								December							
s	м	T	w	T	F	s	S	м	т	w	T	F	S		S	м	Т	w	т	F	S	S	м	Т	W	т	F	S			
-		-		-	-	-						-	_	1	<u> </u>		÷		-	-			-	_		_		_			

January								rebluary								March								Аргіі								
S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S	S	М	Т	W	Т	F	S					
1	2	3	4	5	6	7				1	2	3	4				1	2	3	4							1					
8	9	10	11	12	13	14	5	6	7	8	9	10	11	5	6	7	8	9	10	11	2	3	4	5	6	7	8					
15	16	17	18	19	20	21	12	13	14	15	16	17	18	12	13	14	15	16	17	18	9	10	11	12	13	14	15					
22	22	24	25	26	27	20	10	20	21	22	22	24	25	10	20	21	22	22	24	25	16	17	18	19	20	21	22					
22	23	24	25	20	27	28	19	20	21	22	23	24	25	19	20	21	22	23	24	25	23	24	25	26	27	28	29					
29	30	31					26	27	28					26	27	28	29	30	31		30		_		_		_					
		Γ	1a	у				June								July								August								
S	М	Т	W	Т	F	S	S	М	Ť	W	Т	F	S	S	М	Т	W	Т	F	S	S	М	Т	W	Т	F						
	1	2	3	4	5	6					1	2	3							1			1	2	3	4	5					
7	8	9	10	11	12	13	4	5	6	7	8	9	10	2	3	4	5	6	7	8	6	7	8	9	10	11	12					
14	15	16	17	18	19	20	11	12	13	14	15	16	17	9	10	11	12	13	14	15	13	14	15	16	17	18	19					
21	22	23	24	25	26	27	18	19	20	21	22	23	24	10	24	18	19	20	21	22	20	21	22	23	24	25	26					
28	29	30	31				25	26	27	28	29	30		30	31	25	20	21	20	23	27	28	29	30	31							
	Se	ept	er	nb	er			October								November								en	nbe	er						
S	М	т	W	т	F	S	S	м	т	W	т	F	S	S	М	т	W	т	F	S	S	м	т	W	т	F	S					
					1	2	1	2	3	4	5	6	7				1	2	3	4						1	2					
3	4	5	6	7	8	9	8	9	10	11	12	13	14	5	6	7	8	9	10	11	3	4	5	6	7	8	9					
10	11	12	13	14	15	16	15	16	17	18	10	20	21	12	13	14	15	16	17	18	10	11	12	13	14	15	16					
1.7	10	10	20	21	22	22	22	22	24	25	26	27	20	10	20	21	22	22	24	25	17	18	19	20	21	22	23					
11/	18	19	20	21	22	23	22	23	24	25	26	21	28	19	20	21	22	23	24	25	24	25	26	27	28	29	30					
24	25	26	27	28	29	30	29	30	31					26	27	28	29	30			31											











14Ginola Halifax, United Kingdom

Reviewer



"Not as bad as reviews state."

Reviewed 4 weeks ago

For what you pay for this hotel it's fine. not half as bad as reviews state. It's typically 70's furniture, flowery wallpaper, blown vinyl, toilets are of 70's colours. But then bed linen was clean & comfy. Down stairs was cosy with a nice little bar. Cheap drinks & friendly staff who know they are not running the Hilton &...

More -

Was this review helpful? Yes

< 19



14Ginola Halifax, United Kingdom

Reviewer



"Not as bad as reviews state."

For what you pay for this hotel it's fine. not half as bad as reviews state. It's typically 70's furniture. flowery wallpaper, blown vinyl, toilets are of





14Ginola Halifax, United Kingdom

Reviewer 3 reviews

"Not as bad as reviews state."

For what you pay for this hotel it's fine. not half as bad as reviews state. It's typically 70's furniture, flowery wallpaper, blown vinyl, toilets are of 70's colours. But then bed linen was clean & comfy. Down stairs was cosy with a nice little bar. Cheap drinks & friendly staff who know they

Like the reviewers, I like your manuscript very much. I will not repeat their concerns here. They are relatively minor, and I think you can address them in a revision.

I actually have nothing to add above the points they made. I only have a minor comment. I recommend not making researchers the subjects of your sentences. Put them in parentheses at the end of sentences. Research findings and theories should be the subjects of your sentences, rather than the people who did the research or proposed the theory.

Thus, I am conditionally accepting your manuscript for publication. The condition is that you address the concerns raised by the reviewers. I will send your manuscript back to them. All have agreed to review a revised revision.



"an observation which deviates so much from other observations as to cause suspicions that it was generated by a different mechanism."

(Hawkins, 1980)

"an observation which deviates so much from other observations *as to cause suspicions that it was generated by a different mechanism.*"

(Hawkins, 1980)

"an observation which **deviates** so much from other observations as to cause suspicions that it was generated by a different mechanism."

(Hawkins, 1980)

X_? (a potential outlier)





How do individuals incorporate outliers from a sample into their predictions of the population distribution?

the task



Directional predictions – overweighting

(Tversky & Kahneman, 1974; Lichtenstein, et al., 1978; Sunstein & Zeckhauser, 2011; Rothman, Klein, & Weinstein, 1996; Cruciani, Berardi, Cabib, & Conversi, 2011; Brown & Kulik, 1977; Christianson & Loftus, 1987; Madan, Ludvig, & Spetch, 2014; Ludvig, Madan, & Spetch, 2014).

Directional predictions – overweighting

(Tversky & Kahneman, 1974; Lichtenstein, et al., 1978; Sunstein & Zeckhauser, 2011; Rothman, Klein, & Weinstein, 1996; Cruciani, Berardi, Cabib, & Conversi, 2011; Brown & Kulik, 1977; Christianson & Loftus, 1987; Madan, Ludvig, & Spetch, 2014; Ludvig, Madan, & Spetch, 2014).

Use sample statistics (e.g. mean, median)

(Goldstein & Rothschild, 2014; Griffiths & Tenenbaum, 2006; Leider, Griffiths & Hsu, 2017)

Directional predictions – overweighting

(Tversky & Kahneman, 1974; Lichtenstein, et al., 1978; Sunstein & Zeckhauser, 2011; Rothman, Klein, & Weinstein, 1996; Cruciani, Berardi, Cabib, & Conversi, 2011; Brown & Kulik, 1977; Christianson & Loftus, 1987; Madan, Ludvig, & Spetch, 2014; Ludvig, Madan, & Spetch, 2014).

Use sample statistics (e.g. mean, median)

(Goldstein & Rothschild, 2014; Griffiths & Tenenbaum, 2006; Leider, Griffiths & Hsu, 2017)

Use sample statistics but discount eventually

(Obrecht, Chapman & Suárez, 2010; Obrecht, Chapman & Gelman, 2007)

Directional predictions – overweighting

(Tversky & Kahneman, 1974; Lichtenstein, et al., 1978; Sunstein & Zeckhauser, 2011; Rothman, Klein, & Weinstein, 1996; Cruciani, Berardi, Cabib, & Conversi, 2011; Brown & Kulik, 1977; Christianson & Loftus, 1987; Madan, Ludvig, & Spetch, 2014; Ludvig, Madan, & Spetch, 2014).

Use sample statistics (e.g. mean, median)

(Goldstein & Rothschild, 2014; Griffiths & Tenenbaum, 2006; Leider, Griffiths & Hsu, 2017)

Use sample statistics but discount eventually

(Obrecht, Chapman & Suárez, 2010; Obrecht, Chapman & Gelman, 2007)

Statistical benchmark - Tests of discordancy

Directional predictions – overweighting

(Tversky & Kahneman, 1974; Lichtenstein, et al., 1978; Sunstein & Zeckhauser, 2011; Rothman, Klein, & Weinstein, 1996; Cruciani, Berardi, Cabib, & Conversi, 2011; Brown & Kulik, 1977; Christianson & Loftus, 1987; Madan, Ludvig, & Spetch, 2014; Ludvig, Madan, & Spetch, 2014).

Use sample statistics (e.g. mean, median)

(Goldstein & Rothschild, 2014; Griffiths & Tenenbaum, 2006; Leider, Griffiths & Hsu, 2017)

Use sample statistics but discount eventually

(Obrecht, Chapman & Suárez, 2010; Obrecht, Chapman & Gelman, 2007)

Statistical benchmark - Tests of discordancy

$$T = \frac{excess}{outlier} = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)}}$$

if sample is exponential distribution

$$T = \frac{excess}{range} = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$$

if sample is normal distribution

our contexts





our contexts



~3500 Train Arrival Times from week proceeding survey



~2500 Textbook Prices from quarter survey was offered







IV: latest train manipulated at 10 levels from 6 minutes to 54 minutes



IV: most expensive book manipulated at 9 levels from \$36.95 to \$361.25



IV: latest train manipulated at 10 levels from 6 minutes to 54 minutes

> 5 minutes 54 minutes 2 minutes 0 minutes 3 minutes



IV: most expensive book manipulated at 9 levels from \$36.95 to \$361.25

> \$361.25 \$24.00 \$20.00 \$19.00 \$29.00 \$16.95



IV: latest train manipulated at 10 levels from 6 minutes to 54 minutes

> 5 minutes 54 minutes 2 minutes 0 minutes 3 minutes

Based on this week, when do you think the train will arrive, on average, next week?



IV: most expensive book manipulated at 9 levels from \$36.95 to \$361.25

> \$361.25 \$24.00 \$20.00 \$19.00 \$29.00 \$16.95

Based on this sample, how much would you say a single course book costs, on average?



IV: latest train manipulated at 10 levels from 6 minutes to 54 minutes

> 5 minutes 54 minutes 2 minutes 0 minutes 3 minutes

Based on this week, when do you think the train will arrive, on average, next week?

199 participants995 predictions



IV: most expensive book manipulated at 9 levels from \$36.95 to \$361.25

> \$361.25 \$24.00 \$20.00 \$19.00 \$29.00 \$16.95

Based on this sample, how much would you say a single course book costs, on average?

214 participants1070 predictions

analytic approach

Robust Regression

as a linear mixed-model with random participant intercepts

analytic approach

Robust Regression

as a linear mixed-model with random participant intercepts

Quantile Regression

with bootstrapped SEs clustered by participant predicting the 10th, 50th and 90th percentiles





Latest Train is 12 minutes











Latest Train is 12 minutes



Most Expensive Book is \$53.50



results

Evaluate participants against 4 predictions:

- 1. Overweighting
- 2. Sample mean
- 3. Discordancy tests
- 4. Sample median

Is there evidence of giving outliers extra weight?

Is there evidence of giving outliers extra weight?



Is there evidence of giving outliers extra weight?





Latest Train is 12 minutes





Most Expensive Book is \$53.50



Do participants use the sample mean to predict population mean?



Do participants use the sample mean to predict population mean?



When do participants (on average) start to discount outliers?



When do participants (on average) start to discount outliers?



Do participants use the sample median to predict population mean?

Do participants use the sample median to predict population mean?



Do participants use the sample median to predict population mean?





- about 10% of participants give extra weight to outliers in a sample
- about 10% of participants also always ignore outliers

- about 10% of participants give extra weight to outliers in a sample
- about 10% of participants also always ignore outliers
- On average, participants start to discount early, but don't discount completely

- about 10% of participants give extra weight to outliers in a sample
- about 10% of participants also always ignore outliers
- On average, participants start to discount early, but don't discount completely

open questions

- about 10% of participants give extra weight to outliers in a sample
- about 10% of participants also always ignore outliers
- On average, participants start to discount early, but don't discount completely

open questions

• other distributions?

- about 10% of participants give extra weight to outliers in a sample
- about 10% of participants also always ignore outliers
- On average, participants start to discount early, but don't discount *enough*

open questions

- other distributions?
- individual differences in outlier appraisal?







Barnett, 1978



Barnett, 1978

thank you!

thoughts?

email me at jdannals@stanford.edu

or find me at www.jendannals.com