Package 'samplrData'

October 26, 2026

Title Datasets from the SAMPLING Project

Version 1.0.1

Maintainer Lucas Castillo <lucas.castillo@warwick.ac.uk>

Description Contains human behaviour datasets collected by the SAMPLING project (<https://sampling.warwick.ac.uk>).

License CC BY 4.0

Encoding UTF-8

URL https://github.com/lucas-castillo/samplrData, https://lucas-castillo.github.io/samplrData/

BugReports https://github.com/lucas-castillo/samplrData/issues

Depends R (>= 2.10)

LazyData true

Imports Rdpack

RdMacros Rdpack

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Lucas Castillo [aut, cre, cph] (ORCID: <https://orcid.org/0000-0003-0274-0777>), Yun-Xiao Li [aut, cph] (ORCID: <https://orcid.org/0000-0002-3509-6618>), Adam N Sanborn [aut, cph] (ORCID: <https://orcid.org/0000-0003-0442-4372>), European Research Council (ERC) [fnd]

Repository CRAN

Date/Publication 2026-10-26 13:00:08 UTC

Contents

castillo2024.rgmomentum.e1 2
castillo2024.rgmomentum.e2 3

spicer2022.anchoringrepulsion.e1 5
spicer2022.anchoringrepulsion.e2 6
spicer2022.anchoringrepulsion.e2a 7
sundh2023.meanvariance.e3 8
sundh2023.meanvariance.e4 9
zhu2020.bayesiansampler.e1 10
zhu2020.bayesiansampler.e2 11
zhu2022.coherenceaccuracy.e1 12
zhu2022.coherenceaccuracy.e2 13
zhu2022.structurenoise.animals 14
zhu2022.structurenoise.time 15

Index **16**

`castillo2024.rgmomentum.e1`

Data from Experiment 1 in Castillo et al. (2024)

Description

Participants produced a random sequence of heights of either men or women in the United Kingdom. In one sequence, they sampled heights as distributed according to a uniform distribution (Uniform condition); in the other sequence, heights were distributed following their actual distribution (which is roughly Gaussian). These data are licensed under CC BY 4.0, reproduced from materials in OSF.

id participant id

part_Gender participant's gender (self-reported)

part_Height participant's own height (self-reported)

part_Home participant's home country (self-reported)

RQ_Rep percentage of correct responses in Randomness Questionnaire, for coin toss pairs where one sequence had too many repetitions

RQ_Alt percentage of correct responses in Randomness Questionnaire, for coin toss pairs where one sequence had too many alternations

RQ_GFM percentage of correct responses in Randomness Questionnaire, Gambling Fallacies Measure section

minHeight height participant reports to be the shortest adult in the UK (from target gender)

maxHeight height participant reports to be the tallest adult in the UK (from target gender)

condition whether the participant did the uniform condition first (UN) or not (NU)

target_gender gender they had to generate heights from, either male (M) or female (F)

index position of the item in the sequence, 0 indexed

block whether the item belongs to the first sequence the participant uttered (A) or the second (B)

target_dist whether the instructions asked for heights as distributed in the population (N) or uniformly distributed (U)

label what the participant uttered

unit height unit, either centimetres (cm) or feet and inches (f_in).

value value in cms of the height uttered.

value_in_units value of the height uttered depending on the value of `unit` (either in inches or in centimetres). Used to calculate adjacencies, distances, etc.

starts timestamp of when the utterance starts, in seconds.

delays temporal difference with the start of the previous item (i.e. `starts[index] - starts[index - 1]`)

R whether the item is a repetition of the last

A whether the item is adjacent to the last (after removing repetitions)

TP_full whether the item is a turning point, considering all items (after removing repetitions)

D the Euclidean distance to the previous item (after removing repetitions)

S a measure of how likely the item is in a uniform or gaussian distribution (see text)

expected_* the expectation for measure * derived from reshuffling the participant's sequence 10000 times

Usage

```
castillo2024.rgmomentum.e1
```

Format

An object of class `data.frame` with 5836 rows and 29 columns.

Source

https://osf.io/dw8ez/

References

Castillo L, León-Villagrá P, Chater N, Sanborn AN (2024). "Explaining the Flaws in Human Random Generation as Local Sampling with Momentum." *PLOS Computational Biology*, **20**(1), 1–24. doi:10.1371/journal.pcbi.1011739.

`castillo2024.rgmomentum.e2`

Data from Experiment 2 in Castillo et al. (2024)

Description

Participants first learned a set of syllables arranged in either a single row (one-dimensional condition) or a grid (two-dimensional condition), then produced two random sequences for the same display. These data are licensed under CC BY 4.0, reproduced from materials in OSF.

id participant id

part_Gender participant's gender (self-reported)

part_Age participant's age (self-reported)

index position of the item in the sequence, 0 indexed

id unique identifier for the participant

block whether the item belongs to the first sequence the participant uttered (A) or the second (B)

syll syllable uttered

starts timestamp of when the utterance starts, in seconds.

delays temporal difference with the start of the previous item (i.e. `starts[index] - starts[index - 1]`)

dim whether the participant was allocated to the one-dimensional or two-dimensional condition

seed Which of five possible configurations the participant learned

position The position of the syllable in the array. For 1D arrays, position is left to right. For 2D arrays positions 1-2 correspond to the top 2 cells; 3-5 to the middle 3 cells; and 6-7 to the bottom three cells (always left to right)

R whether the item is a repetition of the last

A whether the item is adjacent to the last in the display (after removing repetitions)

TP_full whether the item is a turning point, considering all items (after removing repetitions)

D the Euclidean distance to the previous item (after removing repetitions)

S a measure of how likely the item is in a uniform or gaussian distribution (see text)

expected_* the expectation for measure * derived from reshuffling the participant's sequence 10000 times

Usage

```
castillo2024.rgmomentum.e2
```

Format

An object of class `data.frame` with 28483 rows and 20 columns.

Source

https://osf.io/dw8ez/

References

Castillo L, León-Villagrá P, Chater N, Sanborn AN (2024). "Explaining the Flaws in Human Random Generation as Local Sampling with Momentum." *PLOS Computational Biology*, **20**(1), 1–24. doi:10.1371/journal.pcbi.1011739.

`spicer2022.anchoringrepulsion.e1`

Data from Experiment 1 in Spicer et al. (2022)

Description

Perceptual judgments. Participants made judgments of numerosity against comparison values or absolute estimates. Comparison values (boundaries) were either similar or dissimilar to the true answer.

Usage

```
spicer2022.anchoringrepulsion.e1
```

Format

An object of class `data.frame` with 9600 rows and 11 columns.

Details

These data are licensed under CC BY 4.0, reproduced from materials in OSF.

Timestamp Date and time of the experimental session

Pt Participant ID

Trial Trial ID based on order of presentation

Boundary Comparison value for that trial

DotCount Number of dots shown on that trial

Region Region for that dot count, being either high or low

Decision Decision made by the participant on whether dot count was higher or lower than the boundary for that trial

Dec_RT Response time for the decision

Accuracy Accuracy of the selected decision

Estimate Direct estimate of the number of dots on that trial made by the participant. NaN is used for trials in which no estimate was requested

Est_RT Response time for the estimate

Source

https://osf.io/95ruy/

References

Spicer J, Zhu J, Chater N, Sanborn AN (2022). "Perceptual and Cognitive Judgments Show Both Anchoring and Repulsion." *Psychological Science*, **33**(9), 1395–1407. doi:10.1177/09567976221089599.

`spicer2022.anchoringrepulsion.e2`
Data from Experiment 2 in Spicer et al. (2022)

Description

Cognitive judgments. Participants answered questions about commonly experienced values. judgments of numerosity against comparison values or absolute estimates. Comparison values (boundaries) were either similar or dissimilar to the true answer.

Usage

```
spicer2022.anchoringrepulsion.e2
```

Format

An object of class `data.frame` with 2960 rows and 13 columns.

Details

These data are licensed under CC BY 4.0, reproduced from materials in OSF.

Timestamp Date and time of the experimental session

Pt Participant ID

Trial Trial ID based on order of presentation

QID ID for the target question of that trial

Question Question text

Region Expected region for that question, being either high or low

Answer Unbiased answer for that question from calibration data

Boundary Comparison value for that trial

Decision Decision made by the participant on whether answer to the question was higher or lower than the boundary

Dec_RT Response time for the decision

Accuracy Accuracy of the selected decision based on calibration data

Estimate Direct estimate of the answer to the question for that trial made by the participant

Est_RT Response time for the estimate

Source

https://osf.io/95ruy/

References

Spicer J, Zhu J, Chater N, Sanborn AN (2022). "Perceptual and Cognitive Judgments Show Both Anchoring and Repulsion." *Psychological Science*, **33**(9), 1395–1407. doi:10.1177/09567976221089599.

`spicer2022.anchoringrepulsion.e2a`

Data from Experiment 2a in Spicer et al. (2022)

Description

Cognitive judgments. Participants answered questions about commonly experienced values. Unlike in Experiment 2, participants viewed each question multiple times, comparing each against both a low (25.5) and high (75.5) comparison value to create 40 trial cases. As in Experiment 1, decisions were requested on all trials, but only 30% of trials were randomly selected to include a direct estimate.

Usage

```
spicer2022.anchoringrepulsion.e2a
```

Format

An object of class `data.frame` with 9920 rows and 13 columns.

Details

This experiment is described in the supplementary materials. These data are licensed under CC BY 4.0, reproduced from materials in OSF.

Timestamp Date and time of the experimental session

Pt Unique ID for that participant

Trial Trial ID based on order of presentation

QID ID for the target question of that trial. Note that these IDs match those of the calibration data.

Question Question text for that trial

Region Expected region for that question, being either high or low

Answer Unbiased answer for that question from calibration data

Boundary Comparison value for that trial

Decision Decision made by the participant on whether answer to the question was higher or lower than the boundary for that trial

Dec_RT Response time for the decision

Accuracy Accuracy of the selected decision based on calibration data

Estimate Direct estimate of the answer to the question for that trial made by the participant. NaN is used for trials in which no estimate was requested

Est_RT Response time for the estimate

Source

https://osf.io/95ruy/

References

Spicer J, Zhu J, Chater N, Sanborn AN (2022). "Perceptual and Cognitive Judgments Show Both Anchoring and Repulsion." *Psychological Science*, **33**(9), 1395–1407. doi:10.1177/09567976221089599.

`sundh2023.meanvariance.e3`

Data from Experiment 3 in Sundh et al. (2023)

Description

Participants made probability judgments of the format: "What is the probability that the weather is [X] on a random day in England?". Various weather events were used, and the queries included both marginal events, conditional events, conjunctions, and disjunctions. The total set of 20 unique queries formed a block within which the presentation order was randomized for each participant. The experiment consisted of three blocks, so that all participants responded to each unique query three times.

Usage

```
sundh2023.meanvariance.e3
```

Format

An object of class `data.frame` with 12420 rows and 10 columns.

Details

These data are licensed under CC BY 4.0, reproduced from materials in OSF.

ID

block 3 blocks in total

trial Trial Number within a block

query, querydetail Verbal descriptions of the query

querytype Type of query: e.g. notBgA = p(¬B|A)

Estimate Estimated probability, in percentages

starttime, endtime

RT

Source

https://osf.io/9kea6/

References

Sundh J, Zhu J, Chater N, Sanborn A (2023). "A Unified Explanation of Variability and Bias in Human Probability Judgments: How Computational Noise Explains the Mean Variance Signature." *Journal of Experimental Psychology: General*, **152**(10), 2842–2860. doi:10.1037/xge0001414.

`sundh2023.meanvariance.e4`

Data from Experiment 4 in Sundh et al. (2023)

Description

Participants made probability judgments about future hypothetical events, of the format: "What is the probability that there will be an early UK general election AND the UK economy will recover this year?". The experiment consisted of three blocks, so that all participants responded to each unique query three times.

Usage

```
sundh2023.meanvariance.e4
```

Format

An object of class `data.frame` with 13320 rows and 7 columns.

Details

These data are licensed under CC BY 4.0, reproduced from materials in OSF.

ID

block 3 blocks in total

query, querydetail Verbal descriptions of the query

querytype Type of query: e.g. not B given A = p(¬B|A)

queryset Whether the query is about biden and 2050 climate goals or UK election and economic recovery

Estimate Estimated probability, in percentages

Source

https://osf.io/9kea6/

References

Sundh J, Zhu J, Chater N, Sanborn A (2023). "A Unified Explanation of Variability and Bias in Human Probability Judgments: How Computational Noise Explains the Mean Variance Signature." *Journal of Experimental Psychology: General*, **152**(10), 2842–2860. doi:10.1037/xge0001414.

`zhu2020.bayesiansampler.e1`

Data from Experiment 1 in Zhu et al. (2020)

Description

Participants made probability judgments of the format: "What is the probability that the weather is [X] on a random day in England?". Various weather events were used, and the queries included both marginal events, conditional events, conjunctions, and disjunctions. The total set of 20 unique queries formed a block within which the presentation order was randomized for each participant. The experiment consisted of three blocks, so that all participants responded to each unique query three times.

Usage

```
zhu2020.bayesiansampler.e1
```

Format

An object of class `data.frame` with 7080 rows and 10 columns.

Details

These data are licensed under CC BY 4.0, reproduced from materials in OSF.

ID

block 3 blocks in total

trial Trial Number within a block

query, querydetail Verbal descriptions of the query

querytype Type of query: e.g. notBgA = p(¬B|A)

Estimate Estimated probability, in percentages

starttime, endtime

RT

Source

https://osf.io/mgcxj/

References

Zhu J, Sanborn AN, Chater N (2020). "The Bayesian Sampler: Generic Bayesian Inference Causes Incoherence in Human Probability Judgments." *Psychological Review*, **127**(5), 719–748. doi:10.1037/rev0000190.

`zhu2020.bayesiansampler.e2`

Data from Experiment 2 in Zhu et al. (2020)

Description

Participants made probability judgments of the format: "What is the probability that the weather is [X] on a random day in England?". Various weather events were used, and the queries included both marginal events, conditional events, conjunctions, and disjunctions. The total set of 20 unique queries formed a block within which the presentation order was randomized for each participant. The experiment consisted of three blocks, so that all participants responded to each unique query three times.

Usage

```
zhu2020.bayesiansampler.e2
```

Format

An object of class `data.frame` with 15120 rows and 10 columns.

Details

These data are licensed under CC BY 4.0, reproduced from materials in OSF.

ID

block 3 blocks in total

trial Trial Number within a block

query, querydetail Verbal descriptions of the query

querytype Type of query: e.g. notBgA = p(¬B|A)

Estimate Estimated probability, in percentages

starttime, endtime

RT

Source

https://osf.io/mgcxj/

References

Zhu J, Sanborn AN, Chater N (2020). "The Bayesian Sampler: Generic Bayesian Inference Causes Incoherence in Human Probability Judgments." *Psychological Review*, **127**(5), 719–748. doi:10.1037/rev0000190.

`zhu2022.coherenceaccuracy.e1`

Data from Experiment 1 in Zhu et al. (2022)

Description

Participants (from Prolific) estimated the frequencies of different 3-card combinations in a 52 card deck and 3-ball combinations in a 52 ball urn (mathematically identical questions). They also answered surveys on poker playing habits and gamblers fallacy questionnaire.

Usage

```
zhu2022.coherenceaccuracy.e1
```

Format

An object of class `data.frame` with 82 rows and 23 columns.

Details

See exact questions in original paper's supplementary materials (Appendix B). These data are licensed under CC BY 4.0, reproduced from materials in OSF.

group Self-reported response on whether they have played poker before

q1-q9 Answers to the poker questions

mq1-mq9 Answers to the ball questions

gfs number of correct answers in gambler's fallacy questionnaire

cs Inferred poker playing time in the last 12 months

RT

taskEqual judged similarity between the Card and Ball task (0=all equal, 1=all differ, 0.5=answers differ but urn and deck were equal)

Source

https://osf.io/cdvkn/

References

Zhu J, Newall PW, Sundh J, Chater N, Sanborn AN (2022). "Clarifying the Relationship between Coherence and Accuracy in Probability Judgments." *Cognition*, **223**, 105022. doi:10.1016/j.cognition.2022.105022.

`zhu2022.coherenceaccuracy.e2`

Data from Experiment 2 in Zhu et al. (2022)

Description

Participants (professional players recruited from twoplustwo.com) estimated the frequencies of different 3-card combinations in a 52 card deck and 3-ball combinations in a 52 ball urn (mathematically identical questions). They also answered surveys on poker playing habits and gamblers fallacy questionnaire.

Usage

```
zhu2022.coherenceaccuracy.e2
```

Format

An object of class `data.frame` with 186 rows and 23 columns.

Details

See exact questions in original paper's supplementary materials (Appendix B). These data are licensed under CC BY 4.0, reproduced from materials in OSF.

group value here is always professional (in contrast to Experiment 1)

q1-q9 Answers to the poker questions

mq1-mq9 Answers to the ball questions

gfs number of correct answers in gambler's fallacy questionnaire

cs Inferred poker playing time in the last 12 months

RT

taskEqual judged similarity between the Card and Ball task (0=all equal, 1=all differ, 0.5=answers differ but urn and deck were equal)

Source

https://osf.io/cdvkn/

References

Zhu J, Newall PW, Sundh J, Chater N, Sanborn AN (2022). "Clarifying the Relationship between Coherence and Accuracy in Probability Judgments." *Cognition*, **223**, 105022. doi:10.1016/j.cognition.2022.105022.

`zhu2022.structurenoise.animals`
Data from Animal Experiment in Zhu et al. (2022)

Description

Participants were asked to type animal names as they came to mind and were explicitly instructed that they could resubmit previous animals, though not consecutively.

Usage

```
zhu2022.structurenoise.animals
```

Format

An object of class `data.frame` with 4967 rows and 7 columns.

Details

These data are licensed under CC BY 4.0, reproduced from materials in OSF.

ID Participant ID

Order Index of the response

Responses Transcribed response

Animal Category the response was allocated to

StartType,EndType Absolute time of starting and ending to type the response

IRI Time between last response's EndType and this response's StartType

Source

https://osf.io/kcfgp/

References

Zhu J, León-Villagrá P, Chater N, Sanborn AN (2022). "Understanding the Structure of Cognitive Noise." *PLoS Computational Biology*, **18**(8), e1010312. doi:10.1371/journal.pcbi.1010312.

`zhu2022.structurenoise.time`

Data from Time Experiment in Zhu et al. (2022)

Description

Participants first listened to a sample of the target temporal interval for 60 seconds. Participants were instructed to reproduce the target by pressing the spacebar when they believed the target interval had elapsed (i.e. perfect performance in the task would mean `IRI == Target`).

Usage

```
zhu2022.structurenoise.time
```

Format

An object of class `data.frame` with 29822 rows and 6 columns.

Details

These data are licensed under CC BY 4.0, reproduced from materials in OSF.

ID Participant ID

Order Index of the response

StartType,EndType Absolute time of starting and ending to type the response

IRI Time between last response's EndType and this response's StartType

Target Whether the participant had to reproduce a 1/3s, 1s or 3s interval

Source

https://osf.io/kcfgp/

References

Zhu J, León-Villagrá P, Chater N, Sanborn AN (2022). "Understanding the Structure of Cognitive Noise." *PLoS Computational Biology*, **18**(8), e1010312. doi:10.1371/journal.pcbi.1010312.

Index

∗ **datasets**
castillo2024.rgmomentum.e1, 2
castillo2024.rgmomentum.e2, 3
spicer2022.anchoringrepulsion.e1, 5
spicer2022.anchoringrepulsion.e2, 6
spicer2022.anchoringrepulsion.e2a, 7
sundh2023.meanvariance.e3, 8
sundh2023.meanvariance.e4, 9
zhu2020.bayesiansampler.e1, 10
zhu2020.bayesiansampler.e2, 11
zhu2022.coherenceaccuracy.e1, 12
zhu2022.coherenceaccuracy.e2, 13
zhu2022.structurenoise.animals, 14
zhu2022.structurenoise.time, 15

castillo2024.rgmomentum.e1, 2
castillo2024.rgmomentum.e2, 3

spicer2022.anchoringrepulsion.e1, 5
spicer2022.anchoringrepulsion.e2, 6
spicer2022.anchoringrepulsion.e2a, 7
sundh2023.meanvariance.e3, 8
sundh2023.meanvariance.e4, 9

zhu2020.bayesiansampler.e1, 10
zhu2020.bayesiansampler.e2, 11
zhu2022.coherenceaccuracy.e1, 12
zhu2022.coherenceaccuracy.e2, 13
zhu2022.structurenoise.animals, 14
zhu2022.structurenoise.time, 15