

Data sets

I've given the source of each data set (not in Harvard style!). All except Set 8 are real published data. Set 8 is a simulation based on real historical means and standard deviations.

Set 1: Blood pressure baselines

The table below provides base line measurements of the supine systolic blood pressure (S) and supine diastolic blood pressure (D) for 19 diabetic patients. This data was collected *before* the patients took part in a drug trial.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
S	124	120	115	134	131	119	124	127	113	132	129	124	112	124	144	134	119	123	122
D	78	70	68	91	93	71	76	82	72	88	73	76	70	80	88	80	77	58	79

Part way through the trial, the same 19 patients had their blood pressures re-measured after all traces of the drug had left their system. This formed a *second* base line measurement.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
S	120	138	111	123	111	113	108	121	107	122	139	127	112	128	156	122	115	114	124
D	68	70	71	83	72	64	76	65	57	78	80	65	72	83	99	81	68	61	76

Source: Favorite Datasets from Early Phases of Drug Research, Bradstreet, T.E,
<http://www.math.iup.edu/~tshort/Bradstreet/part3/part3-table3.html>

Accessed 6th Feb 2010

Set 2: Giant Redwood Trees

Ecologists keep records on the largest Redwood trees in North American forests. Below is a random sample of 100 Breast Height Diameters from a census of large Redwoods. The measurements are in centimetres.

357 421 381 302 411 405 421 332 518 266
390 427 399 329 384 381 487 411 381 402
329 290 268 366 317 287 411 454 524 293
351 296 484 329 402 256 456 381 393 414
338 561 396 402 290 451 456 411 271 402
293 484 381 335 347 338 381 366 472 321
305 442 494 321 296 364 445 405 395 357
381 335 321 335 439 381 747 488 274 351
357 299 417 1530 341 378 408 683 357 579
277 366 381 445 473 426 369 390 503 357

Source: <http://www.landmarktrees.net/redwoods.html>

Accessed: 6th Feb 2010

Set 3: Stroop Effect

“Participants were invited to sit at a desk in a quiet, well illuminated room and were asked to read a set of instructions. They were asked if they fully understood these and a practice list of 24 words was then placed, face down, on the table in front of them. This list was prepared using words from both the test sheets. When ready, the participants turned the paper over and proceeded to name the colours of the ink for each word in order. A timer was used to measure their performance, and a record was kept of the accuracy of their replies. They were given 30 seconds rest before the main test. This consisted of the first list, followed by a 30 second rest, followed by the second list, all in the same manner as before. The order of presentation of the lists was randomised”

The table below has the time taken in seconds to read all of the 24 cards with neutral words (N) and the time taken to read the 24 cards with colour words (C) for each of the 16 subjects. I've missed out the data on errors, but that can be found in the source.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
N	14	18	13	17	16	15	25	21	16	18	16	19	22	25	16	19
C	17	23	16	21	29	19	34	24	20	21	22	22	27	28	18	26

Source: *The Stroop effect and selective attention: intrusion of automatic semantic processing on the performance of a colour identification task.*

URL: <http://www.magenta.u-net.com/open/stroop.htm>

Accessed: 6th Feb 2010

Set 4: Heights of 100 pregnant women

Below is a randomly chosen sample of the heights in centimetres of 100 pregnant women. The sample was drawn from a list of 1794 heights compiled as part of a larger survey by Janet Peacock and used in the textbook *An Introduction to Medical Statistics* by M Bland. This is real data and the heights were measured in metric units directly under normal hospital conditions.

159 173 164 158 163 162 163 156 160 175
161 155 157 152 166 165 167 162 158 169
151 156 158 158 171 156 163 160 159 177
159 159 161 165 157 165 161 157 168 170
156 161 178 168 156 163 168 165 160 164
154 168 161 161 162 165 167 153 157 163
150 164 168 153 167 158 158 160 154 160
161 163 163 164 163 157 165 167 155 165
160 156 170 161 150 156 160 152 147 177
163 162 163 167 163 162 155 159 170 160

Source: <http://www-users.york.ac.uk/~mb55/datasets/datasets.htm#intro>

Accessed: 6th Feb 2010

Set 5: Frequency distribution of heights of 1794 pregnant women

This frequency distribution was compiled from the full data set of Janet Peacock's quoted in Set 4 above.

Height interval	Midpoint	Frequency	fx	Cumulative frequency
$135 \leq h < 140$		0		
$140 \leq h < 145$		2		
$145 \leq h < 150$		42		
$150 \leq h < 155$		183		
$155 \leq h < 160$		455		
$160 \leq h < 165$		579		
$165 \leq h < 170$		346		
$170 \leq h < 175$		144		
$175 \leq h < 180$		40		
$180 \leq h < 185$		3		
$185 \leq h < 190$		0		

Source: <http://www-users.york.ac.uk/~mb55/datasets/datasets.htm#intro>

Accessed: 6th Feb 2010

Set 6: Height and Peak Expiratory Flow Rate (PEFR) for female medical students

The height (H in cm) and PEFR (litres/min) for 15 female medical students has been taken from a larger data set and reformatted into the table below...

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
H	181	168	163	171	177	169	161	170	158	161	163	155	171	172	168
PEFR	522	440	428	537	513	510	383	455	440	461	370	503	430	442	595

Source: <http://www-users.york.ac.uk/~mb55/datasets/datasets.htm#intro>

Accessed: 6th Feb 2010

Set 7: Age of Mothers of Patients with Down's Syndrome for Births in Australia from 1942 to 1952

Age interval of mother	Frequency of births	Frequency of Down's
Less than 20 years	35 555	15
20 to 24 years	207 931	128
25 to 29 years	253 450	208
30 to 34 years	170 970	194
35 to 39 years	86 046	297
40 to 44 years	24 498	240
45 or over	1 707	37

Source: <http://lib.stat.cmu.edu/datasets/Andrews/>

Accessed: 6th Feb 2010

Note: Published in *Data: a Collection of Problems from many Fields for the Student and Research Worker*, D.F. Andrews and A.M. Herzberg, Springer, 1985, but the data set above was adapted from the Web page given.

Set 8: Heights of Dutch Army conscripts in 1890 and 1940

Below is a sample of the heights (cm) of 100 Dutch army conscripts given medicals in 1890.

173 169 168 172 172 162 171 158 168 163 170 167 153 167 173 168 172 174
 161 158 171 167 155 172 145 164 167 169 156 160 160 163 175 164 161 167
 176 162 174 174 167 172 167 175 156 156 179 158 164 152 151 167 167 165
 174 165 167 167 164 145 156 176 169 173 174 155 181 179 166 166 171 169
 163 165 161 177 163 168 174 174 173 170 171 167 161 166 159 163 174 172
 177 174 165 157 162 160 173 179 176 162

Below is a sample of the heights (cm) of 100 Dutch army conscripts given medicals in 1940.

179 170 176 177 187 163 190 172 185 178 177 170 158 171 175 160 192 184
 171 180 176 175 168 177 160 184 167 177 172 166 164 184 178 176 170 179
 174 171 174 171 173 170 174 168 182 170 168 168 181 177 177 167 186 172
 178 175 174 179 159 173 169 164 173 176 169 177 168 177 165 178 175 169
 177 163 160 188 168 176 168 150 177 176 172 181 175 159 169 181 186 170
 162 182 181 183 186 177 179 168 182 178

Source: Based on a time series of median heights in Paradoxes of Modernization and Material Well-Being in the Netherlands during the Nineteenth Century, Drukker and Tassenaar, Chapter 9 of *Health and Welfare during Industrialization*, edited by Steckel and Floud, University of Chicago Press, 1997. Normal samples produced in R using commands `rnorm(100, 167, 7.5)` and `rnorm(100, 173, 7.5)`.

Set 9: Heights of Dutch army recruits: changes over time

Below are the estimated *median* heights (H) for conscripts to the Dutch army each decade for a period spanning 120 years. Heights are in millimetres.

Year	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940
H	1647	1665	1653	1642	1640	1648	1657	1668	1678	1688	1699	1720	1734

Source: Based on a time series of median heights in Paradoxes of Modernization and Material Well-Being in the Netherlands during the Nineteenth Century, Drukker and Tassenaar, Chapter 9 of *Health and Welfare during Industrialization*, edited by Steckel and Floud, University of Chicago Press, 1997, table 9A.2, no pagination.

Set 10: Height at Forced expiratory volume in 1 second (FEV1) for 20 medical students

The table below contains the heights (cm) and Forced expiratory volume in 1 second (FEV1) (litres) for 20 male medical students.

Height (cm)	FEV1 (li)
174.0	4.32
180.7	4.80
183.7	4.68
177.0	5.43
177.0	3.09
172.0	3.78
176.0	3.75
177.0	4.05
164.0	3.54
178.0	2.98
167.0	3.54
171.2	3.42
177.4	3.60
171.3	3.20
183.6	4.56
183.1	4.78
172.0	3.60
181.0	3.96
170.4	3.19
171.2	2.85

Source: <http://www-users.york.ac.uk/~mb55/datasets/datasets.htm#intro>

Accessed: 7th Feb 2010

FEV1 is explained well at <http://www.spirxpert.com/indices7.htm>