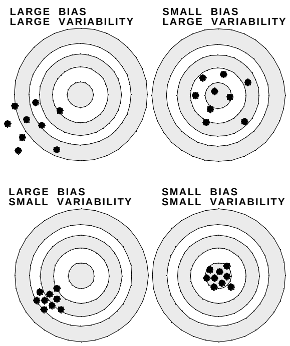


Being approximately correct and being precisely wrong

1. Refer to the descriptions of the SMOG index, the Fry method, the Flesch Reading Ease, and the Flesch-Kincaid Grade Level, for measuring *readability* (under Resources for Measurement/Surveys).[1]

   For the article or text you have chosen (as per discussion in class), randomly select three separate 100 word passages, and use this *set of three passages* to measure the readability (F_1) using the Fry graph. Rather than do so manually, you can use the SMOG calculator to determine the average number of sentences and syllables per hundred words. Repeat the readability measurement (F_2) with a second *different* set of three passages. Repeat once more (F_3), using a *third set*.

   Using these same three sets, calculate the SMOG index, the Flesch Reading Ease, and the Flesch-Kincaid Grade Level.

   For each index, use the 3 estimates to calculate the standard error of measurement, and the coefficient of variation. Comment.

2. Propose a method to assess the *validity* of a readability index.

3. **[m-s]** Derive the link between the standard error of measurement and the (intraclass correlation) reliability coefficient [last line, column 1, p. 7 of notes on "Quantifying Reliability" in Notes on Psychometrics for students in rehabilation sciences in Resources for Measurement/Surveys. *Hint: it's simply a matter of using the definition of R.*

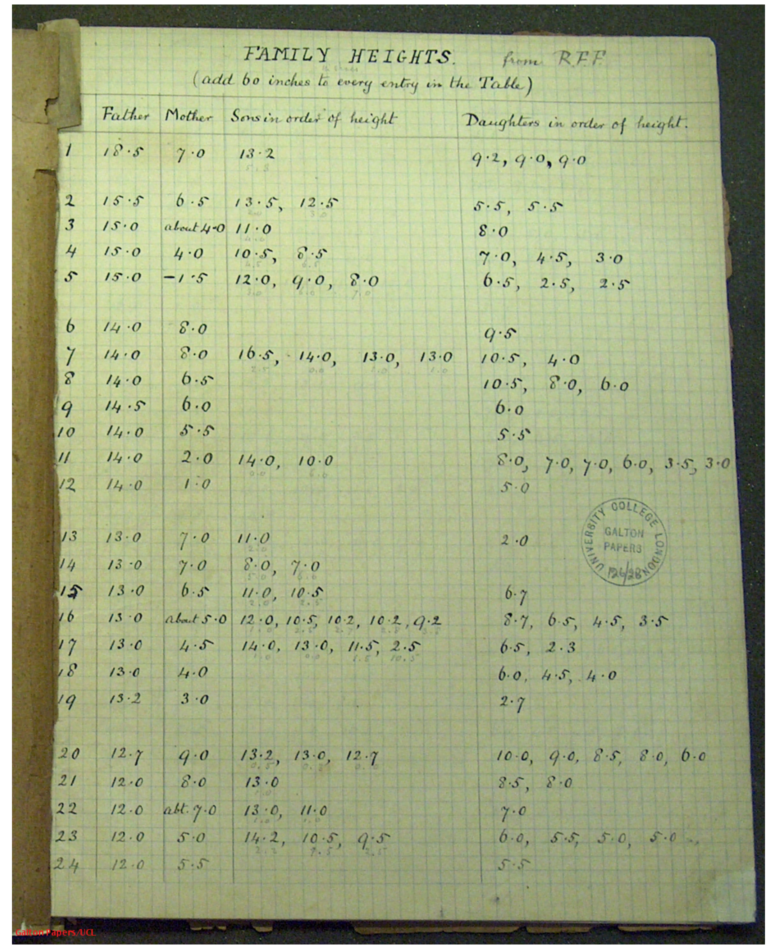
4. **[m-s]** Exercise in section 3: Relationship between test-retest correlation and ICC(X) [In notes on Effect of Errors in X and Y on measured correlation and slope]

[1]ToneCheck (`http://tonecheck.com/`) is another interesting tool. See story at http://www.montrealgazette.com/search/search.html?q=ToneCheck

5. **[m-s]** Exercise section 4: Relationship between correlation(X, X') and ICC(X) [ibid.]

6. Francis Galton (1822-1911) found that the correlation between (*self-reported*) parental and (adult) offspring heights was strongest for the one between father and son [0.396 ± 0.024], and weakest for the one between mother and daughter [0.284 ± 0.028]. Given the way he obtained the measurements, can you imagine why this was?

   [It was 0.302±0.027 for mother & son; 0.360±0.026 for father & daughter.]



FAMILY HEIGHTS. from R.F.F.
(add 60 inches to every entry in the Table)

	Father	Mother	Sons in order of height	Daughters in order of height.
1	18·5	7·0	13·2	9·2, 9·0, 9·0
2	15·5	6·5	13·5, 12·5	5·5, 5·5
3	15·0	about 4·0	11·0	8·0
4	15·0	4·0	10·5, 8·5	7·0, 4·5, 3·0
5	15·0	−1·5	12·0, 9·0, 8·0	6·5, 2·5, 2·5
6	14·0	8·0		9·5
7	14·0	8·0	16·5, 14·0, 13·0, 13·0	10·5, 4·0
8	14·0	6·5		10·5, 8·0, 6·0
9	14·5	6·0		6·0
10	14·0	5·5		5·5
11	14·0	2·0	14·0, 10·0	8·0, 7·0, 7·0, 6·0, 3·5, 3·0
12	14·0	1·0		5·0
13	13·0	7·0	11·0	2·0
14	13·0	7·0	8·0, 7·0	
15	13·0	6·5	11·0, 10·5	6·7
16	13·0	about 5·0	12·0, 10·5, 10·2, 10·2, 9·2	8·7, 6·5, 4·5, 3·5
17	13·0	4·5	14·0, 13·0, 11·5, 2·5	6·5, 2·3
18	13·0	4·0		6·0, 4·5, 4·0
19	13·2	3·0		2·7
20	12·7	9·0	13·2, 13·0, 12·7	10·0, 9·0, 8·5, 8·0, 6·0
21	12·0	8·0	13·0	8·5, 8·0
22	12·0	abt. 7·0	13·0, 11·0	7·0
23	12·0	5·0	14·2, 10·5, 9·5	6·0, 5·5, 5·0, 5·0
24	12·0	5·5		5·5

UNIVERSITY COLLEGE LONDON GALTON PAPERS 126/28

Galton Papers/UCL

Family heights: Page 1/8 of notebook in Galton Papers : see "Galton's family data on human stature" on JH's website

7. *Bridging the physical- and the psycho-metric:* The notes on "Increasing Reliability by averaging several measurements" on the right hand column of page 4 of JH's notes on Quantifying Reliability give the formula for the so-called "Stepped-Up Reliability". In psychometrics (where the number of items on a test serves as the "several measurements") this formula serves as the basis for the "Spearman-Brown prediction formula".[2]

   [m-s] Invert the formula on p.4 to derive the one on the right hand column of p.1 for Spearman-Brown prediction formula relating the reliability of two versions of a test, one with N times more items than the other.

8. You are trying to estimate, from imperfect observations of F and C, the values of the two coefficients B_0 and B_1 in the temperature relation $F = B_0 + B_1 \times C$.

   For each of the following situations, and using the true values of $B_0 = 32$ and $B_1 = 9/5 = 1.8$, simulate[3] 1000 datasets & investigate the behaviour of the 1000 estimates, b_0 and b_1, of B_0 and B_1. In each simulation, use samples of size $n = 4$, with temperatures of $C = 14$, 16, 18 and 20.

   (a) C measured perfectly, F measured with $\epsilon_F \sim$ Gaussian($\mu = 0, \sigma_{\epsilon_F} = 1$) errors that are independent of F. Check – formally, using a test (or CI) based on the mean of the 1000 estimates – for evidence of bias in b_1. Also check whether the empirical variance of b_1 agrees with that given by the theoretical formula, namely

   $$Var(b_1) = \sigma^2_{\epsilon_F} / \sum (x - \bar{x})^2.$$

   (b) F measured perfectly, C measured with $\epsilon_C \sim$ Gaussian($\mu = 0, \sigma_{\epsilon_C} = 1$) errors that are independent of C [*Classical type* error: someone else chose situations when C was indeed exactly 14, 16, etc, but didn't tell you what C was, and instead asked you to independently record C using your own imperfect instrument, and to use *your* recordings of C in your estimation of the equation]. Again, formally test for evidence of bias in b_1.

   Do your findings line up with the predictions in the Notes? If the patterns are difficult to see, you might change the number of simulations, the sizes of the errors, the range of C or the sample size.[4]

[2] `http://en.wikipedia.org/wiki/Spearman-Brown_prediction_formula` .

[3] If new to simulations, see "Computer code to simulate datasets with measurement error" at the bottom of the Resources webpage for measurement/surveys. It gives some 'starter' computer code, which you can modify to suit.

[4] The article by Hutcheon et al. "Random measurement error and regression dilution bias", under 'r e p r i n t s' on JH's home page, tries to explain these patterns intuitively.

9. Before we study how well we can digitize survival curves, here is an exercise on communicating what the curves are meant to convey and the context in which they were generated.

   Refer to the article "Associations between C-reactive protein, coronary artery calcium, and cardiovascular events: implications for the JUPITER population from MESA, a population-based cohort study", available in the Resources link opposite 'Applications' in `bios601`. We digitized the lowermost (green) curve in Figure 2A of that article.

   (a) Read the Abstract and study the Figures in the article. Then, write, *in your own words,* a short news item of 250 words or so (2-3 minutes or so on radio) for your local newspaper and radio station, where you moonlight as a health reporter. In your piece address (i) the rationale for the study (ii) the principal findings and (iii) the implications of these findings. Also suggest a headline for your story. [*You might want to study some health reports to see how they are structured.. the order may not be the (i)-(iii) order listed above.* An interesting but slightly more highbrow website devoted to science reporting in general is `http://www.sciencedaily.com/`.
   The websites
   ... `http://www.cnn.com/HEALTH/`,
   ... `http://www.nytimes.com/pages/health/index.html`,
   ... `http://www.bbc.co.uk/news/health/` and
   ... `http://www.cbc.ca/news/health/`
   are also worth consulting, and indeed monitoring.]

   (b) A 65-year old relative of yours reads your story, looks on the internet and finds that a test that measures coronary artery calcium is available in a private clinic in Montreal, and phones you to ask if it would be worth being tested and getting her "score". What would you say to this relative?

10. Errors in digitization

Refer to the duplicate readings you made of the Kaplan-Meier survival curve in the study entitled “Associations between C-reactive protein, coronary artery calcium, and cardiovascular events: implications for the JUPITER population from MESA, a population-based cohort study” available in the Resources link opposite ‘Applications’ in `bios601`

For now, ignore the point-wise measures of precision, i.e., the standard errors and confidence intervals, that often accompany such curves. These are (decreasing) functions of the numbers of subjects and the numbers of ‘events’; we will cover their calculation later in the term. For now, focus only the loss of precision as a result of your digitization.

Focus on your two measurements of each of the reported y-year risks, where y= 1, 2, 3, 4, 5, 6, 7:

$$\text{y-year CHD risk} = 100 \times (1 - \text{ proportion free of CHD at year y})\%$$

(a) From your two measurements at each of the 7 timepoints, obtain a 7d.f. estimate of the ‘standard error of measurement’. Do so using a ‘canned’ statistical routine and also ‘from scratch’ in `R`

Write out the statistical model that you used to obtain this, and list any assumptions it makes.

(b) The estimate in (a) is an estimate of the ‘within’ observer variation.

In order to estimate the ‘*between*’-observer variation, what is the *minimal* information you would need from each of you co-observers? (since JH has access to all of them, he will supply each of them once you email him with your specific request: he can supply the full raw data that could be then put into a canned statistical routine, but he would prefer that you do the calculations ‘from scratch’ in `R`).

Again, write out the statistical model that you used to obtain this, and list any assumptions it makes.

(c) Here the ‘objects’ to be measured were 7 very specific (fixed) time-points. Assume for the sake of this exercise that the 7 objects were 7 randomly selected human subjects and that we were interested in calculating an intra-class correlation coefficient to serve as a reliability measure. Carry out the ICC calculation. Restrict you attention to years 1-5 and recalculate the new ICC. Comment on why the ICC becomes smaller.

XXI. *Experiments to determine the Density of the Earth. By* Henry Cavendish, *Esq. F.R.S. and A.S.*

Read June 21, 1798.

MANY years ago, the late Rev. JOHN MICHELL, of this Society, contrived a method of determining the density of the earth, by rendering sensible the attraction of small quantities of matter; but, as he was engaged in other pursuits, he did not complete the apparatus till a short time before his death, and did not live to make any experiments with it. After his death, the apparatus came to the Rev. FRANCIS JOHN HYDE WOLLASTON, Jacksonian Professor at Cambridge, who, not having conveniences for making experiments with it, in the manner he could wish, was so good as to give it to me.

The following Table contains the Result of the Experiments.

Exper.	Mot. weight	Mot. arm	Do. corr.	Time vib.	Do. corr.	Density.
1	m. to +	14,32	13,42	′ ″	-	5,5
	+ to m.	14,1	13,17	14,55	-	5,61
2	m. to +	15,87	14,69	-	-	4,88
	+ to m.	15,45	14,14	14,42	-	5,07
3	+ to m.	15,22	13,56	14,39	-	5,26
	m. to +	14,5	13,28	14,54	-	5,55
4	m. to +	3,1	2,95		6,54	5,36
	+ to −	6,18	-	7,1	-	5,29
	− to +	5,92	-	7,3	-	5,58
5	+ to −	5,9	-	7,5	-	5,65
	− to +	5,98	-	7,5	-	5,57
6	m. to −	3,03	2,9	-	-	5,53
	− to +	5,9	5,71		-	5,62
7	m. to −	3,15	3,03	7,4 by mean.	6,57	5,29
	− to +	6,1	5,9			5,44
8	m. to −	3,13	3,00		-	5,34
	− to +	5,72	5,54	-	-	5,79
9	+ to −	6,32	-	6,58	-	5,1
10	+ to −	6,15	-	6,59	-	5,27
11	+ to −	6,07	-	7,1	-	5,39
12	− to +	6,09	-	7,3	-	5,42
13	− to +	6,12	-	7,6	-	5,47
	+ to −	5,97	-	7,7	-	5,63
14	− to +	6,27	-	7,6	-	5,34
	+ to −	6,13	-	7,6	-	5,46
15	− to +	6,34	-	7,7	-	5,3
16	− to +	6,1	-	7,16	-	5,75
17	− to +	5,78	-	7,2	-	5,68
	+ to −	5,64	-	7,3	-	5,85

`http://en.wikipedia.org/wiki/Cavendish_experiment`: Cavendish found that the Earth’s density was 5.448 ± 0.033 times that of water (due to a simple arithmetic error, found in 1821, the erroneous value 5.48 ± 0.038 appears in his paper).