

Bland-Altman plot

Measuring agreement between methods

(compiled by Eric Fèvre, with apologies for
“physical” absence)

Example with application

- The ability of smallholder farmers and animal health workers to estimate live bodyweight can critically affect the likelihood of under- or over-dosing of veterinary compounds.
- Decentralised systems where farmers administer a significant proportion of veterinary treatments.
- Project run by Noreen Machilla, formerly of Edinburgh University, based in Kenya at the time.
 - Machila N, Fèvre EM, et al (in preparation). Farmer estimation of live bodyweight of cattle: implications for veterinary drug dosing in East Africa. *Preventive Veterinary Medicine*.

The data (1)

- Survey of three hundred and twenty four (324) cattle owned by 170 farmers was conducted in Busia District, Kenya.
- Cattle were weighed on
 - a standard calibrated digital scale
 - owners were asked to estimate the live weight of their animals
 - estimated by a weigh-band
 - estimated by local animal health workers.
- How well do these methods agree?



WeighBand	Scale
101	42
165	78
134	64
152	73
225	110
117	58
201	103
165	86
122	64
106	56
277	149
217	119
286	161
209	118
159	90
209	119
286	163
194	111
194	111
225	129
194	111.5
106	61
194	112
277	160
51	29.5
128	75
277	163
209	124
106	63
	Etc...

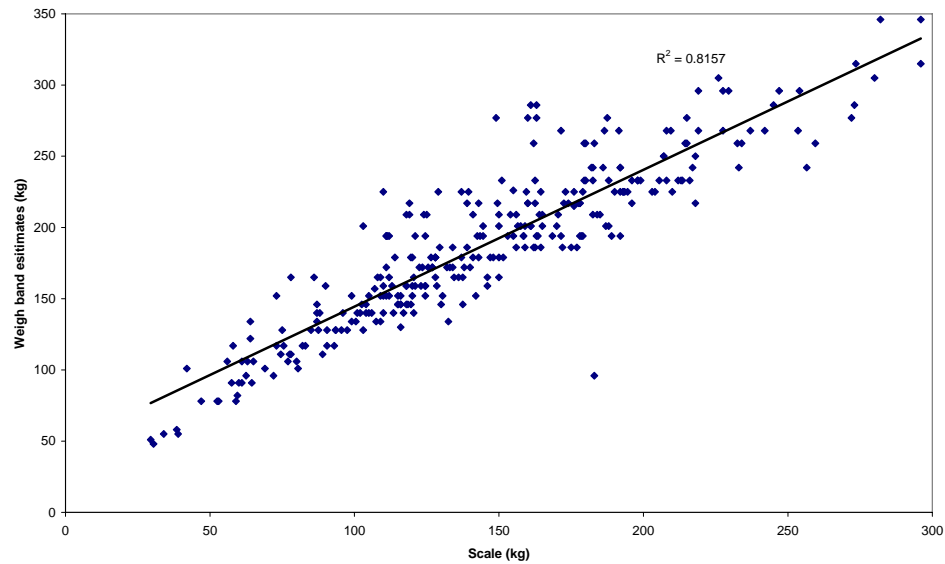
The data (2)

324 measurements on non-pregnant cattle of all ages

Let's use the weigh band and the scale data...

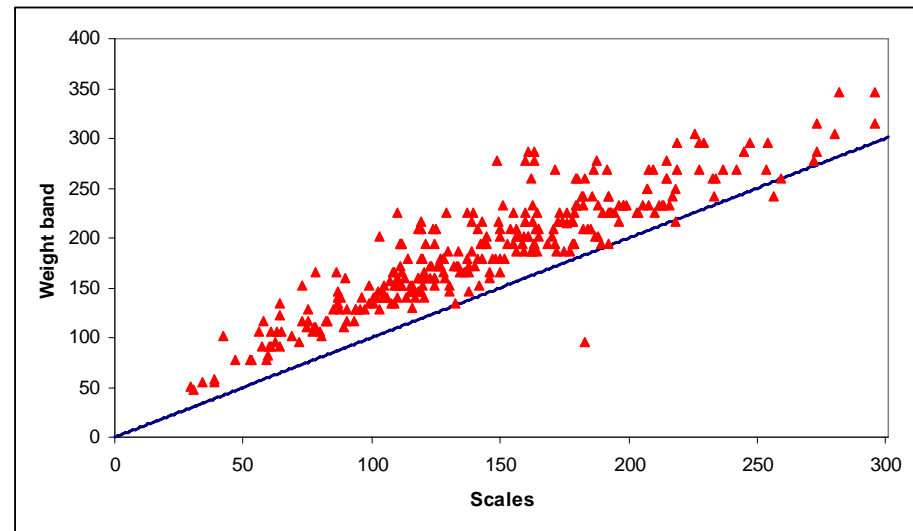
How do the weigh band and the digital scale compare?

- $R =$ Pearson correlation of Weigh-Band and Scale = 0.903 $p = 0.000$.
- The measurements are closely related.
- But do they agree?
- Perfect correlation if the points lie perfectly along any straight line. But perfect AGREEMENT only if they lie perfectly along the line of equality.



Regression line – illustrates correlation...

Line of equality – illustrates agreement – hmm, not quite the same!



Bland and Altman say:

Bland MJ, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327: 307-310.

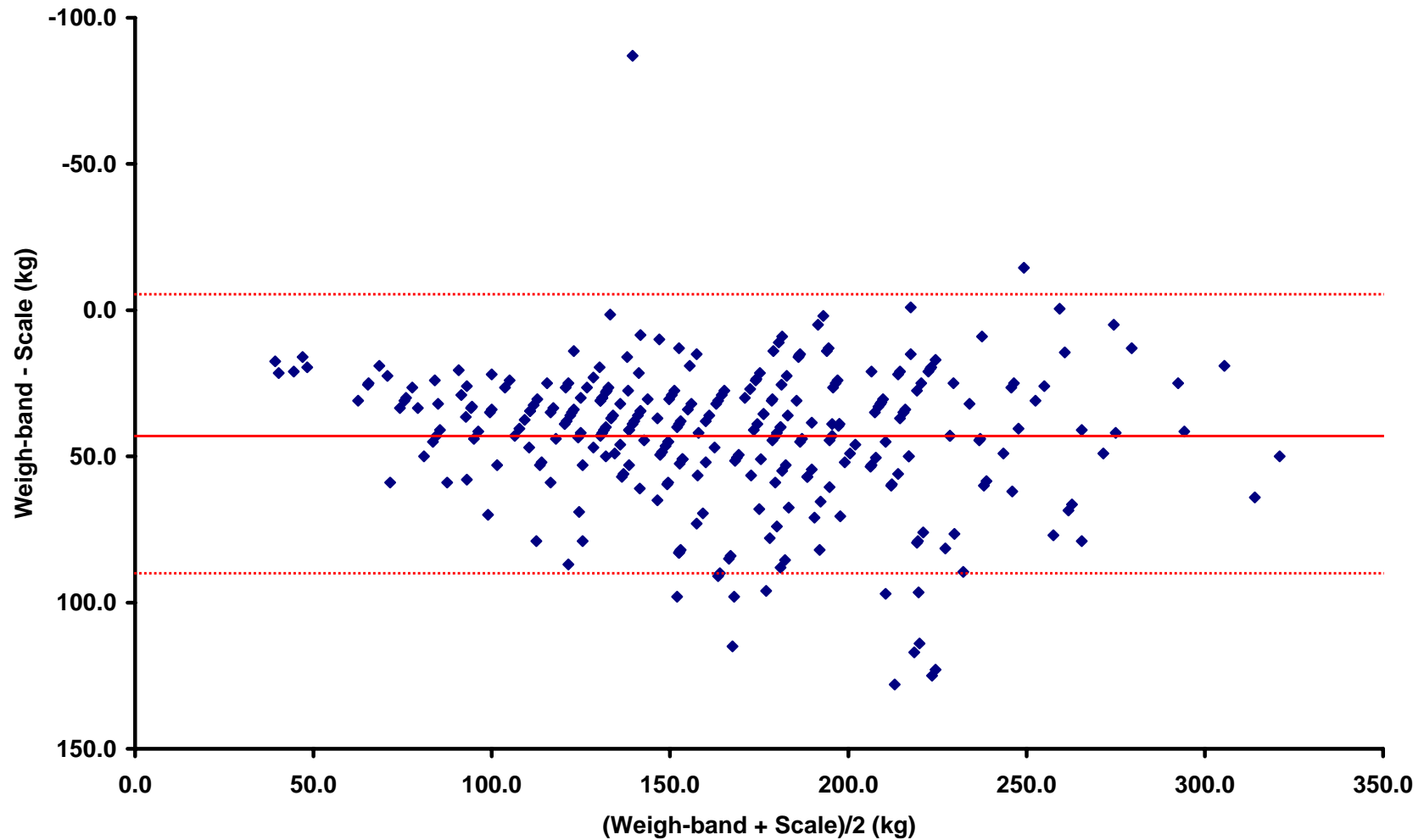
- A change in scale of measurement does not affect the correlation, but it affects the agreement.
- Correlation depends on the range of the true quantity in the sample. If this is wide, the correlation will be greater than if it is narrow.
- **The test of significance may show that the two methods are related, but it would be amazing if two methods designed to measure the same quantity were not related. The test of significance is irrelevant to the question of agreement.**
- Data which seem to be in poor agreement can produce quite high correlations.

Bland and Altman say:

- How far apart measurements can be without causing difficulties will be a question of judgment. Ideally, it should be defined in advance to help in the interpretation of the method comparison and to choose the sample size.

A Bland-Altman plot...

- Aka a “Tukey mean-difference plot”
- Two variables (eg data from 2 tests)
- X-axis: mean of the two measurements
- Y-axis: difference between the two values
 - What is the difference between the two methods vs the best measure of the “true value” we have, which is the mean of the two methods.



- 3 lines: Mean, mean +2SD and mean -2SD = limits of agreement (here, quite wide ~ 100kg, with many outside this range).

Interpretation (1)

- Decide in advance what error is acceptable: here, as error leads to wrong drug dosing, we choose +/- 20% of true weight to be acceptable.
- Weigh-band overestimated the true weight of 98.7% of cattle by an average of 35.3% (range: 1.0%-140.5%; median: 30.5%).
- Only 25.8% (78/302) of cattle were weighed to within +/- 20% of the true weight by the weigh-band.
- Limits of agreement of +91.3 kg and -6.0 kg (see plot)

Interpretation (2)

- The weigh-band used over-estimated weight.
- Designed and calibrated on European breeds (*Bos taurus*) of cattle.
- Stockier and coats thicker than that of small East African zebu (*Bos indicus*) breeds.
- Relying on the weigh-band for estimating weight for dosing = many cattle receiving an overdose.
- Overuse or frequent treatment of cattle can increase the risk of drug toxicity, especially when animals are in poor body condition.

Finally...

- Farmers consistently under-estimated weight (the weights they know relate to carcass weight, not live body weight, as this is what they get for their cash) ~ under-dosing.
- Animal health assistants started by under-estimating weight, but got better with practice.