Errata for The Foundations of Statistics: A Simulation-based Approach

Shravan Vasishth

July 16, 2011

Here is a list of errata for the book *The Foundations of Statistics: A Simulation-based Approach*. We are grateful to Matt Goldrick, students in Potsdam, Wolfgang Schwarz, and Christian Robert for corrections. If you find any other errors, please let me know (vasishth@uni-potsdam.de). We hope to bring out a corrected edition soon.

For each corrected item, I credit the person who found the error (MG: Matt Goldrick; CR: Christian Robert; WS: Wolfgang Schwarz; S: groups of students).

1. (**WS**): We should have mentioned in the book that the entire discussion in chapter 3 ignores the properties of the Cauchy distribution. This is especially relevant for the Central Limit Theorem.

2. (**MG**) p. 13 the function multiplot is used before it's defined (p. 15) in the book. It's defined in the vb.R code on the book's web page. You should source this code before running the code step by step. The vb.R file is here:

   `http://www.ling.uni-potsdam.de/~vasishth/Misc/vb.R`

3. (**S**): p. 22, line 20.

   "This shows us how the total probability is distributed among all the possible results of an experiment"

   Period missing after experiment.

4. (**MG**): p. 32 figure 2.10 left panel should be titled "sample size 40".

5. (**CR**): p. 52 and elsewhere, the statement in the book "s is an unbiased estimator of σ" is incorrect. It should be:

   "s is a standard estimator of σ"

6. (**MG**) pooled variance is mentioned on p. 92 but not defined. See here:

   `http://en.wikipedia.org/wiki/Pooled_variance`

7. (**CR**) On p. 75 we say:

> In one-sample situations our null hypothesis is that there is no difference between the sample mean and the population mean:
>
> $$H_0 : \bar{x} = \mu \tag{1}$$

The above equation doesn't make any sense. Our null hypothesis is that the population mean is some point value (for example, zero).

8. (**MG**) Page 92 onwards: In Chapter, formulas for the standard error in the equivalence tests section are wrong. Equations 4.3 - 4.6 should all have a multiplication, not division, sign in the denominator:

Equation 4.3 (corrected):

$$t = \frac{d - \Theta}{SE} = \frac{d - \Theta}{s_{\text{pooled}} \times \sqrt{(1/n_1 + 1/n_2)}} = -2.616 \tag{2}$$

Equation 4.4 (corrected):

$$t = \frac{d + \Theta}{SE} = \frac{d + \Theta}{s_{\text{pooled}} \times \sqrt{(1/n_1 + 1/n_2)}} = 5.384 \tag{3}$$

Equations 4.5 and 4.6 (corrected):

$$t_{d \leq \Theta_L} = \frac{d - \Theta}{s_{\text{pooled}} \times \sqrt{(1/n_1 + 1/n_2)}} \tag{4}$$

$$t_{d \geq \Theta_U} = \frac{d + \Theta}{s_{\text{pooled}} \times \sqrt{(1/n_1 + 1/n_2)}} \tag{5}$$

Equations 4.7-4.10:

$$\begin{aligned} CI =& d \pm 1.6565 \times SE & (6) \\ =& d \pm 1.6565 \times (\sigma\sqrt{(1/n_1 + 1/n_2)}) & (7) \\ =& 0.1085 \pm 1.6565 \times (0.4533\sqrt{(1/64 + 1/70)}) & (8) \\ =& 0.1085 \pm 0.1299 & (9) \end{aligned}$$

9. (**CR**) p. 128: We wrote: "We have seen that a perfect correlation is perfectly linear, so an imperfect correlation will be 'imperfectly linear'."

   What we meant to say is:

   "We have seen that the perfect correlation seen above is perfectly linear; by contrast, the correlation between midterm and final scores is not perfect."

10. (**MG**) p. 152 First occurrence of negative covariance should be positive covariance.

11. (**S**): p. 152, the equation for covariance should be:

$$\text{Covariance} = Cov(X,Y) = \frac{\sum(X-\bar{x})(Y-\bar{y})}{n-1}. \tag{10}$$

12. (**WS**) On p. 167, equation A.29: the operator in the middle term should be Var rather than E.