

Comments on and Errata for
A Methodology for the Health Sciences (2nd Edition)
by van Belle et al.

August 27, 2008

- page 4, line -19. “us” should be “use” (occurs twice)
- page 13. Mention of the Salk polio vaccine trial should be removed since this Example was deleted when the text was revised from first to second edition.
- page 26, line -13. Should be: “with ordering of qualitative...”
- page 32, table 3.7. In order for the table to be exhaustive, “>166” should be “> 164” or “≥166
- page 40, line. The whiskers should be defined as extending to the largest observation less than the 75th percentile plus 1.5 x IQR and the smallest observation greater than the 25th percentile min 1.5 x IQR. In other words, the whiskers are not directly related to the median.

- page 51. The formula for skewness should be

$$a_3 = \sqrt{n} \frac{\sum_i (y_i - \bar{y})^3}{\{\sum_i (y_i - \bar{y})^2\}^{3/2}} = \frac{m_3}{(m_2)^{3/2}}$$

i.e., the \sqrt{n} term is missing after the first equality. Also note that, contrary to the claim in the text, skewness can be negative when the mean is greater than the mode. For example, consider the data set with five observations $\{0, 2, 2, 3, 4\}$, which has mean 2.2, mode 2, and skewness $a_3 = -0.37$.

- page 51. The formula for kurtosis should be

$$a_4 = n \frac{\sum_i (y_i - \bar{y})^4}{\{\sum_i (y_i - \bar{y})^2\}^2} = \frac{m_4}{(m_2)^2}$$

i.e., the n term is missing after the first equality.

- page 54, question 3.9 (b). the acronym ECDF is not previously defined. Presumably this stands for the *empirical cumulative distribution function* as given in Definition 3.9.
- page 66, line 7 should read “Consider the outcomes “overweight” or “high blood pressure”...”

- page 72, comment 3 should refer to Note 4.10, not Note 4.9. Also, the term *symmetry* is not formally defined.
- page 76, line -2. Z_p should be z_p
- page 80. The term *quantile* is not previously defined. This term is often used synonymously with *percentile*.
- page 82, line 3-5. The sentence “The vertical scale has been stretched...” appears to be a holdover from the first edition. This sentence should have been deleted as it no longer applies to figure 4.15.
- page 89, line 3. Should refer to Note 4.16.
- page 97, middle of the page. Reference to Figure 4.14 should be to Figure 4.28.
- page 100. The Chebyshev inequality is incorrect. It should be

$$P \left[-K < \frac{Y - E(Y)}{\sigma} < K \right] \geq 1 - \frac{1}{K^2}$$

- page 123. Since s/\sqrt{n} is an estimate of the standard error, bullet 2 should read “The estimated standard error, $633.0/\sqrt{15}, \dots$ ”
- page 129, line 21. Should refer to Example 5.5, not Example 5.6.
- page 133, section 5.7.2, line 5 should refer to Table A.6, not Table A.7.
- page 141, Problem 5.1(b). Change second sentence to “Calculate the estimated standard error of the difference.”
- page 157, Procedure 5. Should read “For large n , say $n\hat{p}(1 - \hat{p}) \geq 10 \dots$ ”
- page 163, Figure 6.2 is not reproducible from the formulae given in the text. Rather, Figure 6.2 is based on the arcsin sample size formula given in Feigl (1978).
- page 169, line -6. Should read “The estimator of $\ln \omega \dots$ ”
- page 186, first equation. It is not clear why $z_{.975}$ is used here instead of $t_{.975;6}$ since $n = 7$ and the variance is being estimated.
- page 191, line 4. $p_i(\hat{\mu}, \hat{\omega})$ should be $p_i(\hat{\mu}, \hat{\sigma})$.
- page 195, Note 6.8. In the description of the exact confidence interval, replace x with Y .
- page 195, Note 6.9. The description of the rule of threes is confusing since both 90% and 95% confidence bounds are discussed. Further, whether these confidence bounds correspond to one or two-sided confidence intervals is not clear. I believe the rule can be restated completely in terms of two-sided 90% confidence intervals.

- page 259, section 8.5.3, Step 1 of signed rank statistic. Observations equal to zero should be dropped prior to the ranking step.
- page 262, line 18. Should be Table A.1
- page 270. Sentence four, change “odd” to “even.”
- page 278, Note 8.2, line 3 should be “...two possible estimators...”
- page 301, page 301, line 3 should refer to Figure 9.5.
- page 322, line -8. Change 0:6891 to 0.6891.
- page 363. Def 10.3 should let $\alpha_i = \mu_i - \mu$
- page 363. The last equation should be

$$Y_{ij} = \bar{Y}_{.j} + (\bar{Y}_{i.} - \bar{Y}_{.j}) + (Y_{ij} - \bar{Y}_{i.})$$

- page 401, Example 10.5 (continued) refers to Table 10.8. However Table 10.8 does not correspond to the mice example. It is not clear to what table the authors are referring.
- page 402, Cochran’s test should be

$$C = \frac{s_{\text{maximum}}^2}{\sum s_i^2}$$

- page 404, line -7 should read “For example, the smallest deviation is...” Two other comments on this page. First, several ideas are redefined here that have appeared earlier in the book (e.g., order statistics). Why define again? Second, the presentation here does not make it clear that what we are looking for in Figure 10.8 is linearity.
- page 406, line 2. Using the data from Example 10.1, I get $r = 0.9602$.
- page 438, line -2. Should be “be one if variability in Y can be perfectly predicted.”
- page 525. $\sigma_{\hat{\theta}^2}$ should be $\sigma_{\hat{\theta}}^2$.
- page 535, line 7. The *overall significance level* is not defined, but appears to be the same as the *per experiment error rate*.
- page 535, result 12.3. The quantile of the t distribution to be used is ambiguous as written. Instead, use $t_{m, 1-\alpha/(2N)}$.
- page 640, line -6. Should be: “The *incidence* of a disease in a fixed time interval is the number of new cases diagnosed during the time interval divided by the person-time at risk.”

- page 640, line -5. Should be: “The *prevalence* of a disease is the **proportion** of people with the disease at a fixed time point.”
- page 642. Prevalance is not a rate since it does not involve the change of a quantity over time.
- page 654, problem 15.3. Part a should refer to a “rate” not “rates”. Parts b and c should refer to a “ratio” not “rates”.
- page 821, Table A.3. In the footnote, D denotes the degrees of freedom (d.f.)
- Table A.5 and A.6. The columns should be labeled 1 2 3 4 5 6 7 8 9 10 12 15 20 24 30 40 60 125 ∞ .
- Table A.9. This table is not consistent as to whether the critical values are in the critical region. For example, for two sided $\alpha = 0.05$ and $n = 6, 7, \dots, 14$, the critical values should be 0, 2, 3, 4, 8, 10, 13, 17, 21, 25 according to Rosner’s text. More specifically, for $n = 6$, $\Pr[S^+ \leq 0] = 0.016$ and $\Pr[S^+ \leq 1] = 0.03125$. Under the convention that the critical value be in the rejection region, this would imply the critical values is 0. This is consistent with Rosner and different from the text. Now consider $n = 7$ where $\Pr[S^+ \leq 2] = 0.0234$. Again under the convention that the critical value is in the rejection region, this would imply the critical value is 2. This is consistent with Rosner and the text.