

**List of Corrections for Applied Survey Data Analysis-August 17, 2020**

---

Page 18, SRSWOR should be SRSWR.

Page 19, equation 2.1, the  $y_i$  in the numerator should be  $Y_i$ .

Page 33, line 2, the formula  $f(h)=n(n)/N(h)=n/N=f$ , should be  $f(h)=n(h)/N(h)=n/N=f$ .

Page 36 equation 2.14, replace  $\bar{y}$  by  $\bar{y}_w$  in  $s_w^2$

Page 45 equation 2.21, replace  $\sigma^2(w)$  by  $s^2(w)$

Page 54, equation 3.1, note that this assumes a model with an intercept (see Figure 3.1).

Page 57, equation 3.4, note that the linear regression coefficient is assumed to be from a model without an intercept.

Page 59, Table 3.1, Unweighted model, coefficient for Suburban should be 0.05878 not 0.5878.

Page 60, a recommended reference for informative sampling, Eideh, A. H. and Nathan, G. (2009). Two-Stage Informative Cluster Sampling with application in Small Area Estimation. *Journal of Statistical Planning and Inference*.139, pp 3088-3101.

Page 77, first sentence "The second JRR replicate is formed by dropping the second PSU..." should be "The second JRR replicate is formed by dropping the **first** PSU..."

Page 77, the estimated population means for the four replicates should be: 0.4752174, 0.4779167, 0.4732000, 0.4827273 not as listed in 3.6.3.1.3.

Page 77, Sect 3.6.3.1.4 the variance of the mean under JRR should be 0.00007720 (not 0.00005790). It is correct in the standard error expression on the following line.

Page 84, Table 3.3,

$$\bar{y} = \bar{Y}_0 \text{ BY } \bar{Y} = \bar{Y}_0$$

$$\bar{y}_1 - \bar{y}_2 = 0 \text{ BY } \bar{Y}_1 - \bar{Y}_2 = 0$$

$$\sum_j a_j p_j = 0 \text{ BY } \sum_j a_j P_j = 0$$

Page 158, first section of Stata syntax should be:

```
egen meanbpsy = rowmean( bpxsy1 bpxsy2 bpxsy3 bpxsy4)
```

```
egen meanbpdi = rowmean( bpxdi1 bpxdi2 bpxdi3 bpxdi4)
```

```
generate bp_cat = .
```

```

replace bp_cat = 1 if (meanbpsy<120 & meanbpdi<80)

replace bp_cat = 2 if (missing(bp_cat) & ((meanbpsy >= 120 & meanbpsy < 140 ) ///
| (meanbpdi >= 80 & meanbpdi < 90 )))

replace bp_cat = 3 if (missing(bp_cat) & ((meanbpsy >= 140 & meanbpsy < 160 ) ///
| (meanbpdi >= 90 & meanbpdi < 100 )))

replace bp_cat = 4 if (missing(bp_cat) & ((160 <= meanbpsy & meanbpsy != .) ///
| (meanbpdi >= 100 & meanbpdi !=. )))

```

Page 160, Stata syntax for Example 6.5 (Bar Chart) should be:

```

graph bar (mean) bp_cat1 bp_cat2 bp_cat3 bp_cat4 [pweight=wtmec2yr] if age18p==1 , percentages ///
bar(1,color(gs12)) bar(2,color(gs4)) bar(3,color(gs8)) bar(4,color(black)) ///
blabel(bar, format(%9.3g)) bargap(7) scheme(s2mono) ///
legend (label(1 "Normal") label(2 "Pre-Hypertensive") label(3 "Stage1 Hypertensive") ///
label(4 "Stage 2 Hypertensive")) ytitle ("Percentage")

```

Page 167, Final line "(Stater, 2008)" should be "(Stata, 2008)".

Page 169, Stata code to produce Table 6.6 should be *subpop(if 18<=age & age < 29)*

Page 175, ChiSquare value for CMH test should be 94.26 not 92.46

Page 191, equation 7.22, note that the linear regression coefficient is assumed to be from a model without an intercept.

Page 201, in expression (7.35) on page 201, the expression for the Overall test should not have the subscript "2" following the B(hat) matrix symbol. It should just be B(hat). The partial test expression is correct.

Page 203, last paragraph "(P Manual, p. 611, StataCorp, 2008)" should be  
(Reference Manual [R] I-P, p. 611, StataCorp, 2008)".

Page 219, in the footnote for Table 7.4,  $v < 0.001$  should be  $p < 0.001$ .

Page 230, at the bottom of page 230, the | in  $E(y | x)$  is cut off at the end of a line. In the text,  $E(y x)$  should be  $E(y | x)$  as in the equation above.

Page 250, Table 8.4: column Logit:  $z_j$  should be:

Logit: z <sub>j</sub>	Odds Ratio
0.126	0.82
0.046	0.76
-0.332	0.52
0.781	1.57
0.504	1.19
0.327	1.00

Page 250, equation 8.24 first term under the radical sign should be squared:  $(x_{1j}-x_{2j})^2$ . The remainder of equation 8.24 is correct.

Page 296, the first line of code specifying survey variables says [pweight = kwqtr] when it should be [pweight = kwgtr].

Page 329, syntax to expand data set should be:

```
expand intwage
bysort caseid: gen pyr = _n
```

rather than:

```
expand intwage
sort caseid
gen pyr=_n
```

Page 366, Table 11.5: Complete Case Analysis n=4,308 not n=4,305.

Page 367, line 6: n=4,305 should be n=4,308.

Page 403, first row of Table A.2 for TSL, WesVar "Yes" should be "No".

Page 440, there is a missing reference that should be inserted after Rust and Hsu: Rust KF, Rao JN: Variance estimation for complex surveys using replication techniques. Stat Methods Med Res 1996, 5:283-310.

Page 463 and 464, 1. Index keyword: SPSS (page 463) with pages named in index 433-439 should be pages, 421-427 and 2. Index keyword: SUDAAN (page 464) with pages named in index 427-433 should be pages 414-421.