

- 2.1 (i) Bar chart. Median = 1.
 (ii) 0.569 (0.528, 0.611).

- 2.2 (i)

Education	Marital status			Total
	Single	Separated	Married	
Males				
Illiterate	9	4	122	135
Literate	57	14	357	428
Total	66	18	479	563
Females				
Illiterate	1	4	13	18
Literate	13	18	70	101
Total	14	22	83	119
Persons				
Illiterate	10	8	135	153
Literate	70	32	427	529
Total	80	40	562	682

- (ii) Test statistic = 0.86 ($p = 0.65$). 39.9 (36.2, 43.6).

- 2.3 All show positive skewness.

- 2.4 (ii)

Variable	Minimum	Q1	Q2	Q3	Maximum
Protein C	80	92.5	109	143.5	188
Protein S	63.91	93.375	103.75	117.86	142.76

- (iii), (iv)

Variable	Mean	Standard deviation	Skewness
Protein C	119.1	31.89	0.78
Protein S	104.9	19.22	-0.17

- (v) Protein S is not highly skewed. Protein C is, but the negative reciprocal transformation removes most of this skewness (and will be used subsequently).
 (vii) Protein C (after using the transformation): (103.8, 121.2). Protein S: (98.76, 111.06).
 (viii) (0.502, 0.798).
 (x) Preliminary F test statistic = 1.62 ($p = 0.29$). Pooled t test statistic = 0.62 ($p = 0.54$).
 (xi) Wilcoxon test $p = 0.68$.
 (xii) Preliminary F test statistic (after transformation) = 1.21 ($p = 0.66$).

Pooled t test statistic (after transformation) = 2.41 ($p = 0.02$).

Wilcoxon test $p = 0.03$.

- (xiii) Protein C, but not Protein S, varies with drinking status. Nondrinkers have higher Protein C.
- (xiv) $(-9.0, 17.0)$. Yes, because it includes zero.
- (xv) t test statistic = -1.31 ($p = 0.2$). No evidence of a real difference.
- 2.5 (i) Test statistic = 0.04 ($p = 0.97$).
(ii) 0.06 $(-2.40, 2.52)$.
(iii) No effect.
- 2.6 The major features are as follows:
- The bulk of sugar intake is extrinsic.
 - Extrinsic (and thus total) intake decreases with obesity.
 - Intrinsic intake increases with obesity.
 - Lactose intake is largely unrelated to obesity.
 - Sugar intake patterns are very similar for men and women.
 - Women derive slightly more energy from nonextrinsic sugar, and less from extrinsic sugar, than do men.
- 2.7 0.644 (0.532, 0.757).
- 2.8 (i) 0.196 (0.147, 0.245): many users are missed.
(ii) 0.952 (0.914, 0.989): nonusers are usually identified.
(iii) 0.891 (0.809, 0.973): positive test is fairly reliable.
(iv) 0.370 (0.317, 0.423): negative test is unreliable.
- 2.9 (i) 100. (ii) 80. (iii) 120.
- 2.10 For consistency, the cigarette smoking groups and the orientation of smoking habit could be standardised. The best orientation would have smoking groups labelling the rows since the major comparisons will be made between these groups and because the order of magnitude of the numbers differs (e.g., in Table 1.2).
- 3.1 (i) 0.100 (0.030, 0.170) and 0.033 (0.024, 0.042).
(ii) 3.02 (1.42, 6.41).
(iii) 0.111 and 0.034.
(iv) 3.25 (1.42, 7.43).
(v) Continuity-corrected test statistic = 6.83 ($p = 0.009$).
(vi) 0.079 (0.027, 0.214).
- 3.2 $p = 0.018$ (one-sided). The more severe the asthma, the more likely are parents to provide adequate treatment.
- 3.3 (i) 2.40 (1.11, 5.15).
(ii) 3.16 (1.04, 9.56).
(iii) Continuity-corrected test statistic = 3.14 ($p = 0.08$). Fisher's exact test is better (an expected value is below 5). The two-sided test (found from the CSI procedure in Stata) has $p = 0.08$.
(iv) 0.140 (0.039, 0.395).