Chapter 13: When and why might I need to use non-parametric statistics?

Full answers to study questions

1. You could have suggested a large number of variables for this task, but below are some of my ideas.
    1.1. A nominal variable would need to define categories that people can belong to, such as which political party the person voted for in the last general election: Conservative, Labour, Liberal Democrats, Other.
    1.2. An ordinal variable would require data that exist in a particular order, but with variable distances between the orders. I could list ten different aspects of political policy and ask participants to rank order them according to how important each one is when they decide which party to vote for. For each of the ten policies, I would then have ranked ordinal data.
    1.3. An interval variable needs to be continuous, with the same distance between each number, and where negative values are possible. I could devise a questionnaire where participants rate the extent to which they agree with different political statements on a seven point Likert scale. When the scores are summed across all of the items, they give a score from -70 indicating more leftwards political opinions, through to +70 indicating more rightwards political opinions.
    1.4. A ratio variable needs to be continuous, with the same distance between each number, and where negative values are not possible. For this example I might look at people's social media accounts for the three months before a General Election, and count the number of times they post about political issues.
2. Calculations for these questions are given below.
    2.1. Lower boundary: mean - 1SD
    Lower boundary: 27 - 5 = 22
    Upper boundary: mean + 1SD
    Lower boundary: 27 + 5 = 32
    Percentage calculation: (155/200) * 100 = 77.5%
    Percentage calculation: 0.775 * 100 = 77.5%
    Scores would range from 22 to 32. 77.5% of the sample fall within ± 1 SD around the mean. You would expect 68% of the sample to fall within ± 1 SD, so this dataset includes more participants than would be expected with normally distributed data.
    2.2. Lower boundary: mean - 2SD
    Lower boundary: 27 - 10 = 17
    Upper boundary: mean + 2SD

Lower boundary: 27 + 10 = 37

Percentage calculation: (162/200) * 100 = 81%

Percentage calculation: 0.81 * 100 = 81%

Scores would range from 17 to 37. 81% of the sample fall within ± 2 SD around the mean. You would expect 95% of the sample to fall within ± 2 SD, so this dataset includes fewer participants than would be expected with normally distributed data.

3. Calculations for these questions are given below.

Boys			Girls		
Aggression score (x)	$x - \bar{x}$	$x - \bar{x}^2$	**Aggression score (x)**	$x - \bar{x}$	$x - \bar{x}^2$
37	11	121	26	0.75	0.5625
25	-1	1	21	-4.25	18.0625
21	-5	25	19	-6.25	39.0625
24	-2	4	28	2.75	7.5625
16	-10	100	29	3.75	14.0625
38	12	144	24	-1.25	1.5625
29	3	9	30	4.75	22.5625
18	-8	64	25	-0.25	0.0625

Boys, mean:

$$Mean = \frac{37 + 25 + 21 + 24 + 16 + 38 + 29 + 18}{8}$$

$$Mean = \frac{208}{8}$$

$$Mean = 26$$

Boys, variance:

$$Variance = \frac{121 + 1 + 25 + 4 + 100 + 144 + 9 + 64}{8 - 1}$$

$$Variance = \frac{468}{7}$$

$$Variance = 66.86$$

Girls, mean:

$$Mean = \frac{26 + 21 + 19 + 28 + 29 + 24 + 30 + 25}{8}$$

$$Mean = \frac{202}{8}$$

$$Mean = 25.25$$

Girls, variance:

$$Variance = \frac{0.5625 + 18.0625 + 39.0625 + 7.5625 + 14.0625 + 1.5625 + 22.5625 + 0.0625}{8 - 1}$$

$$Variance = \frac{103.5}{7}$$

$$Variance = 14.79$$

3.1. Boys: mean = 26, variance = 66.86. Girls: mean = 25.25, variance = 14.79.

3.2. The smallest variance is 14.79 in the girl's condition. Four times this is 59.16 (4*14.79 = 59.16). The variance in the boy's condition is 66.86, which is greater than 59.16, and therefore more than four times the size of the variance in the girl's condition. This means that there is evidence of heterogeneity of variance in this dataset (different variances), and the assumption of homogeneity of variance has been violated.

3.3. A non-parametric test would need to be used as the parametric assumption has been violated.