6.3.2 Spearman rank correlation and test choice

Name: _____ Class: _____ Date: _____ **Total: 19 marks**

KEY WORDS rank correlation • critical value • t-test t

Objective

Run and interpret Spearman's rank correlation: rank, difference, the formula, the critical value and the three-sentence conclusion – and choose t-test versus correlation for the plan question. Original practice for the Unit 6 statistics demands.

1 Key knowledge and worked examples

■ Ranking, d and d²

Rank each variable separately, highest value = rank 1 (state your convention). The difference between the ranks in each pair is d; square each (the sign disappears). For the oil data the ranks give $d = -6, -4, 0, -1, 1, 5, 5$ and:

$$d^2 = 36, 16, 0, 1, 1, 25, 25 \qquad \sum d^2 = 104$$

■ The formula and the sign

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 104}{7 \times (49-1)} = 1 - \frac{624}{336} = -0.857$$

Keep the minus sign in the answer: oil falls as time passes – a negative correlation. Accept −0.86 or −0.9; the sign must be present.

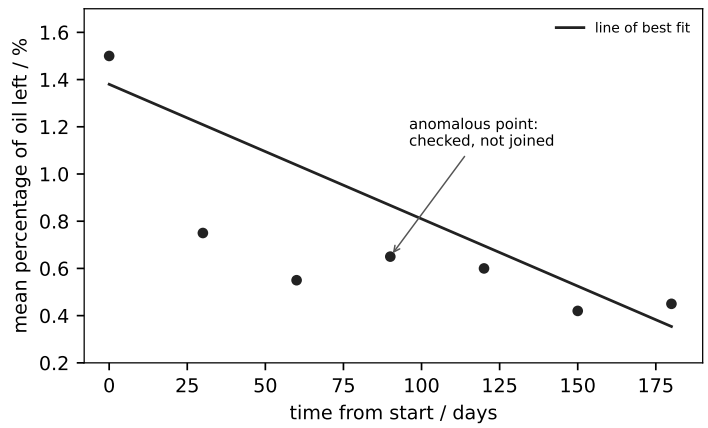
■ Three sentences to the conclusion

1. The calculated value (0.857, magnitude) exceeds the critical value at $p = 0.05$ for $n = 7$, which is 0.786.
2. Therefore reject the null hypothesis.
3. There is a significant negative correlation between the percentage of oil remaining and storage time.

Below the critical value: accept the null hypothesis – no significant correlation.

■ Which test when

Difference between two group means (bar chart) → **t-test** t . Association between two measured variables on the same subjects (scatter/line graph) → correlation, any number of levels. Never test a correlation with a t-test.



2 Practice

2.1 Two variables are ranked for n = 8 pairs; $\Sigma d^2 = 42$. Calculate r_s. [2]

2.2 With $r_s = -0.71$ and a critical value of 0.738 at p = 0.05 (n = 8), write the full conclusion. [2]

2.3 State the null hypothesis for: "students' recovery time shortens as weekly training hours increase". [1]

2.4 For each study choose t-test or correlation: (a) mean blood pH of athletes versus non-athletes; (b) heart rate against running speed over five speeds. [2]

3 Exam-style questions

3.1 The oil-decay ranks are: time ranks 1–7 (day 0 highest); oil ranks 7, 6, 3, 5, 4, 1, 2; $d = -6, -4, 0, -1, 1, 5, 5$. Calculate Σd^2 and r_s for $n = 7$, and deduce the conclusion using the critical value 0.786 at $p = 0.05$. [5]

3.2 A student writes: "$rs = -0.857$ is more negative than -0.786, so the correlation is significant." Correct the comparison being made, and give the two remaining conclusion sentences. [3]

3.3 In a plan question you compare mean recovery time for two groups, and separately the recovery times of students across five weekly-training levels. Name the graph and the statistical test for each part, and justify one choice. [4]

Solutions

Original-task marking: M1 method, A1 accuracy, B1 reasoning.

2.1 $r_s = 1 - (6 \times 42)/(8 \times 63) = 1 - 252/504 = 1 - 0.5 = 0.5$.

2.2 The calculated value 0.71 is less than the critical value 0.738, so accept the null hypothesis: no significant correlation at p = 0.05.

2.3 There is no correlation between recovery time and weekly training hours.

2.4 (a) t-test – comparing the means of two groups. (b) Correlation – association between two measured variables across levels.

3.1 $d^2 = 36, 16, 0, 1, 1, 25, 25$, so $\sum d^2 = 104$. $r_s = 1 - 624/336 = -0.857$. The magnitude 0.857 exceeds the critical value 0.786; reject the null hypothesis – there is a significant negative correlation between oil remaining and time.

3.2 The comparison uses the magnitude (0.857 versus 0.786), not the signed values – the sign records direction, not strength. Remaining sentences: therefore reject the null hypothesis; there is a significant negative correlation between the percentage of oil remaining and storage time.

3.3 Two groups: bar chart of mean recovery time per group with a t-test; five levels: scatter or line graph of recovery time against training hours with a correlation test such as Spearman's rank.