PO91Q: Fundamentals in Quantitative Research Methods

Worksheet Week 7 - Solutions

Dr Florian Reiche

F.Reiche@warwick.ac.uk

1 | Crosstabulations – Calculations by Hand

Expected Values:

	Mode of Transport			
Department	Bike	Bus	Car	Total
PAIS	7.2	6.8	6	20
CIM	10.8	10.2	9	30
Total	18	17	15	50

- Calculate the χ^2-value
  - **(2.8867)**
- How many degrees of freedom does this table have? Why?
  - **(2)**
- Using the χ^2 Table, what is the p-value?
  - **(0.236, or between 0.250 and 0.100)**
- Are mode of transport and departmental association independent in the population?
  - **(Yes)**

2 | Correlation

Linear relationships

1. Increasing n from 50 to 300 makes the sample r cling more tightly to the target (≈0.60), narrows its confidence interval, and yields a much smaller p-value. This shows that larger samples reduce sampling variability and increase power.
2. With a small true effect (r ≈ 0.20) and n = 60, the observed r can bounce widely across seeds and may even flip sign occasionally. This instability comes from high sampling error relative to the weak signal.

Non-linear and non-monotonic (Quadratic)

3. Often yes: r can be near zero despite a clear U-shape. The positive and negative slopes on either side of the minimum/maximum cancel in a linear summary, masking the strong but non-linear association.
4. When Noise is large enough that the two arms blur together (threshold depends on n and the scale), both the straight line and LOESS become close to flat and similarly uninformative.

Non-linear but monotonic (Cubic)

5. Yes. The straight line averages across the curve and often understates the visible relationship (sometimes showing a small slope). LOESS follows the bend and reveals a strong monotonic trend.
6. As Noise increases, r typically declines quickly toward zero because curvature plus added noise erodes linear association. The LOESS curve still shows the underlying shape longer, revealing curvature even when r is small.

Comparing fits (lines)

7. For a curved monotonic pattern, the straight line can imply weak or even misleading direction, while LOESS shows the true increasing pattern with curvature (steeper ends, flatter middle).
8. In Linear mode (truly linear data), LOESS and the least-squares line essentially coincide, especially as n grows, because the best smooth for a straight-line signal is a straight line.

Sampling variability and reproducibility

9. With r = 0.40 and n = 80, r typically varies noticeably across seeds (often on the order of ±0.1 or so). With n = 400, the variability shrinks substantially (often to just a few hundredths), illustrating how larger n stabilizes estimates.
10. For Cubic with low Noise, r is moderately stable but still more variable than in a truly linear case because the linear summary is sensitive to where the curve is sampled. Stability improves with larger n and lower Noise.

Same r, different story

11. You can tune Cubic (noise) to get a Pearson r similar to a Linear scenario (e.g., both around 0.5). The plots differ: Linear looks elliptical with a constant slope; Cubic is curved. LOESS highlights this curvature, showing why the same r can represent different relationships.

Reflections

12. Pearson's r is adequate when the relationship is approximately linear and residual variation is roughly symmetric and homoscedastic. Prefer a non-linear description when you see curvature or non-monotonic structure that a straight line cannot capture.
13. Justify a non-linear model when LOESS departs clearly from the straight line, residuals show systematic patterns, and a non-linear fit improves predictive performance (e.g., lower residual error, better cross-validated metrics) while matching the visible shape in the plot.

3 | Applied Exercises in R (Core & Going Further)

See RScript in the Online Companion