2. (a) IV (b) III (c) V (d) I

3. (b) and (c) only

4. (a) `gf_point(y ~ x, data = xyPairs) |> gf_lm()`

(c) The appearance of the residuals-vs.-fitted-values plot is that of a random (unpatterned) scatter of points about the zero line with no tendency to expand/contract in distance from the zero line as x changes. This confirms the independence of residuals, as well as the one uniform σ applying at all x, assumed in the SLM.

The appearance of the normal quantile plot of residuals is that of a straight line, as it should be if residuals follow a normal distribution.

(d) The coefficient of determination (R^2) tells what fraction of variability in (observed) y-values is explained by the linear model in x.

(e) $r = -\sqrt{0.686} = -0.828$

(f) $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$ (ρ can appear instead of β_1)

(g) $t = (-0.828)\sqrt{\dfrac{82}{1-(0.828)^2}} = -13.372$

P-value: `2 * pt(-13.372, 82)`

(h) The "prediction" one is to locate the likely range of a single y-value at x = 27. The "confidence" one is to locate the likely range of the mean y-value at x = 27. The "prediction" one is wider.

5. (a) H_0: Variables river and group are independent
H_a: The variables have an association

(b) C-Yellow is smallest: $\dfrac{(135)(141)}{454} = 41.92$

(c) $\dfrac{(O-E)^2}{E} = \dfrac{(37-41.92)^2}{41.92} = 0.5774$

(d) 1 - pchisq(5.595, 4)

(e) It is valid to use pchisq(), since all expected counts are ≥ 5.

6. (a)

Source	df	SS	MS	F
Group	2	43.935	21.968	4.119
Error	59	314.627	5.333	

(b) H_0: $\mu_A = \mu_B = \mu_C$

H_a: At least two means are different

(c) We are told the samples are independent. ✓

The populations (each) are normal $\Rightarrow$ each $\bar{x}_A, \bar{x}_B, \bar{x}_C$ are normal ✓

$s_{max}/s_{min} = 2.549/1.922 < 2$ ✓ Yes, it is valid.

(d) 1 - pf(4.119, 2, 59)

(e) $\mu_A \neq \mu_B$, significant at the 10% (even the 5%) level.