Form B

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

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) $r = -\sqrt{0.818} = -0.9044$

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

(g) $t = (-0.904)\sqrt{\frac{82}{1-(0.904)^2}} = -16.55$

P-value: 2 * pt(-16.55, 61)

(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 "confidence" one is narrower.

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

(b) B-Nile is smallest: $\frac{(168)(104)}{577} = 30.28$

(c) A-Amazon has expected count $\frac{(205)(159)}{577} = 56.49$ $\frac{(47-56.49)^2}{56.49} = 1.594$

(d) `1 - pchisq(9.153, 6)`

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

6. (a)

Source	df	SS	MS	F
Group	2	36.793	18.397	3.973
Error	82	379.753	4.631	

(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 ✓

$s_{max}/s_{min} = 2.403/1.955 < 2$ ✓ Yes, it is valid.

(d) `1 - pf(3.973, 2, 82)`

(e) $\mu_A \neq \mu_C$, significant at the 5% level