1. (a) 3 (b) 1 (c) 2

2. It means you are predicting a y-value at an x-value "within" the x-values contained in your data — that is, with $x_{min} \le x \le x_{max}$. It is safer to interpolate than to extrapolate, because it is quite possible our linear association between y and x is invalid/inappropriate outside of our sampled x-values.

3. (a) Choose the cell with minimal row/column totals: (Case, Pet Store)

(b) $\frac{(285)(75)}{1013} \doteq 21.10$

(c) That all expected counts should be ≥ 5. It is met here.

4. (a) $R^2 = 0.433$, which we interpret to say that 43.3% of the variability in sampled voplus-values is explained by our linear model.

(b) $r = \sqrt{0.433} \doteq 0.658$ (choosing positive root since $b_1 > 0$)

This coefficient addresses the strength of the linear model. It is a relatively strong linear association.

(c) $\widehat{\text{voplus}} = 316.37 + 19.86\,(\text{oc})$

(d) Noteworthy(?)

- Some large residuals, particularly one large positive one,
- There may be a pattern (undesirable) emerging at the right side, where we end with three straight negative residuals.

To the extent these are not artifacts of sampling, the cast some doubt on the validity of the model.

(e) This t^* is appropriate for a 92% confidence level. The interval is

$b_1 \pm (t^*)(SE_{b_1}) = 19.86 \pm (1.8166)(4.295)$, or

$(12.06,\ 27.66)$

(f) It is $t = \frac{b_1}{SE_{b_1}} = \frac{19.86}{4.295} \doteq 4.623$. It can be computed as $r\frac{\sqrt{n-2}}{\sqrt{1-r^2}}$, too.

(g) $[1 - pt(4.623, df = 28)] * 2$.

5. (a) $\boxed{H_0: p_N = 0.179,\ p_M = 0.217,\ p_S = 0.371,\ p_W = 0.233}$

Category	Observed	Expected
Northeast	173	179 = (1000)(0.179)
Midwest	205	217 = (1000)(0.217)
South	384	371 = (1000)(0.371)
West	238	233 = (1000)(0.233)

$$\chi^2 = \frac{(173-179)^2}{179} + \frac{(205-217)^2}{217} + \frac{(384-371)^2}{371} + \frac{(238-233)^2}{233} \doteq \boxed{1.428}$$

(b) $1 - pchisq(1.428, df = 3)$

(c) If not significant, then the data is consistent with people continuing to live according to 2010 census proportions.

(d) A significant P-value means we reject that people continue to reside according to 2010-census proportions.

6. (a) Let μ, with subscripts indicating flux type, be population mean hardness.
$H_0: \mu_A = \mu_B = \mu_C = \mu_D$, and H_a: at least two group means are unequal.

(b) The rule of thumb, $2 \geq \frac{s_{max}}{s_{min}} = \frac{9.757}{5.403}$ is met.

(c)

df	SS	MS	F
3	743.4		4.842
20	1023.6	51.18	
23			

(d) t (2-sample t)

(e) $1 - pf(4.842, 3, 20)$

(f) We conclude at least two group means are different.

(g) TukeyHSD( )
Looking over the pairwise comparisons, the only one for which the adjusted P-value is under 5% is the one comparing means for flux A and flux C. We conclude $\mu_A - \mu_C \neq 0$. We are unable to conclude that any other group mean pairs are different.