1. (a) H_0: The variables color and group are independent

   H_a: There is an association between color and group

   (b) (White, group A) – it combines the smallest row and column totals

   (c) $\frac{(293)(284)}{1988} = 41.857$

   (d) It is justified, as every expected count is at least 5.

   (e) 1 - pchisq(19.906, df = 4)

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

3. (a) $n = 5+9+10+15+7+11 = 57$

   Expected for Blue: $(57)(0.24) = 13.68$

   Expected for Red: $(57)(0.13) = 7.41$

blue	13.68	orange	11.4
brown	7.98	red	7.41
green	9.12	yellow	7.98

   (b)

$$\chi^2 = \frac{(5-13.68)^2}{13.68} + \frac{(9-7.98)^2}{7.98} + \frac{(10-9.12)^2}{9.12} + \frac{(15-11.4)^2}{11.4} + \frac{(7-7.41)^2}{7.41} + \frac{(11-7.98)^2}{7.98} = 8.025$$

   (c) 1 - pchisq(8.025, df = 5)

   (d) "We fail to reject that the counts of colors are distributed differently than asserted on the website."

   Or, "We have insufficient evidence to reject the company's claim concerning the proportions of colors it produces."

4. (a) That these are independent samples is likely, but cannot be verified by this table. The 2:1 ratio is in place, since $0.794/0.442 < 2$. We don't know that the three population distributions are normal (or even symmetric) so we wish our sample sizes (10 in each group) were higher (30 or more).

   (b)

df	SS	MS	F
2	3.7664		4.846
27		0.3886	

   (c) 1 - pf(4.846, 2, 27)

   (d) TukeyHSD(weight ~ group, data = PlantGrowth)

(e) There is no evidence of a significant difference between the control group's mean and those of either treatment – 0 lies inside those confidence intervals. 0 is absent from the last confidence interval, so μ is significantly different for the two treatments.

5. (a) gf_point(width ~ length, data = KidsFeet)

(b) Graph A has points adhering closely to a line, helping to confirm the SLM. The points in B are fairly random (unpatterned) on either side of the zero line, without an obvious change in how far from the zero line at different choices of "length".

(c) Following

```
> lmResult = lm( width ~ length, data = KidsFeet),
```

the plots are produced by

```
A: gf_qq( ~ residuals( lmResult ))
B: gf_point( residuals( lmResult ) ~ fitted( lmResult ))
```

(d) $$R^2 = \frac{4.0557}{4.0557 + 5.812} \doteq 0.411$$

Interpret as saying about 41.1% of the variability in widths is explained through the linear model with length.

(e) $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$.

The F-statistic (25.819) corresponds to a P-value of 1.097×10^{-5}, smaller than all the usual choices of α. We reject H_0 in favor of the alternative H_a.

(f) II (7.7694, 9.3608)
III (8.3874, 8.7428)
I (7.8732, 9.2569)