1. (a) H_0: There is no association between group and color.

H_a: There is an association between group and color.

(b) In the (A, white) cell, the observed count is 41, while the expected count is $\frac{(148)(76)}{292} = 38.5206$

so, its contribution to the χ^2-statistic is

$$\frac{(41-38.5206)^2}{38.5206} = \boxed{0.1596}$$

(c) All expected counts have 292 in the denominator. We get the smallest one when our numerator pairs the smallest row total with the smallest column total, which occurs at the (C, Red) cell. The expected count for that cell is

$$\frac{(38)(51)}{292} = 6.637.$$

(d) It is justified, since all expected counts exceed 5.

(e) We would consult a chi-square distribution, the one with $(3-1)\times(3-1) = 4$ degrees of freedom.

(f) 1 − pchisq(4.3166, df = 4)

2. (a) length is explanatory, as it is plotted along the horizontal axis.

(b) Overall pattern does seem linear. Residuals are near zero much more often than far from zero (like the standard normal distribution). The one thing we might note is that residuals have a smaller spread about the line on the left end than on the right end.

(c)

	df	SS	MS	F
length	1	3666.2	3666.2	196.76
Residuals	52	968.9	18.633	
Total	53	4635.1		

(d) $$R^2 = \frac{\text{SSModel}}{\text{SSTotal}} = \frac{3666.2}{4635.1} = 0.791$$

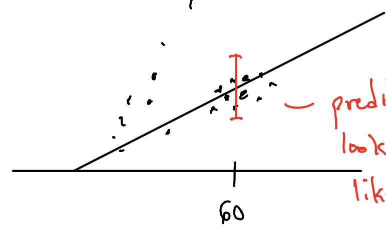
This indicates 79.1% of the variability in responses is "explained" by the model.

$$\Rightarrow \text{Correlation } r = \sqrt{0.791} = 0.8894$$

(e) The Total dfs, 53, equals $n-1$, where n is the sample size.
So there are 54 cases/bears.

(f) Since our F test statistic, 196.76, is well in the rejection region,
we would reject $H_0: \beta_1 = 0$ in favor of $H_a: \beta_1 \neq 0$.

(g) This prediction interval attempts to capture the chest size/value of the next bear we see whose length is 60.



prediction interval looks something like this

(h) We get b_1, SE_{b_1} from the output of

summary(lm(chest ~ length, data = bears))

We get our critical value from

qt(0.95, df = 52)

3. (a) $H_0: p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$

H_a: At least one of these proportions is not $1/6$.

(b)

Roll	1	2	3	4	5	6
Observed Count	5	7	17	16	8	7
Expected Count	10	10	10	10	10	10

← each is $(60)(\frac{1}{6}) = 10$

$$\chi^2 = \frac{1}{10}\left[(5-10)^2 + (7-10)^2 + (17-10)^2 + (16-10)^2 + (8-10)^2 + (7-10)^2\right]$$

$$= \frac{1}{10}(25 + 9 + 49 + 36 + 4 + 9) = \frac{1}{10} \cdot 132 = \boxed{13.2}$$

(c) 1 - pchisq(13.2, df = 5)

4. (a) We choose $df = \min(27, 19) - 1 = 18$. The command:

qt(0.97, df = 18)

(b) The point estimate is $\bar{x}_1 - \bar{x}_2 = 27.9 - 32.3 = -4.4$

We have $$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{6.8^2}{27} + \frac{5.1^2}{19}} \doteq 1.7554$$

So, our CI is $-4.4 \pm (2.0071)(1.7554)$, or $(-7.923, -0.877)$.

(c) We have followed a procedure that began with acquiring two independent samples, one from each group, ultimately leading to an interval centered on the point estimate $\bar{X}_1 - \bar{X}_2$, which has a 94% success rate in enclosing $\mu_1 - \mu_2$ inside the two endpoints. We do not know if $\mu_1 - \mu_2$ is inside the interval from (b). Our confidence lies in the process, not the result.

(d) There is no indication either way on whether the two underlying populations from which our samples are drawn are themselves normally distributed. If we knew they were, then any sample sizes would be adequate. In the absence of such knowledge, however, we would prefer both n_1 and n_2 be at least 30, which they are not. So, there is some lack in evidence to be assured that $\bar{X}_1 - \bar{X}_2$ has an approximately normal distribution.