Answers

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

2. (a) It say that 85.8% of the variability in response values has been explained by the linear model.
   (b) The fitted value at $x = 39$ is $\widehat{y} = 181.645 - (3.645)(39) = 39.49$.
   (c) We know the correlation r is the square root of the coefficient of determination, but as there are two square roots (a positive root and a negative one), we use the fact that b_1 is negative to choose which sign: $r = -\sqrt{.858} \doteq -0.926$.
   (d) $\mathbf{H}_0: \quad \beta_1 = 0 \quad$ (though $\rho = 0$ is equally good here)

       $\mathbf{H}_a: \quad \beta_1 \neq 0 \quad$ (though $\rho \neq 0$ goes with the above variant)
   (e) $n = 27$ and $r = -0.926$, so $t = (5)(-0.926)/\sqrt{1-(.926)^2} = -12.26$. The right command is

```
pt(-12.26, df=25) * 2
```

3. (a)

Source of Variation	df	SS	MS	F
Group	3	77.6	25.87	0.536
Error	89	4293.7	48.24	
Total	92	4371.3		

   (b) The researchers compared $3 + 1 = 4$ different diets.
   (c) $92 + 1 = 93$ subjects completed the study.
   (d) $\mathbf{H}_0:$ the (population) mean weight loss is the same for all four diets

       $\mathbf{H}_a:$ at least one diet's population mean is different from another
   (e)

```
1 - pf(0.536, df1=3, df2=89)
```

   (f) You might list the ideal conditions, namely that
       - that the amount of weight lost is normally-distributed within each population of dieters (i.e., those on Diet 1, those on Diet 2, etc.).
       - that within each population of dieters, the standard deviation of the amount of weight lost is the same.

       Or, being more practical, you might instead list rules of thumb, *based* on the ideal conditions, but that can be checked:
       - that the sample size of dieters from each program is at least 30, and
       - the ratio of largest sample standard deviation to smallest sd is no larger than 2.

       Noting you want the value of α to dictate the amount of area in the *right* tail, use
   (g)

```
qf(0.94, df1=3, df2=89)
```

   (h) (iv)
   (i) For both rows, the `lwr` confidence limit is negative and the `upr` confidence limit is positive, meaning that a zero-difference of means is entirely plausible. Thus, our evidence does not lead us to conclude a significant difference in means $\mu_1 - \mu_2$ in either case:

       where Group 1 is dieters on the Ornish plan, and Group 2 is Atkins dieters.
       where Group 1 is dieters on the Zone plan, and Group 2 is Atkins dieters.

4. (a) $\mathbf{H}_0$: There is no association between time of day and whether you yawn in response to someone else yawning.

   $\mathbf{H}_a$: There is association between time of day and whether you yawn in response to someone else yawning.

   (b) Its the ("2-5 pm", "No yawn") cell, with expected count 13.6.

   (c) That cell's contribution is approximately

$$\frac{(38-38.8)^2}{38.8} = 0.0165.$$

   (d) Yes. The smallest expected count is larger than 5.

   (e) The one with 3 degrees of freedom.

5. (a) The expected counts are as follows.

   no dropped calls: $(100)(0.11) = 11$
   one dropped call: $(100)(0.23) = 23$
   two dropped call: $(100)(0.27) = 27$
   three-or-more dropped calls: $(100)(0.39) = 39$

   (b)
```
pchisq(5.387, df=3, lower.tail=FALSE)
```

   (c) At none of the usual significance levels ($\alpha = 0.01, 0.05, 0.1$) do we have strong enough evidence to reject the null hypothesis. Thus, our data is consistent with the premise that the proportions of 0-, 1-, 2-, and 3-or-more-dropped calls are 0.11, 0.23, 0.27, and 0.39 respectively.

   (d) You might take 100 slips of paper and write "No dropped calls", "one dropped call", "two dropped calls", "3-or-more dropped calls" so that the number of each exactly matches the expected counts of part (a). You could place these slips in a bag and draw from that bag 100 times, each time placing the drawn slip back in the bag and mixing well before the next draw. On each draw, you would record the words on the slip so that, when finished, you had 100 results.