Answers (Problem set 12)

Online S520

1. Guess the correlation game

2. The psychologists Daniel Kahneman and Amos Tversky described the following situation:

   > The instructors in a flight school adopted a policy of consistent positive reinforcement recommended by psychologists. They verbally reinforced each successful execution of a flight maneuver. After some experience with this training approach, the instructors claimed that contrary to psychological doctrine, high praise for good execution of complex maneuvers typically results in a decrement of performance on the next try.[1]

   Is there a simpler explanation for the decreased performance following praise? What does this have to do with chapter 15?

   Solutions:

   > Regression is inevitable in flight maneuvers because performance is not perfectly reliable and progress between successive maneuvers is slow. Hence, pilots who did exceptionally well on one trial are likely to deteriorate on the next, regardless of the instructors' reaction to the initial success. The experience flight instructors actually discovered the regression but attributed it to the detrimental effect of positive reinforcement. This true story illustrates a saddening aspect of the human condition. We normally reinforce others when their behavior is good and punish them when their behavior is bad. By regression alone, therefore, they are most likely to improve after being punished and most likely to deteriorate after being rewarded. Consequently, we are exposed to a lifetime schedule in which we are most often rewarded for punishing others, and punished for rewarding.

3. Trosset chapter 15.7 exercise 4

   (a) We answer this question using the normal distribution:

   ```
   sister = c(69, 64, 65, 63, 65, 62, 65, 64, 66, 59, 62)
   brother = c(71, 68, 66, 67, 70, 71, 70, 73, 72, 65, 66)
   m = mean(brother)
   s = sd(brother)
   1 - pnorm(70, mean = m, sd = s)
   ```

   gives an estimated proportion of brothers at least 5' 10" of 0.36 or 36%.

   (b)
   ```
   r = cor(sister, brother)
   slope = r * sd(brother) / sd(sister)
   intercept = mean(brother) - slope * mean(sister)
   ```
   The regression line is $31.2 + 0.59\times$ sister's height. Plugging in 61 inches:

[1] Reprinted in *Judgement Under Uncertainty: Heuristics and Biases* (1982).

```
prediction = intercept + slope * 61
```

gives a predicted brother's height of 67.2 inches.

(c) One solution:

```
1 - pnorm(70, mean = prediction, sd = s*sqrt(1-r^2))
```

which gives 0.11 or 11%.

An alternative solution (which is more work) is:

```
n = 11
residuals = brother - (intercept + slope * sister)
SSE = sum(residuals^2)
prediction.error = sqrt(SSE/(n-2))
1 - pnorm(70, mean = prediction, sd = prediction.error)
```

which gives 0.12 or 12%, which isn't much different.

4. Trosset chapter 15.7 exercise 5

(a) This is just r^2, which is 0.31.

(b) We test the null hypothesis that the slope is 0:

```
std.error = sd(brother)/sd(sister) * sqrt((1-r^2)/(n-2))
t.stat = slope / std.error
# Two-sided P-value
2 * (1 - pt(abs(t.stat), df = n-2))
```

The P-value is 0.07, which is some evidence, but I wouldn't consider it "convincing." Of course, it's a small sample, so it's hard to obtain convincing evidence of anything.

(c)
```
slope - qt(0.95, df=n-2) * std.error
slope + qt(0.95, df=n-2) * std.error
```
A 90% confidence interval runs from 0.05 to 1.13.

(d) The width of the confidence interval is

$$2 \times q_t \times \frac{s_y}{s_x}\sqrt{\frac{1-r^2}{n-2}}.$$

Set this equal to 0.1 and make n the subject:

$$\begin{aligned} 0.1 &= 2 \times q_t \times \frac{s_y}{s_x}\sqrt{\frac{1-r^2}{n-2}} \\ \frac{0.05 s_x}{q_t s_y} &= \sqrt{\frac{1-r^2}{n-2}} \\ \frac{0.0025 s_x^2}{q_t^2 s_y^2} &= \frac{1-r^2}{n-2} \\ n &= \frac{q_t^2 s_y^2 (1-r^2)}{0.0025 s_x^2} + 2. \end{aligned}$$

Using the observed values of s_x, s_y, and r, and making the approximation $q_t \approx 1.96$, we get $n \approx 1189$.

5. Trosset chapter 15.7 exercise 8

   Solution:

   (a) This is a bad suggestion: for a start, Test 2 has lower scores than Test 1. We can assign a score using regression. The slope of the regression line is $0.5 \times 12/10 = 0.6$, and the intercept is $64 - 0.6 \times 75 = 19$. The prediction is $19 + 0.6 \times 80 = 67$.

   (b) This is a bad suggestion because of the regression effect. The regression effect states that individuals that do well on one test (in terms of standard units) will tend to do somewhat less well on another moderately correlated test (again, in terms of standard units). So somebody one standard deviation above the mean on one test will, on average, do somewhat less well on another test. Instead, we fit another regression line: the slope is $0.5 \times 10/12 = 5/12$, and the intercept is $75 - (5/12)64 = 48 + 1/3$. The prediction is $48 + 1/3 + (5/12)76 = 80$: in other words, half a standard deviation above the mean.