Statistics for Astronomers
Solutions to Homework #9

Prof. Sundar Srinivasan

February 2, 2021

Note: the solutions below use the script `hw9.py`.

1. (a) The script `hw9q1a` prints out

```
-----------------
Best-fit intercept:  7.14 ± 32.6
Best-fit slope:  2.38 ± 0.18
Correlation coefficient rho_bm = -0.99
-----------------
```

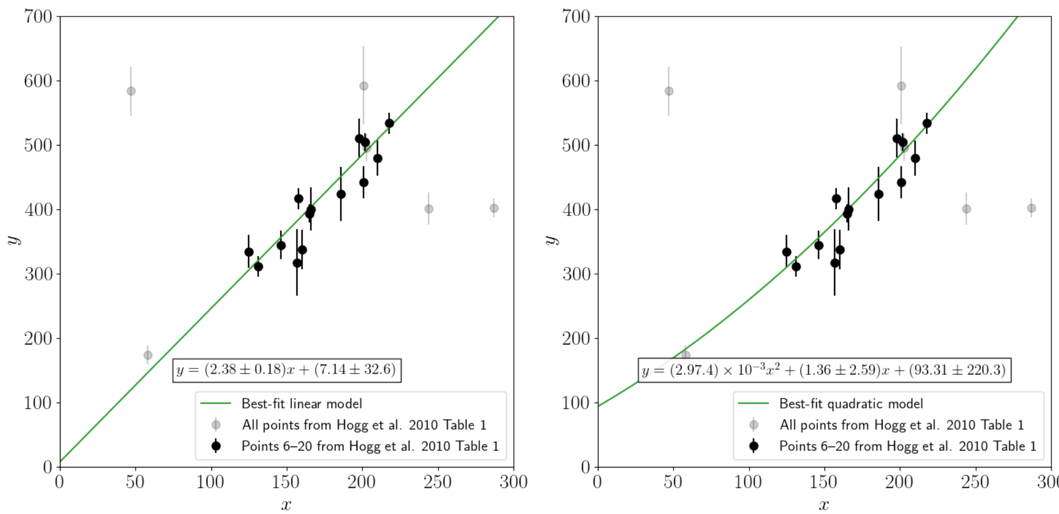

Figure 1: Linear (*left*) and quadratic (*right*) fits to the Hogg et al. data.

(b) The script `hw9q1b` prints

```
-----------------
Best-fit constant term b:  93.31 ± 220.3
Best-fit linear term m:  1.36 ± 2.59
Best-fit quadratic term q:  (2.936 ± 7.425)e-03
Correlation matrix:
4.85e+04 -5.69e+02 1.62e+00
-5.69e+02 6.70e+00 -1.92e-02
```

```
1.62e+00 -1.92e-02 5.51e-05
-----------------
```

Figure 1 compares the linear and quadratic fits to the data.

(c) The model with the lower reduced χ^2 (the χ^2 per degree of freedom) is preferred. The code `hw9q1c` prints

```
Reduced chi-square for linear model:  17.52
Reduced chi-square for quadratic model:  17.36
The reduced chi-squared for the quadratic model is lower.
The quadratic model is a better fit.
```

An improvement on the reduced χ^2 is the Bayesian Information Criterion, which is defined in terms of the best-fit χ^2 as

$$\mathrm{BIC} = p\ \ln N + \chi^2_{\mathrm{best}},$$

where $p =$ the number of parameters. The smaller the BIC, the better the model. The code prints out

```
BIC for linear model:  22.8
BIC for quadratic model:  25.28
The BIC for the linear model is lower, the linear model is better.
```

2. The code `hw9q2` computes 1000 bootstrap resamples of the uncensored data and computes a linear fit for each resample. If b is the vector of resampled intercepts and b_0 is the best-fit intercept, we define $\widehat{\sigma_b} = \sqrt{\mathrm{mean}[(b - b_0)^2]}$. A similar expression can be written down for $\widehat{\sigma_m}$. These are the bootstrap estimates for the standard deviations of the parameters. The code prints out

```
-----------------
Best-fit intercept:  213.27
Standard deviation in intercept:  86.59
Best-fit slope:  1.08
Standard deviation in slope:  0.49
-----------------
```

The standard deviations are higher than the estimates derived from the covariance matrix for the problem (see Figure 2 in Hogg et al.), but they are more realistic.