**Just Enough Data**

**Science and**

**Machine Learning**

Corrections for 08/01/2025

|  |  |  |  |
| --- | --- | --- | --- |
| **Printing No.** | **Page No.** | **Error** | **Correction** |
|  | 12 | Chapter 2:  Reads:  Figure 2.6 Histogram and PDF (overlaid in red) of the distribution of the heights of women; a normal distribution with parameters *μ* = 63.709 and *σ* = 2.696 is fitted to the data, *X ∼* N(63.709, 2.696). | Update numbers, should read:  Figure 2.6 Histogram and PDF (overlaid in red) of the distribution of the heights of women; a normal distribution with parameters μ = 161.820 and σ = 6.848 is fitted to the data, X ∼ N(161.820, 6.848). |
|  | 14 | Chapter 2:  Reads:  Histograms and PDFs (overlaid in red) of heart surgery survival rates (left), and HackerNews posts (right); an exponential distribution with *λ* = 223.28 is fitted to  the heart data, that is, *X ∼* Exp(222.28), while an exponential distribution with *λ* = 0.99 is fitted to the HackerNews data, that is, *X ∼* Exp(0.99). | Update numbers, should read:  Figure 2.7 Histograms and PDFs (overlaid in red) of heart surgery survival rates (left), and HackerNews posts (right); an exponential distribution with λ = 223.28 is fitted to the heart data, that is, X ∼ Exp(223.28), while an exponential distribution with λ = 0.99 is fitted to the HackerNews data, that is, X ∼ Exp(0.99). |
|  | 14 | Chapter 2:  Figure 2.7 The same plot is shown on both sides of the image. | Figure 2.7 [Corrected left plot] |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 15 | Chapter 2:  Reads:  **Example 2.8.** Consider two data sets, one recording the salaries of football players from FIFA 2019 (see Table 2.8), and the second recording the counts of words in the Brown corpus (see Table 2.9). A Pareto distribution is fitted to both data sets through maximum likelihood, resulting in the scale and shape parameters *xm* = 2*.*46 and *α* = 0*.*13, respectively, for the first data set representing the salaries of football players, and the scale and shape parameters *xm* = 72*.*192 and *α* = 1*.*0435, respectively, for the second data set  representing the counts of words. That is, *X ∼* Pareto(2*.*46*,* 0*.*13), where *X* is a random variable indicating the salary of a football player, and *X ∼* Pareto(72*.*192*,* 1*.*0435), where *X* is a random variable indicating the count of a word, respectively. The histograms and the PDFs (overlaid in red) for the probability of a salary of a football player (left) and for the count of a word (right) are shown in Figure 2.8. | Updated numbers, should read:  **Example 2.8.** Consider two data sets, one recording the salaries of football players from FIFA 2019 (see Table 2.8), and the second recording the counts of words in the Brown corpus (see Table 2.9).  A Pareto distribution is fitted to both data sets through maximum likelihood, resulting in the scale and shape parameters xm = 1000 and α = 0.59, respectively, for the first data set representing the salaries of football players, and the scale and shape parameters xm = 72 and α = 1.042, respectively, for the second data set representing the counts of words.  That is, X ∼ Pareto(1000, 0.59), where X is a random variable indicating the salary of a football player, and X ∼ Pareto(72, 1.042), where X is a random variable indicating the count of a word, respectively.  The histograms and the PDFs (overlaid in red) for the probability of a salary of a football player (left) and for the count of a word (right) are shown in Figure 2.8. |
|  | 15 | Chapter 2:  Reads:  Table 2.8 Sample of FIFA 2019 data showing the salary information in Euros for each player with the top 20 removed. | Removed “with the top 20 removed”, should read:  Table 2.8 Sample of FIFA 2019 data showing the salary information in Euros for each player. |
|  | 16 | Chapter 2:  Reads:  Figure 2.8 Histogram plot of FIFA 2019 player salaries (left), and a further plot of the top 1,500 words and their frequency counts in the Brown corpus (right); a Pareto distribution with parameters *xm* = 2.46 and *α* = 0.13 is fitted to the football salary data, that is, *X ∼* Pareto(2.46, 0.13), while a Pareto distribution with *xm* = 72.192 and *α* = 1.0435 is fitted to the word count data, that is, *X ∼* Pareto(72.192, 1.0435), where *X* is a random variable indicating the count of a word in the corpus. | Updated numbers, should read:  Figure 2.8 Histogram plot of FIFA 2019 player salaries (left), and a further plot of the top 1,500 words and their frequency counts in the Brown corpus (right); a Pareto distribution with parameters xm = 1000 and α = 0.59 is fitted to the football salary data, that is, X ∼ Pareto(1000, 0.59), while a Pareto distribution with xm = 72. and α = 1.042 is fitted to the word count data, that is, X ∼ Pareto(72, 1.042), where X is a random variable indicating the count of a word in the corpus. |
|  | 108 | Chapter 4:  Figure 4.30 Incorrect y-axis label | Figure 4.30 [Correct y-axis label in plot -> 'WCSS' to 'SSE'] |
|  | 162 | Chapter 5:  Reads:  To find the exponent *α* that satisfies the 80-20 rule, we first need to solve *SX*(*x*) = 0*.*8 to get an expression for *x*, and then solve *fX*(*x*) = 0*.*2 to get a second expression for *x*. We then equate both expressions, as given by  (0*.*8(*xmin*)*α*)*−* 1*α* = (0*.*2*α* (*xmin*)*α*)*−* 1*α*+1*,*  and solve this equation for *α*. We can replace 0*.*8 and 0*.*2 in the preceding equation by any fractions as long as they add to 1, and obtain *α*, for example, for the 90-10 rule. | Replace the highlighted text with the following text, from "To find the exponent ..." until " for the 90-10 rule."  Should read (latex format):  In order to find the exponent $\alpha$ that satisfies the 80-20 rule, we first assert that $F\_X(x\_{80}) = 0.8$ and solve for $x\_{80}$, noting that $S\_X(x\_{80}) = 1 - F\_X(x\_{80}) = 0.2$. We can then find the proportion of wealth above $x\_{80}$ which should be equal to $0.8$. The general solution for $\alpha$, is then given by  \begin{displaymath}  \alpha = \frac{\log(p)}{\log(p) - \log(1-p)},  \end{displaymath}  with $0 < p < 0.5$, noting that for $p$ in that range $\alpha > 1$ holds.  We replace $0.8$ and $0.2$ in the equation above by any fractions as long as they add to $1$, and obtain $\alpha$, for example, for the 90-10 rule.  Output: |