

Statistics Project

Statistics course Project

Joel Rodriguez | Probability and Statistics for Engineers | 4/15/2019

Z23402515

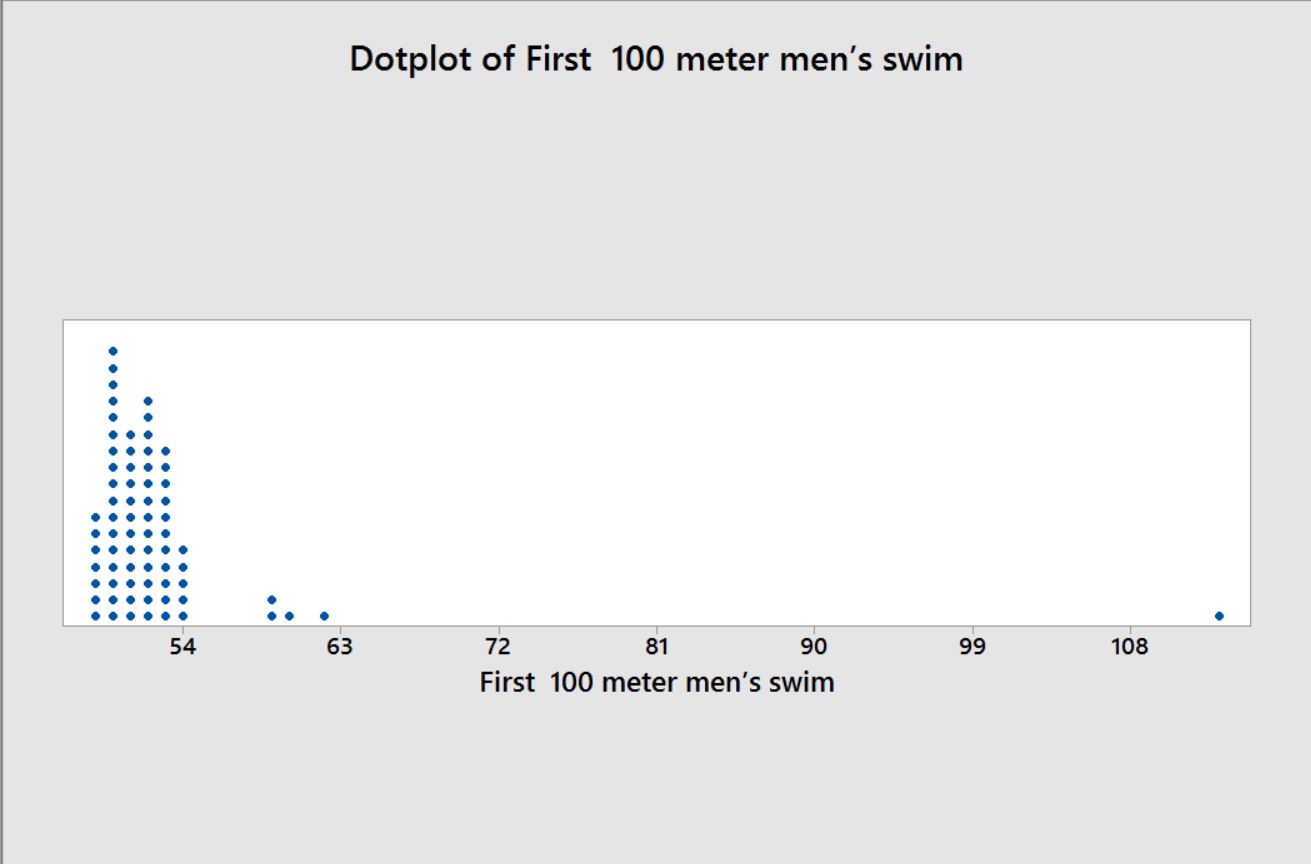
**6-24. In the 2000 Sydney Olympics, a special program initiated by IOC president Juan Antonio Samaranch allowed developing countries to send athletes to the Olympics without the usual qualifying procedure. Here are the 71 times for the first round of the 100 meter men’s swim (in seconds).**

60.39 49.93 53.40 51.82 50.46 51.34 50.28 50.19 52.1450.56 52.72 50.95 49.74 49.16 52.57 52.53 52.09 52.40 49.75 54.06 53.50 50.63 51.93 51.62 52.58 53.55 51.07 49.76 49.73 50.90 59.26 49.29 52.78 112.72 49.79 49.83 52.43 51.28 52.22 49.76 49.70 52.90 50.19 54.33 62.45 51.93 52.24 52.82 50.96 48.64 51.11 50.87 52.18 54.12 50.49 49.84 52.91 52.52 50.32 51.52 52.0 52.85 52.24 49.45 51.28 49.09 58.79 49.74 49.32 50.62 49.45

1. **Find the sample mean and sample standard deviation of these 100-meter swim times.**

|  |  |
| --- | --- |
| [[1]](#footnote-1)First 100 meter men’s swim. | |
|  |  |
| Mean | **52.64760563** |
| Standard Error | 0.908597355 |
| Median | 51.34 |
| Mode | 50.19 |
| Standard Deviation | **7.655977397** |
| Sample Variance | 58.6139899 |
| Kurtosis | 55.94693946 |
| Skewness | 7.142038881 |
| Range | 64.08 |
| Minimum | 48.64 |
| Maximum | 112.72 |
| Sum | 3737.98 |
| Count | 71 |
| Largest (1) | 112.72 |
| Smallest (1) | 48.64 |
| Confidence Level (95.0%) | 1.812140284 |

1. **Construct a dot diagram of the data.**

**[[2]](#footnote-2)**

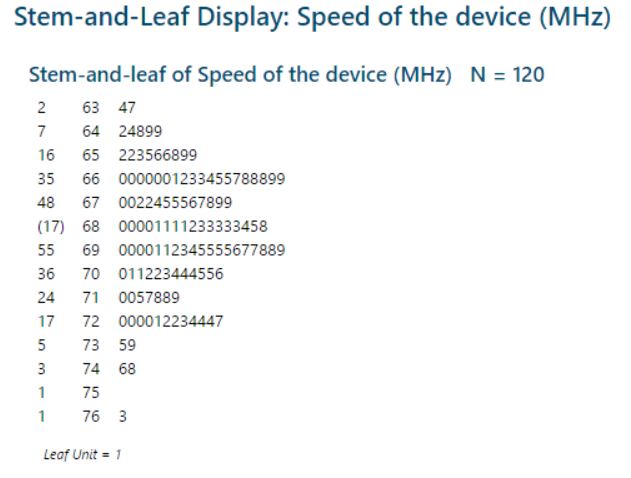
1. **Comment on anything unusual that you see.**

This graph shows that most swimmers swim 100 meters not faster than 54 seconds, it also shows that four of those swimmers can swim 100 meters faster than the average swimmer. Finally, it also shows that one swimmer who is an outlier, meaning this swimmer can swim a bit more than twice faster than the average swimmer.

**6-42. A semiconductor manufacturer produces devices used as central processing units in personal computers. The speed of the devices (in megahertz) is important because it determines the price that the manufacturer can charge for the devices. The following table contains measurements on 120 devices.**

680 669 719 699 670 710 722 663 658 634 720 690 677 669 700 718 690 681 702 696 692 690 694 660 649 675 701 721 683 735 688 763 672 698 659 704 681 679 691 683 705 746 706 649 668 672 690 724 652 720 660 695 701 724 668 698 668 660 680 739 717 727 653 637 660 693 679 682 724 642 704 695 704 652 664 702 661 720 695 670 656 718 660 648 683 723 710 680 684 705 681 748 697 703 660 722 662 644 683 695 678 674 656 667 683 691 680 685 681 715 665 676 665 675 655 659 720 675 697 663

1. **Construct a stem-and-leaf diagram for these data and comment on any important features that you notice.**

****

**[[3]](#footnote-3)**

1. **Compute the sample mean, the sample standard deviation, and the sample median.**

|  |  |
| --- | --- |
| *Speed of the device[[4]](#footnote-4)* | |
|  |  |
| Mean | **686.775** |
| Standard Error | 2.343160453 |
| Median | **683** |
| Mode | 660 |
| Standard Deviation | **25.66803672** |
| Sample Variance | 658.8481092 |
| Kurtosis | -0.19983589 |
| Skewness | 0.379864154 |
| Range | 129 |
| Minimum | 634 |
| Maximum | 763 |
| Sum | 82413 |
| Count | 120 |
| Largest (1) | 763 |
| Smallest (1) | 634 |
| Confidence Level (95.0%) | 4.639691724 |

**What percentage of the devices has a speed exceeding 700 megahertz?**

To answer this question, we already know that our sample consists of 120 devices. We also know that only 35 devices out of those 120 have a speed exceeding 700 megahertz.

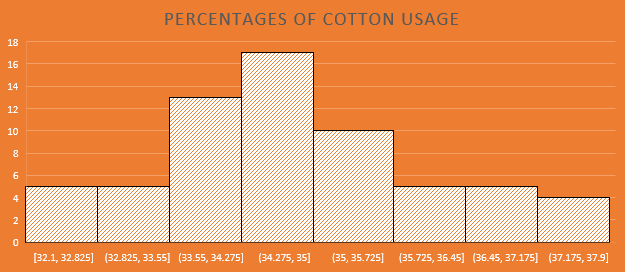
|  |  |  |  |
| --- | --- | --- | --- |
| [[5]](#footnote-5)719 | 724 | 702 | 710 |
| 720 | 724 | 702 | 715 |
| 720 | 727 | 703 | 717 |
| 720 | 735 | 704 | 718 |
| 720 | 739 | 704 | 718 |
| 721 | 746 | 704 |  |
| 722 | 748 | 705 |  |
| 722 | 763 | 705 |  |
| 723 | 701 | 706 |  |
| 724 | 701 | 710 |  |

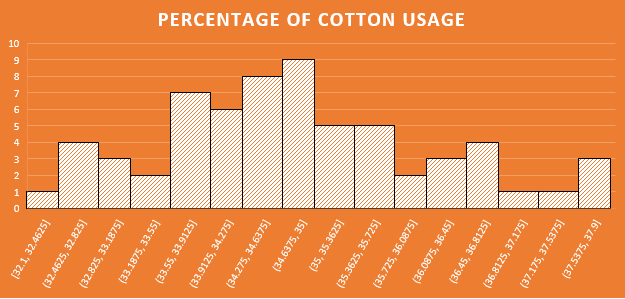
So, at this point we can just apply simple arithmetic;

%= (35\*100)/120= **29.17%**

**6-52. Construct histograms with 8 and 16 bins for the data in Exercise 6-32. Compare the histograms. Do both histograms display similar information?**

34.2 37.8 33.6 32.6 33.8 35.8 34.7 34.6 33.1 36.6 34.7 33.1 34.2 37.6 33.6 33.6 34.5 35.4 35.0 34.6 33.4 37.3 32.5 34.1 35.6 34.6 35.4 35.9 34.7 34.6 34.1 34.7 36.3 33.8 36.2 34.7 34.6 35.5 35.1 35.735.1 37.1 36.8 33.6 35.2 32.8 36.8 36.8 34.7 34.0 35.1 32.9 35.0 32.1 37.9 34.3 33.6 34.1 35.3 33.5 34.9 34.5 36.4 32.7

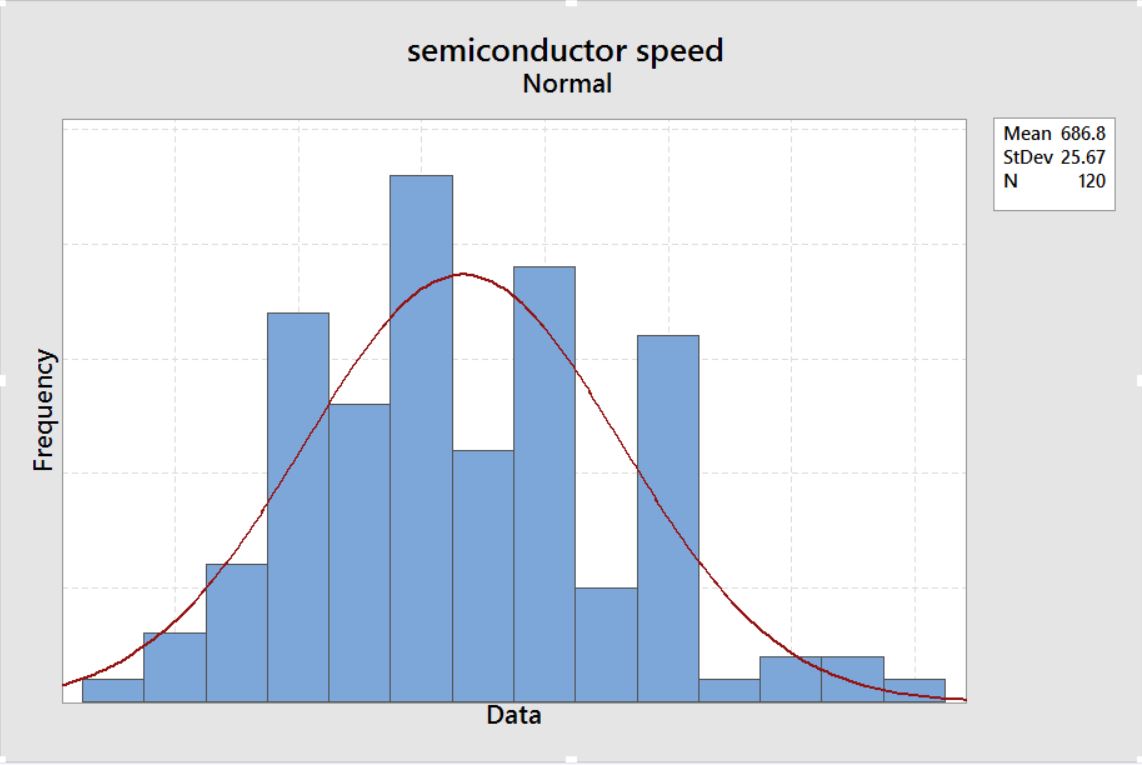


[[6]](#footnote-6)

These two histograms represent the entire data which is the percentages of cotton used to manufacture men’s shirt. The difference between the two histograms is that the first one with only eight bins shows how the data is normally distributed going from the lowest points to the highest points, but at the same time it is not detailing the data as much as the second graph is.

The downside to the second diagram is that it shows the data more spread out and it makes it very difficult to see the normal distribution curve as good as it is presented in the first graph.

**6.58 Construct a histogram for the data in Exercise 6-42. Comment on the shape of the histogram. Does it convey the same information as the stem-and leaf display?**

**[[7]](#footnote-7)**

The data from the stem-and-leaf plot shows a more evenly distributed data. The stem leaf (shown above) presents the data positively skewed, whereas this histogram shows the data skewed to the left. The stem-and-leaf plot in this case shows a more accurate transition compared to the one shown in the histogram.

**6-67. Using the data from Exercise 6-22 on cloud seeding,**

**Unseeded:**

81.2 26.1 95.0 41.1 28.6 21.7 11.5 68.5 345.5 321.2 1202.6 1.0 4.9 163.0 372.4 244.3 47.3 87.0 26.3 24.4 830.1 4.9 36.6 147.8 17.3 29.0

**Seeded:**

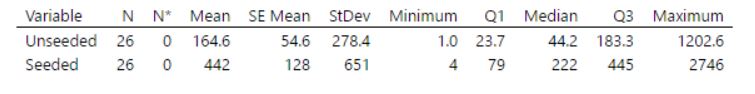
274.7 302.8 242.5 255.0 17.5 115.3 31.4 703.4 334.1 1697.8 118.3 198.6 129.6 274.7 119.0 1656.0 7.7 430.0 40.6 92.4 200.7 32.7 4.1 978.0 489.1 2745.6

Find the median and quartiles for the unseeded cloud data.

1. **Find the median and quartiles for the seeded cloud data**.

From Minitab, we get the following data

[[8]](#footnote-8)



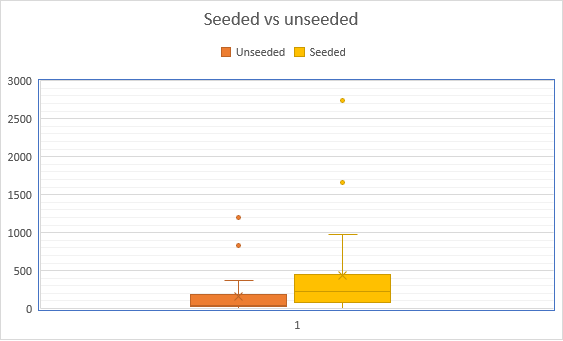
So,

The median: 44.20

The first quartile: 23.70

The Third quartile: 183.3

1. **Make two side-by-side box plots, one for each group on the same plot.**

[[9]](#footnote-9)

1. **Compare the distributions from what you can see in the side-by-side box plots.**

By looking at these two plots side by side, we can see that the data from “seeded” contains much bigger data that the one from unseeded. Both data also show outliers that seem to be equally proportional to the size of the data. Both data also seem to be right skewed with two outliers each.

**6-68. Using the data from Exercise 6-24 on swim times, find the median and quartiles for the data.**

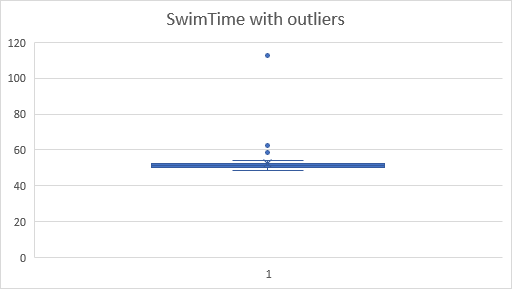
|  |  |
| --- | --- |
| ***Swim Time data[[10]](#footnote-10)*** | |
| Mean | 52.64760563 |
| Standard Error | 0.908597355 |
| Median | **51.34** |
| Mode | 50.19 |
| Standard Deviation | 7.655977397 |
| Sample Variance | 58.6139899 |
| Kurtosis | 55.94693946 |
| Skewness | 7.142038881 |
| Range | 64.08 |
| Minimum | 48.64 |
| Maximum | 112.72 |
| Sum | 3737.98 |
| Count | 71 |
| Largest (1) | **112.72** |
| Smallest (1) | **48.64** |
| first quartile | **50.06** |
| 2nd quartile | **51.34** |
| Third quartile | **52.575** |
| Confidence Level (95.0%) | 1.812140284 |

**Median:** 51.34

**1st quartile:** 50.06

**3rd quartile:** 52.58

1. **Make a box plot of the data.**



1. **Repeat (a) and (b) for the data without the extreme outlier and comment.**

The data will now be generated without the following five pieces of data;

|  |  |  |
| --- | --- | --- |
| 58.79 | 60.39 | 112.72 |
| 59.26 | 62.45 |  |

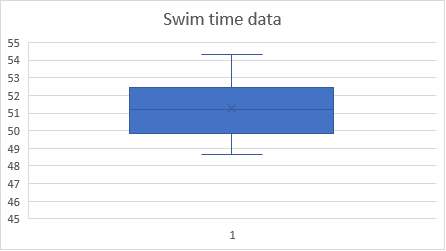
|  |  |
| --- | --- |
| ***Swim Time without outliers[[11]](#footnote-11)*** | |
| Mean | 51.31892308 |
| Standard Error | 0.17402933 |
| Median | 51.28 |
| Mode | 49.45 |
| Standard Deviation | 1.403069317 |
| Sample Variance | 1.96860351 |
| Kurtosis | -0.986619151 |
| Skewness | 0.220027136 |
| Range | 5.24 |
| Minimum | 49.09 |
| Maximum | 54.33 |
| Sum | 3335.73 |
| Count | 65 |
| Largest (1) | 54.33 |
| Smallest (1) | 49.09 |
| first Q | 49.8625 |
| Third Q | 52.4225 |
| Confidence Level (95.0%) | 0.347663554 |

**Median:** 51.28

**1st quartile:** 49.09

**3rd quartile:** 52.42

Now, this is the boxplot\* based on the data without the five outliers indicated above.

**[[12]](#footnote-12)**

My comment on the boxplot data: The data without the outliers, shows a normal and very consistent distribution. This data obviously has no outliers because it was plotted like that intentionally.

**C. Compare the distribution of the data with and without the extreme outlier.**

The first set of data indicates that there are five outliers present in the data, which were indicated above. The data without outliers does not look as consistent and normally distributed as the second set of data, without outliers, does.

On the other hand, the second set of data without outliers shows a more consistent picture and also shows a set of data that is normally distributed with a bell shape and so on.

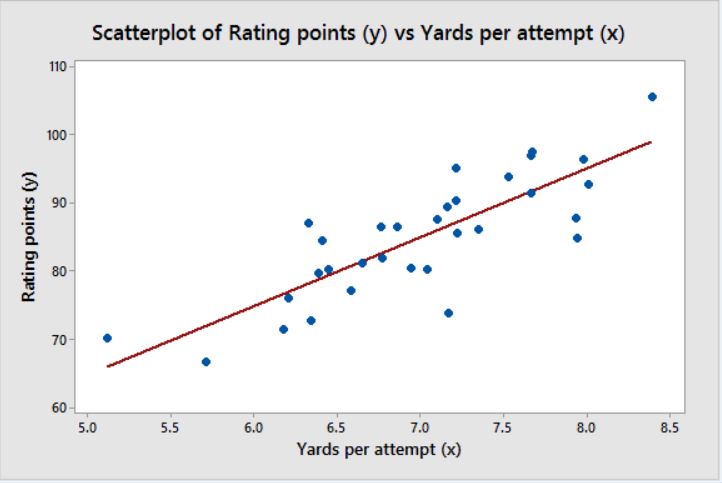
11-5. See Table E11-1 for data on the ratings of quarterbacks for the 2008 National Football League season (*The Sports* *Network*). It is suspected that the rating (*y*) is related to the average number of yards gained per pass attempt (*x*).





|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Yards per attempt (x)* | |  | *Rating points (y)[[13]](#footnote-13)* | |
|  |  |  |  |  |
| Mean | 6.9978125 |  | Mean | 84.815625 |
| Standard Error | 0.128663833 |  | Standard Error | 1.584171386 |
| Median | 7.07 |  | Median | 85.7 |
| Mode | 7.66 |  | Mode | 86.4 |
| Standard Deviation | 0.727832551 |  | Standard Deviation | 8.961426635 |
| Sample Variance | 0.529740222 |  | Sample Variance | 80.30716734 |
| Kurtosis | 0.142682976 |  | Kurtosis | -0.197162486 |
| Skewness | 0.326815945 |  | Skewness | 0.013453228 |
| Range | 3.27 |  | Range | 39 |
| Minimum | 5.12 |  | Minimum | 66.5 |
| Maximum | 8.39 |  | Maximum | 105.5 |
| Sum | 223.93 |  | Sum | 2714.1 |
| Count | 32 |  | Count | 32 |
| Largest (1) | 8.39 |  | Largest (1) | 105.5 |
| Smallest (1) | 5.12 |  | Smallest (1) | 66.5 |
| Confidence Level (95.0%) | 0.262411618 |  | Confidence Level (95.0%) | 3.230938843 |

1. Calculate the least squares estimates of the slope and intercept. What is the estimate of σ2? Graph the regression model.



Based on all of the data generated above

σ2= SSe/(n-2)

σ2= (817.0568)/(32-2)

σ2= 27.23522667 or **27.235.**

1. Find an estimate of the mean rating if a quarterback averages 7.5 yards per attempt.

E(Y, given 7.5) =β0+β1(7.5)

= 14.19548847+10.0917446(7.5) = 89.88357299 or **89.884**

1. What change in the mean rating is associated with a decrease of one yard per attempt?

ΔE (Y given x) = β1(Δx)

= 10.0917446 (-1) =-10.0917446 or **-10.092**

1. To increase the mean rating by 10 points, how much increase in the average yards per attempt must be generated?

ΔE = (Y given x)/(β1(Δx))

=10/10.0917446

= 0.990908946 or **0.99091**

(e) Given that *x* = 7.21 yards, find the fitted value of *x* and the corresponding residual.

X=7.21

y-hat= 14.19548847+10.0917446 (7.21)

= 86.95696704

=86.9570

**References**

1. Data pasted from the textbook. Montgomery, Douglas C., and George C. Runger. *Applied statistics and probability for engineers, (With CD)*. John Wiley & Sons, 2007.
2. Levine, David M., et al. *Statistics for managers using Microsoft Excel*. Vol. 660. Upper Saddle River, NJ: Prentice Hall, 1999.
3. Meyer, Ruth, and David Krueger. *Minitab guide to statistics*. Prentice Hall PTR, 2001.

1. **Data generated through the “analysis data feature on excel.** [↑](#footnote-ref-1)
2. **Data was generated on Minitab.** [↑](#footnote-ref-2)
3. **This Data was generated on Minitab.** [↑](#footnote-ref-3)
4. **Data generated on excel.** [↑](#footnote-ref-4)
5. **Data pasted from the textbook. Montgomery, Douglas C., and George C. Runger. *Applied statistics and probability for engineers, (With CD)*. John Wiley & Sons, 2007.** [↑](#footnote-ref-5)
6. **Data generated on excel.** [↑](#footnote-ref-6)
7. **Data generated on Minitab.** [↑](#footnote-ref-7)
8. **Data generated on Minitab** [↑](#footnote-ref-8)
9. **Data generated on excel.** [↑](#footnote-ref-9)
10. **Data generated on excel.** [↑](#footnote-ref-10)
11. **Data generated on excel.** [↑](#footnote-ref-11)
12. **Data was generated on excel.** [↑](#footnote-ref-12)
13. **These tables were generated on excel.** [↑](#footnote-ref-13)