Solutions to the Problems in Elementary Statistical Quality Control by John T. Burr

John Weatherwax

To my family.

Chapter 2 (Characteristics of Data)

Questions 2.1-2.8

In the R code for this chapter we compute these statistics on this given data. With a computer these type of exercises are quite simple.

Questions 2.9-2.10

Please see the output of the R code that comes with this chapter.

Questions 2.11-2.12

These would be worked in the same way as Question 2.13.

Question 2.13

Running the R code for this problem I get

```
[1] 2.504280808 0.001283146
```

Question 2.14

Running the R code for this problem I get

```
[1] 0.833876923 0.001718536
```

Question 2.15

Running the R code for this problem I get

```
[1] 52.860000  1.002724
```

Question 2.16

If we assume that any tubes with values less than or equal to 66 (the lower specification limit) or larger than or equal to 72 (the upper specification limit) we would find three such tubes which is a 50% out of specification rate.

Question 2.17

If the upper specification limit is 100 then I get *none* of the samples violate this condition giving a zero violation percentage.

Question 2.18

If we assume that a value of heat less than or equal to 70 (the lower specification limit) or larger than or equal to 90 (the upper specification limit) are out of specification then I find 12 such data points which is a 15% out of specification rate.

Question 2.19

If we take the given numbers as the specification limits I find two samples are too low and no samples are too large. These will give about a 2% out of specification rate.

Question 2.20

If we take the given numbers as the specification limits, I find one sample is too low and no samples are too large. These will give about a 1.5% out of specification rate.

Question 2.21

From the given data I find only two samples with a value below 50 to give a 2% out of specification rate.

Chapter 3 (Simple Probability and Probability Distributions)

Question 3.1

In the R code for this question I simulate an unbiased coin and compute the empirical probability that the coin flip will land heads from the from the number of heads observed When you run that code I plot this probability and you can see it converging to the true value of $p_0 = 0.5$. Let x be the random variable representing the number of heads observed in n coin flips. Then as x is a binomial random variable the variance of this random variable is $np_0(1 - p_0)$. Evaluating this I compute $\sigma = 5$.

Question 3.2

We do the same as in the previous question but now using $p_0 = 1/6$ we compute $\sigma = 3.72678$ for the standard deviation of the number of sixes (in 100 rolls).

Questions 3.3-3.5

The number of nonconforming found when inspecting n is a binomial random variable. Using the command `dbinom` we can easily compute these numbers.

Question 3.6

The probability that we draw zero or one bad gauges from the two we draw is given by a hyperbolic random variable with form

$$P(d) = \frac{\binom{1}{d}\binom{7}{2-d}}{\binom{8}{2}},$$

for $0 \leq d \leq 1$. Evaluating this gives

```
[1] 0.75 0.25
```

for $P(0)$ and $P(1)$ respectively.

Question 3.7

Here we get

```
[1] 0.625 0.375
```

Question 3.8

Here we get

```
[1] 0.53571429 0.42857143 0.03571429
```

Question 3.9

Here we get

```
[1] 0.3571429 0.5357143 0.1071429
```

Question 3.10

The number X of nonconforming parts is a binomial random variable which has a well defined mean and variance. Computing these using the numbers given I get

```
[1] "mean= 40.000000; sigma= 6.000000; 3*sigma= 18.000000"
```

Question 3.11

These can be computed with the `R` function `choose` and `factorial`. We find these are given by

```
[1] 210 210 5040
```

Question 3.12

Part (a): This would be $\binom{52}{4}$.

Part (b): The rank is the "number" on the card. There are only thirteen ways to get a four hand set of cards all of the same rank. This hand is also called "four-of-a-kind".

Part (c): There are 13 ways to pick the rank that the three cards share. There are $\binom{4}{3}$ ways to pick these three cards. There are then $52 - 4 = 48$ other cards that can be the last card. Thus there are

$$13\binom{4}{3}(48) = 2496\,,$$

hands of this type. This hand is also called "three-of-a-kind".

Part (d): We would divide the above two numbers by the number in Part (a). This gives

```
[1] 4.801921e-05 9.219688e-03
```

Problem 3.13

Errors on a page are often modeled as Poisson with the given average as the parameter. Using the `R` function `ppois` we can compute these two probabilities as

```
[1] 0.6626273 0.1205129
```

Problem 3.14

These are given by

```
[1] 0.5152161 0.2935616
```

Problem 3.15

These are given by

```
[1] 0.6766764 0.2706706 0.5939942
```

Problem 3.16

These are given by

```
[1] 0.09475787 0.17834131
```

Problem 3.17

This is given by 0.5898322. The true distribution is a hypergeometric distribution. We can use a binomial probability distribution as an approximate distribution with $p = 100/1000 = 0.1$ and $n = 5$. Then with the approximate distribution the probability requested in Part (a) would be given by 0.5904900 very close to the exact probability.

Problem 3.18

This is given by 0.9043821. The correct distribution is the binomial. We can approximate this distribution using a Poisson approximate with $c_0 = np = 0.1$. Under this approximate the probability requested in Part (a) would be given by 0.9048374 very close to the exact probability.

Problem 3.19

Part (a-b): The exact distribution is a binomial with $n = 100$ and p_0 as given and the probability requested is given by

```
[1] 0.81856680: 0.16404144 0.01627265
```

Part (c): We can approximate this binomial distribution using a Poisson distribution. Under that approximate the probabilities above would be

```
[1] 0.81873075 0.16374615 0.01637462
```

Problems 3.20 – 3.22

For each of these problems we will compute the value of $c_0 + 3\sigma_c$ and then evaluate the given probability This is quite easy as all of the densities are Poisson. The results of this calculation for each of the problems can be found in the `R` code for this chapter and we get

```
[1] 0.004024267 0.004533806 0.003834707
```

Problem 3.23

The true density in this case is hypergeometric but we can do much of the same for this problem as we did in the previous ones. Following the same steps as above we find this probability given by 0.00836578.

Problem 3.24

We can use the `dbinom` command to find

```
[1] 0.4096 0.4096 0.1536 0.0256 0.0016
```

Problem 3.25

We can use the `dbinom` command to find

```
[1] 0.59049 0.32805 0.07290 0.00810 0.00045 0.00001
```

Problem 3.26

We can use the `dhyper` command for this problem where I find

```
[1] 0.47619048 0.47619048 0.04761905
```

Problem 3.27

We can also use the `dhyper` command for this problem where I find

```
[1] 0.23809524 0.53571429 0.21428571 0.01190476
```

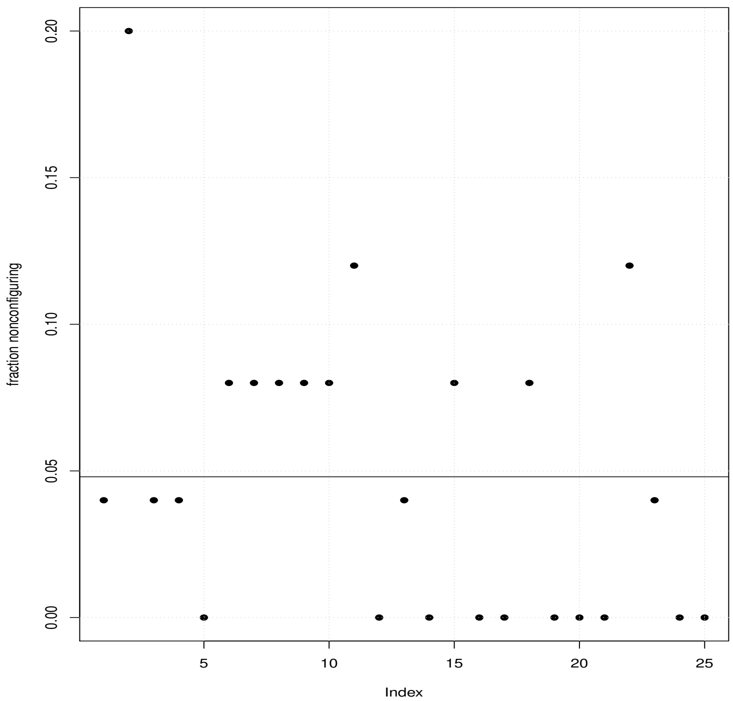

Figure 1: The control plot for Problem 2.

Chapter 4 (Control Charts in General)

Question 4.2

We plot the individual samples and the full sample mean in Figure 1. There we see that the second sample is significantly larger than the others. A cause for this sample should be investigated.

Question 4.3

We plot the individual samples and the full sample mean in Figure 2 (left). There we see that most of the samples seem to be clustered around the mean indicating that this process is looks to be in control.

Question 4.4

We plot the individual samples for $\bar{x}$ and the range Figure 2 (right). There we see that the third range sample and the 12th samples of $\bar{x}$ are very large seem and should be investigated.

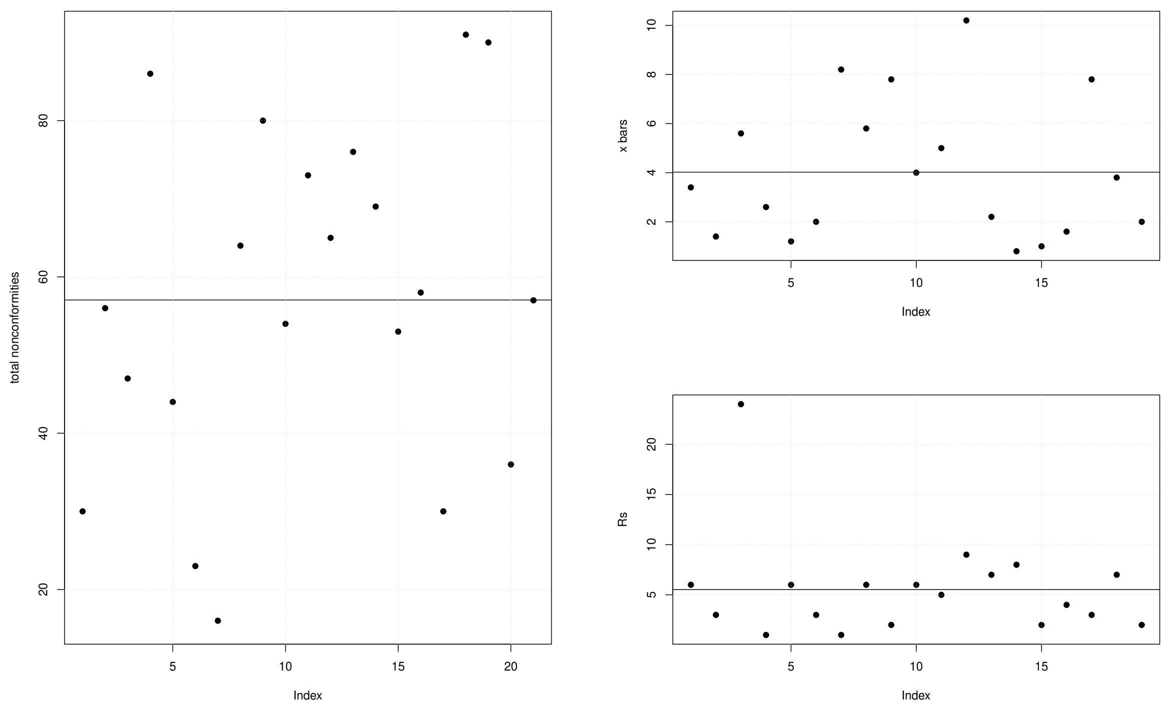

Figure 2: **Left:** The control plot for Problem 3. **Right:** The control plot for Problem 4.

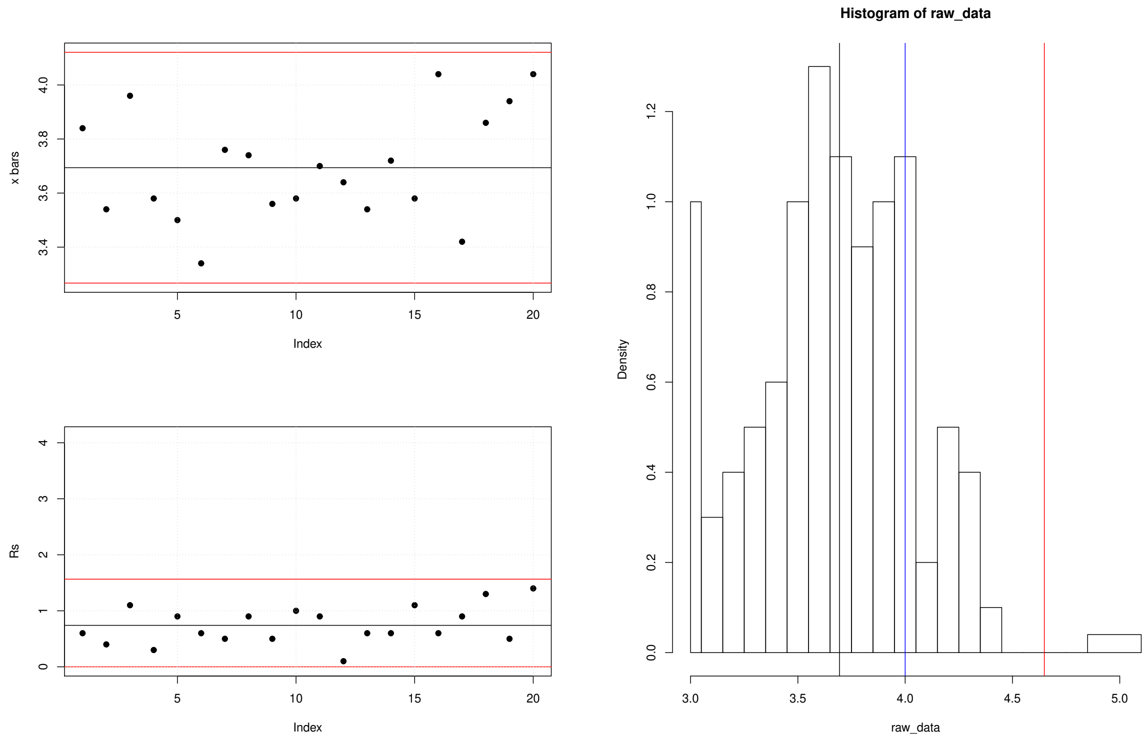

Figure 3: **Left:** The control plot for Question 1. **Right:** A histogram of the raw data for Question 1.

Chapter 6 (Control Charts for Measurements: Process Control)

Question 6.1

As each of the $\bar{x}$ is made up of samples with five elements we have

$$A2 = 0.577$$
$$D3 = 0$$
$$D4 = 2.115\,,$$

using the table in the back of the book. Using these values we get the control charts given in Figure 3 (left). Notice that this process appears to be in control.

If we then take $\mu = \bar{x}$ and estimate σ using $\bar{R}$ we find $\sigma = 0.318143$. Then the limits of the x values that make up the sample are given by $\mu - 3\sigma$ and $\mu + 3\sigma$. For these two values I get

```
[1] "confidence limits x: (2.739572, 4.648428)"
```

Notice that the right-most limit is significantly larger than the maximum specification limit

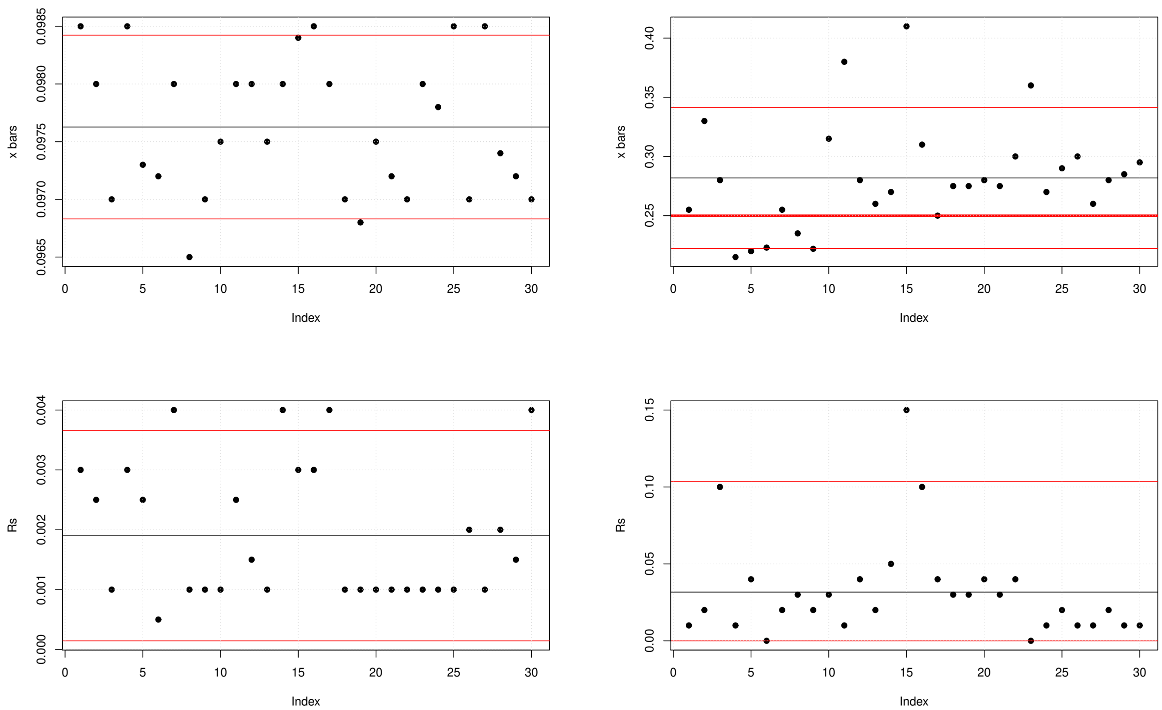

Figure 4: **Left:** The control plot for Question 2. **Right:** The control plot for Question 3.

of 4.0. We can place these limits on the histogram of the individual x measurements see Figure 3 (right). Note that from this plot we see that one sample has a very large value (of 4.9) that is even beyond the 3σ limits. Note that also many of the individual oat cereal packages had more than the allowed 4.0 (the maximum specification value and drawn as a blue vertical line in the histogram) of moisture content. To make this process better match what is required we would need to lower the process average value μ.

Question 6.2

In Figure 4 (left) I plot the control curves for $\bar{x}$ and R for this process. Notice that we have many points beyond the control limits in both $\bar{x}$ and R. This process does not appear to be in control. If we desire to hold to the low side of the diameter it looks like we accomplish that in that more points are below the value of $\bar{x}$ than are above it. We seem to have have several (5-6 values) that are above the upper control limit of $\bar{x}$.

Question 6.3

In Figure 4 (right) I plot the control curves for $\bar{x}$ and R for this process. As is standard, in the top plot I have drawn the control plot for $\bar{x}$. In that plot I've also drawn the upper and lower control limits in light red lines and the the minimum specification value (of 0.25) in a thicker red line. Notice that we see several points in both control plots that indicate that the process is out-of-control.

As there are several value of $\bar{x}$ that are beyond the minimum specification of 0.25 it looks like our minimum specification will *not* be met in all cases. To correct for this we need to raise the process mean μ. This can also be observed by assuming that the process *was* in control (even though it looks like it is not) and then estimating μ and σ from the given data. The three sigma limits from these are given by

```
[1] "confidence limits x: (0.197613, 0.366053)"
```

Notice that the lower confidence limit of 0.198 is quite a bit smaller than the minimum specification of 0.25 and in fact there is a almost a 13% probability that a sample would fall below that value.

As $n = 2$ if we are given $\bar{x}$ and R for the first sample we can assume that $x_2 > x_1$ and solve

$$\bar{x} = 0.255 = \frac{1}{2}(x_1 + x_2)$$
$$R = 0.01 = x_2 - x_1 \,,$$

to get x_1 and x_2. We find $x_1 = 0.26$ and $x_2 = 0.25$.

Question 6.4

In Figure 5 (left) I plot the control curves for $\bar{x}$ and s for this process. In both plots I've also drawn the upper and lower control limits in light red lines. Notice that we see several points in both control plots that indicate that the process is out-of-control. I'm not sure we would expect control as potentially this experiment is looking to find additives that help to reduce corrosion. Thus some additives might work better than others and by doing the experiments we would find out which ones. The points (or more specifically the underlying element) far from the control center are ones that are significantly different than the others and should be studied further.

Question 6.5

In Figure 5 (right) I plot the control curves for $\bar{x}$ and R for this process. I do this for the three different runs of 20 samples each. In each run I estimate $\bar{x}$ and $\bar{R}$. We have the following observations about each run.

- For the first run: A few points outside of the control bounds for $\bar{x}$ and three points close to the upper control limit for R. Too many points are too far from the estimated value of $\bar{x}$ to ignore.

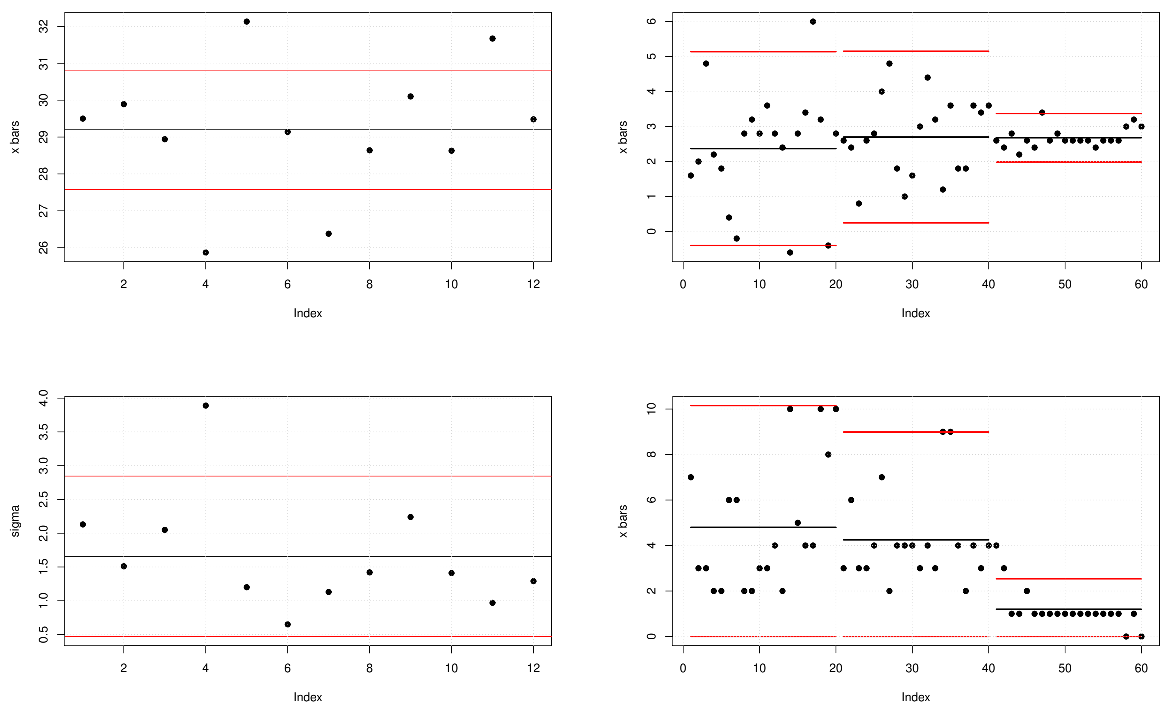

Figure 5: **Left:** The control plot for Question 4. **Right:** The control plot for Question 5.

- For the second run: Some adjustments are made and control limits are recomputed. Notice that now the process means seem in more control but that the control chart for the range has two sample that are on the upper control limit. Thus it seems like the process range is not under control.
- For the third run: More adjustments are made and both control limits are much tighter. The process seems to now be in control.

We can see the progression of the estimates of $\bar{x}$ and $\bar{R}$ over each batch

```
[1] "xbar_1= 2.370000; rbar_1= 4.800000"
[1] "xbar_2= 2.700000; rbar_2= 4.250000"
[1] "xbar_3= 2.680000; rbar_3= 1.200000"
```

Question 6.6

In Figure 6 (left) I plot the control curves for $\bar{x}$ and R for this process. As we are not told the number of samples in each batch used to compute the sample $\bar{x}$ I assumed $n = 2$. Note that in both of these plots the process appears to be in control. In the control plot of the process mean, the fact that the lower and upper confidence intervals are so far from the sample values of $\bar{x}$ indicates that my assumption that $n = 2$ is probably incorrect.

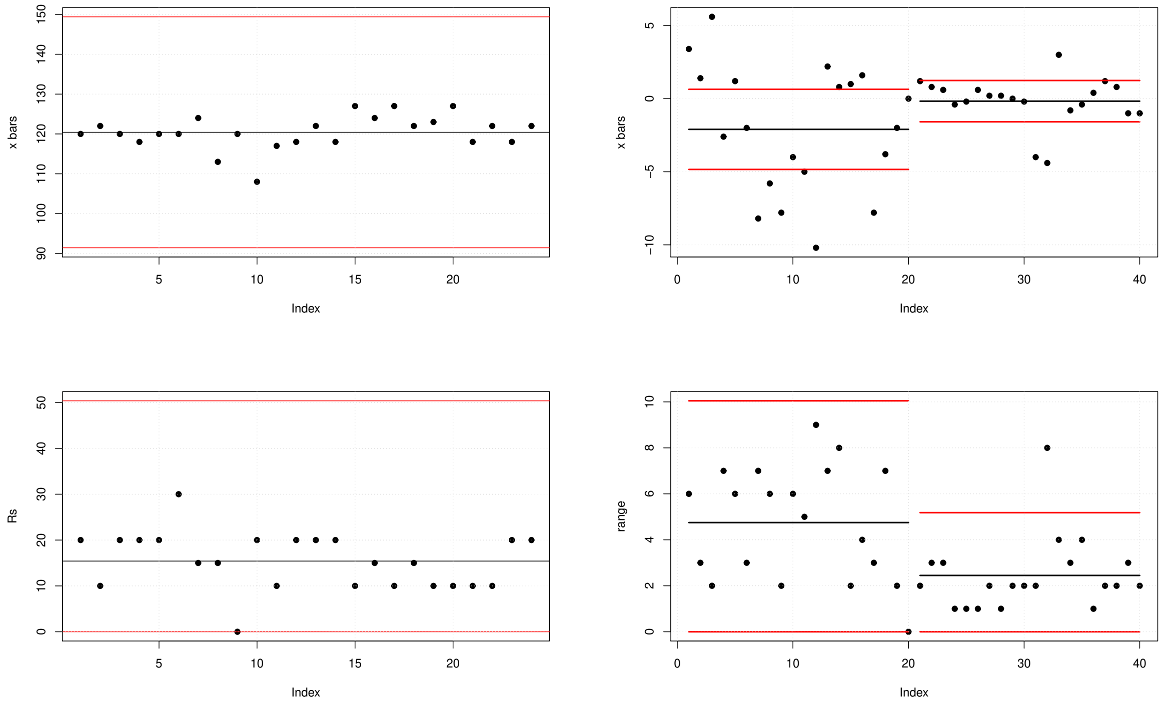

Figure 6: **Left:** The control plot for Question 6. **Right:** The control plot for Question 7.

Question 6.7

In Figure 6 (right) I plot the control curves for $\bar{x}$ and R for this process.

Chapter 7 (Process Capability)

Notes on the Text

Specifications or Tolerance Limits: Case (2)

If we find we have dye that is 97.8% pure which is under our specification limit of 98%. If we take a volume X of this dye and mix it with another volume X of dye with a purity of p to have the combined total meet our specification we need to have

$$\frac{0.978X + pX}{2X} > 0.98\,.$$

Solving this for p gives $p > 0.982$ which is how the book gets the number quoted.

Manufacturing Tolerances vs. Specification limits

From my reading of the book, here is how I would describe some of the terms introduced.

Specification Limits: are how accurate the parts need to be for them to be correctly used downstream of their manufacturing i.e for them to be used for their intended purpose.

Manufacturing Tolerances: would be constructs that specify how accurately a part needs to be made. One would want the manufacturing limits to be at least as stringent as the specification limits but sometimes they might be *more* stringent. By this I mean that the part is required to be manufactured to a tighter specification than is actually required. In the book this is argued to not be a good idea.

Question 7.1

Running the exercise 6.1 we see that the process appears to be in control and we should certainly be able to compute the process capability indices.

The definition of C_P is

$$C_P = \frac{\text{US} - \text{LS}}{6\sigma_X}\,, \tag{1}$$

Here US and LS are the upper and lower specification limits and we can estimate σ_X using either

$$\sigma_X = \frac{\bar{R}}{d_2}$$
$$\sigma_X = \frac{\bar{s}}{c_4}\,.$$

From the problem statement we have LS = 0 and US = 4 and using the first formula above I estimate $\sigma_X = 0.318143$. These together give $C_P = 2.0954958$.

To compute C_{PK} recall that its definition is

$$C_{PK} = \min\left(\frac{\bar{x} - \text{LS}}{3\sigma_X}, \frac{\text{US} - \bar{x}}{3\sigma_X}\right) . \tag{2}$$

Using the numbers for this problem I compute $C_{PK} = 0.320611$.

Next to compute P_{PK} I estimate $\sigma_{\bar{X}}$ using the standard deviation of the $\bar{x}$ values. I then multiply this by $\sqrt{n}$ to get σ_X. This procedure gives $P_{PK} = 1.479004$.

I think the take away from these calculations is that both C_P and P_{PK} are "large" (larger than one) indicating that we are able to meet our specifications. The fact that C_{PK} is "small" (less than one) indicates that the mean of the process is too large relative to the value of US and should be controlled lower.

Question 7.2

Modifying the previous exercise to compute the requested process specifications metrics I find

```
[1] "C_p= 0.000162"
[1] "x_bar= 2.680000; LS= 1.176000; US= 1.176500; C_pk= -0.971428"
[1] "sigma_xbar= 0.285804; sigma_x= 0.639078; P_PK= 0.000130"
```

Notice that the process appears to be in control but based on the metrics above is unable to meet the required specifications.

Question 7.3

Modifying the previous exercise to compute the requested process specifications metrics I find

```
[1] "C_p= 1.805867"
[1] "x_bar= 4.879062; LS= 2.000000; US= 6.000000; C_pk= 1.012132"
[1] "sigma_xbar= 0.664333; sigma_x= 1.150658; P_PK= 0.579379"
```

Given the question statement I was not exactly sure how to compute the specification limits LS and US in this case.

Question 7.4

Modifying the previous exercise to compute the requested process specifications metrics I find

```
[1] "C_p= 0.522159"
[1] "x_bar= 78.276923; LS= 70.000000; US= 90.000000; C_pk= 0.432187"
[1] "sigma_xbar= 3.313283; sigma_x= 5.738775; P_PK= 0.580844"
```