Homework 3

Here we present solutions to the problems posted in the third homework assignment. The solutions and related commentary are put in italics. Remember that problems can have several different but correct ways of solving them.

Multiple choice questions

1. Which one of the following statements is true:

   A the area within 2 units of 0 under the standard Normal curve is .4772;

   B the proportion of values in a Normal distribution within two standard deviations of the mean is .4772;

   ***C** 2.5% of the values in a Normal distribution are at least 1.96 standard deviations above the mean;

   D 5% of the values in a Normal distribution are at least 1.96 standard deviations above the mean;

   E one standard deviation of the Normal distribution is .3413 above the mean.

   Solution: *Since 95% of the values following the normal model are within 1.96 standard deviation of the mean, 5% are either above or below one standard deviation of the mean and by symmetry there is 2.5% of the values that are above one standard deviation of the mean.*

2. In a population of values whose frequency distribution is the standard Normal distribution, 68

   ***A** 68% of the values in a population whose frequency distribution is Normal fall within one standard deviation of the mean;

   B 32% of the values in a population whose frequency distribution is Normal are less than 1;

   C 68% of the values in a population whose frequency distribution is Normal are greater than -1 or less than +1;

   D the standard deviation of a Normal distribution is the value assumed by 68% of the values;

   E the difference between plus and minus one standard deviation in a Normal distribution is .68.

   Solution: *If a variable X follows a general normal distribution with the mean μ and standard deviation σ, then, by standardization, the variable $Z = (X - \mu)/\sigma$ follows the standard normal distribution. So saying that the values of Z are between -1 and +1 is equivalent to saying that the values of X fall within one standard deviation of the mean.*

3. The weekly drink sales in a licensed premises tend to follow a Normal distribution with mean £5,000 and standard deviation £500. This means that

***A** sales will fall below £4,500 in about 16% of weeks;

B sales will be above £4,500 in about 68% of weeks;

C sales will fall between £4,500 and £ 5,500 in about 84% of weeks;

D sales will be £1,000 below average in about 5% of weeks;

E sales will deviate from £5,000 by about £500 in about 5% of weeks.

Solution: *It follows from the normal model that sales will be between £4,500 and £5,500 about 68% of the time. By symmetry they have to be below £4,500 about 32/2% percent of the time.*

4. The time to complete each of 100 bank counter transaction was measured. The average transaction time was 5 minutes and the standard deviation was 1 minute. Assuming that the Normal model for statistical variation applies to bank counter transaction times, which one of the following is true:

A 68% of transactions take more than 5 minutes;

B 68% of transaction times take more than 6 minutes;

C 95% of transaction times take between 4 and 6 minutes;

D 5% of transaction times take more than 7 minutes;

***E** 95% of transaction times take between 3 and 7 minutes.

Solution: *About 95% of values fall within 2 standard deviation of the mean, which translates to falling between 3 and 7 minutes.*

5. Till reconciliations at the end of a day's trading in a retail outlet invariably show small cash errors, either a surplus or a shortfall. Records over a long period suggest that the average error is a cash shortfall of £1, with a standard deviation of £3. Assuming that the Normal model for statistical variation applies, the chances of getting a shortfall in excess of £10 are

A approximately 95%;

B approximately 5%;

***C** less than 3 in 1,000;

D the same as the chances of a surplus of £10;

E the same as the chances of not getting a shortfall in excess of £10.

Solution: *Data following the normal model fall within 3 standard deviations more than 99.7% of the time, so they would fall above 3 standard deviations of the mean less 0.3% of the time.*

6. Given specification limits of 70 to 90 for a process whose mean may vary between 80 and 85 and whose standard deviation may vary between 4 and 8:

A the non-conformance rate improves as the standard deviation is increased from 4 to 8;

B the non-conformance rate improves as the mean is increased from 80 to 85;

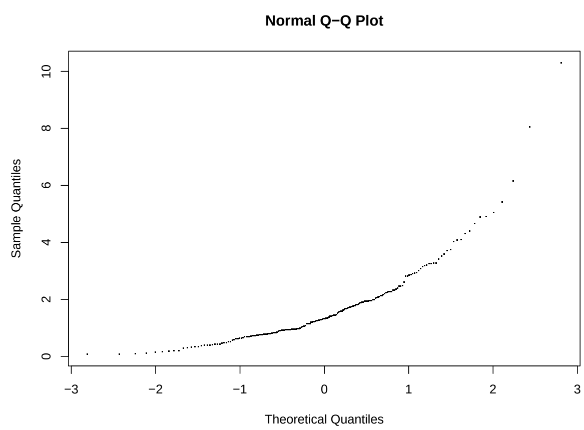
***C** the non-conformance rate exceeds 1.24% for all parameter values in the given ranges;

D the non-conformance rate never exceeds 1.24% for all parameter values in the given ranges;

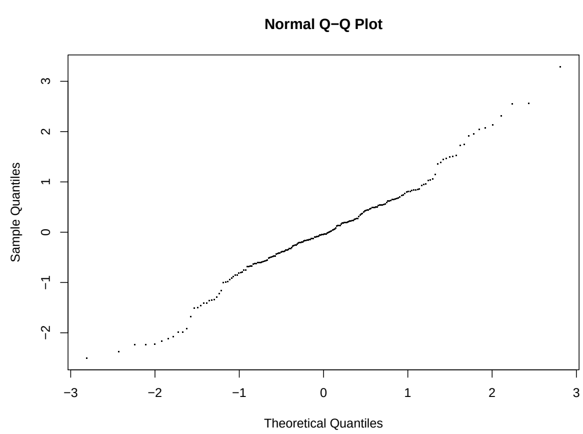
E the non-conformance rates exceed the specification limits for all parameter values in the given ranges.

Solution: *The smallest non-conformance rate is attained for the case where the mean is at 80 and the standard deviation is at 4. For this case, the non-conformance rate is equivalent to the percentage of values following the normal model that fall outside of the range 2.5 standard deviations of the mean. The percentage of such values is (1-0.9875807)100*

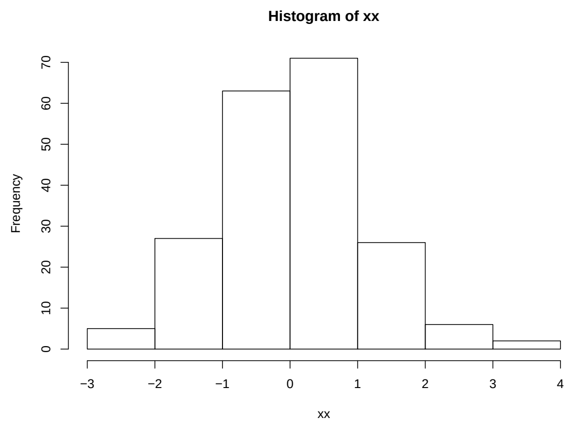
7. For which of the following data that are represented on the graphs, could be some doubts about applying the normal model assumptions:

***A** 

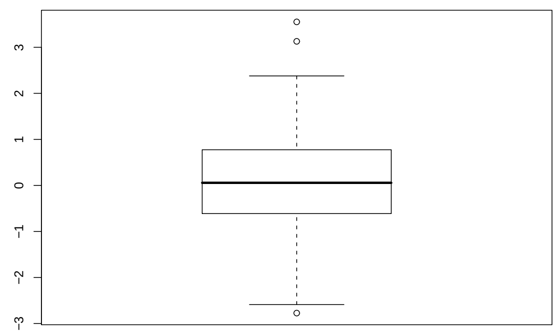
;

B 

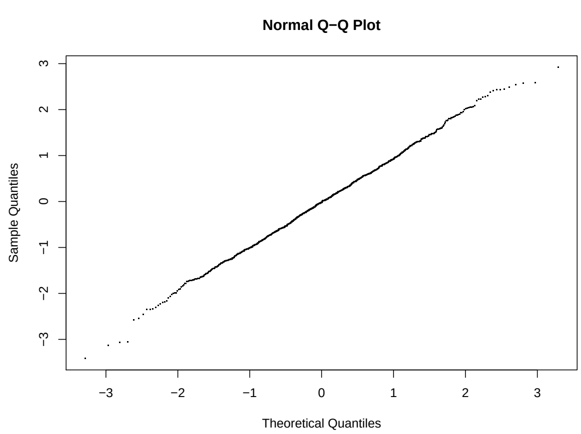
;

C 

;

D 

;

E 

.

Solution: *The qq-normal plot that curves accross entire range of values exposes evident deviation from the normal model.*

8. The Normal diagnostic plot or qq-normal plot is

***A** a scatter plot of actual data values against corresponding theoretical values that conform to the Normal model;

B the plot most normally used when data have too large spread to plot a histogram;

C a histogram which shows by the pattern of rectangles whether the plotted data conform to the Normal model or not;

D a histogram with a Normal curve superimposed;

E a plot of sample data in which the chance variation has been normalised so that it conforms to the Normal model.

Solution: *The qq-normal plot represents scattered points where the first coordinate is the data value and the second the value that would have the same percentage data below it if the percentage would be according to the normal model.*

Problems

1. During the lecture, the non-conformance rate for the first press has been computed based on the normal model for the tennis ball diameter data. The non-conformance rates for other presses have been reported but calculations have not been presented. Do the calculation on your own and confirm that the reported values are correct.
Solution: *The following R-code computes the means and standard deviations for the tennis ball data (there are also given in the textbook, page 102)*

```
Tennis=read.csv("../Data/Tennis.csv")
mean(Tennis)

# Press.1  Press.2  Press.3  Press.4
# 90.52151 88.28495 85.80645 76.80108

sd(Tennis)
#  Press.1   Press.2   Press.3   Press.4
# 8.305782  8.929725 12.138176  6.994067
```

Using these value one can compute the standardized tolerance limits $(70-91)/8.3 = -2.530$, $(70-88)/8.9 = -2.022$, $(70-86)/12.1 = -1.322$, $(70-77)/7 = -1$ *and* $(100-91)/8.3 = 1.084$, $(100-88)/8.9 = 1.348$, $(100-86)/12.1 = 1.157$, $(100-77)/7 = 3.285$*, respectively. Now, using the tables or computer software one can obtained the frequencies being within standardized tolerance limits*

```
1-pnorm(1.08)+pnorm(-2.53)
# 0.1457742
1-pnorm(1.34)+pnorm(-2.02)
# 0.1118144
1-pnorm(1.15)+pnorm(-1.32)
# 0.2184894
1-pnorm(3.28)+pnorm(-1)
# 0.1591743
```

2. We have seen in the lecture that after adjusting the mean value for each press and by slightly limiting the percentage of production assigned to the third press, it is possible to get the non-conformance rate within 10%. Here is another approach to improve quality standards. Suppose that by discarding the balls from the matrix edges of the third press (26 of them), one can reduce standard deviation for this press by 20%. Assume that additionally to setting the mean value to 85 for each press, the management decides also to discard these balls but to keep 25% contribution of each press to the total production. Compute the overall non-conformance rate that follows from such an approach. Discuss pros and cons of both approaches.
Solution: *From the condition in the problem we have reduced the standard deviation of the third press to* $12.1 * .8 = 9.68$. *The mean value for each of the press has been set to 85. The following code present the fast computation of non-conformance rates (otherwise proceed as shown in the previous solution by using tables)*

```
1-pnorm(100,85,8.3)+pnorm(70,85,8.3)
# 0.0707266
1-pnorm(100,85,8.9)+pnorm(70,85,8.9)
# 0.09191272
1-pnorm(100,85,9.68)+pnorm(70,85,9.68)
# 0.1212407
1-pnorm(100,85,7)+pnorm(70,85,7)
# 0.03212457
```

for overall non-conformance rate 7.9%. *As we see the non-conformance is much improved over the other approach which targeted 10% non-conformance rate. The cons are that the press number 3 has to be treated in a special way affecting the uniformity of the production flow. This may require a special training of workers and thus may cost more.*

3. A company produces power circuits with a nominal output of 115[V]. Quality control records show that the actual output follows a normal distribution with mean 115[V] and standard deviation 0.8[V]. The company has been approached by a Japanese company interested in purchasing large volumes of these units for their European subsidiary. Their purchasing policy requires that before taking on a new supplier, the supplier must show that 99% of his production lies within 2[V] of the nominal output. Can the company meet these standards?
Solution: *The standardization leads to the following conformance rate for the company*

$$(113 - 115)/0.8 < Z < (117 - 115)/0.8,$$
$$-2.5 < Z < 2.5.$$

From the tables we get 0.9938-0.0062=0.9876 which is slightly below the requirements set by the Japanese partner.

4. The amount of oxygen dissolved in rivers and streams depends on the water temperature and on the amounts of decaying organic matter from natural processes or human disturbances that are present in the water. The Council on Environmental Quality (CEQ) considers a dissolved oxygen content of less than 5 milligrams per litre (mg/L) of water to be undesirable because it is unlikely to support aquatic

life. At a certain location of Shannon River the daily content of oxygen is normally distributed with mean equal to 6.3[mg/L] and a standard deviation of 0.6[mg/L]. The CEQ is making control measurements at randomly selected days at this location. What percentage of these measurements would be considered undesirable by the CEQ?

Solution: *The standardization leads to the following*

$$\begin{aligned}(5-6.3)/0.6 &> Z\\ -2.17 &> Z.\end{aligned}$$

From the tables we get 0.015, i.e. 1.5% of the measurements will be undesirable.

5. Daily changes in the FTSE 100 were displayed in Figure 1.10 of Chapter 1. The mean change and standard deviation for 1996 were 2.00 and 20.82, respectively. On how many trading days would you expect a daily change of more than 50, more than 75, more than 100, assuming the normal model for variation in daily change. Note that there were 265 trading days on the London Stock Exchange in 1996.

   For the second half of 1997, the mean and standard deviation of daily changes in the FTSE 100 were 3 and 56, respectively. On how many trading days would you expect a daily change of more than 50, more than 75, more than 100, assuming the normal model for variation in daily change. Note that there were 130 trading days on the London Stock Exchange in the second half of 1997.

   Comment on the effect of increasing volatility.

   Solution: *The standardization leads to the following for 1996*

$$\begin{aligned}(50-2)/20.82 &< Z\\ 2.31 &< Z,\\ (75-2)/20.82 &< Z\\ 3.51 &< Z,\\ (100-2)/20.82 &< Z\\ 4.71 &< Z.\end{aligned}$$

   From the tables we get 0.011, 0.0002, 1.24e-06, leading to the expected number of days 2.915, 0.53, 0.0003.

   For the second half of 1997

$$\begin{aligned}(50-3)/56 &< Z\\ 0.84 &< Z,\\ (75-3)/56 &< Z\\ 1.29 &< Z,\\ (100-3)/56 &< Z\\ 1.73 &< Z.\end{aligned}$$

   From the tables we get 0.20, 0.1, 0.04, i.e. there should be about 26 days with a daily change above 50, about 13 days with a daily change more than 75, and about 5-6 days with a daily change more than 100.

We observe a dramatic change in the number of days with a high daily change and this can be attributed almost solely to the change in standard deviation (volatility).