Solutions to Homework 3

Solutions to Multiple Choice Questions

1. **C** – Since 95% of the values following the normal model are within 1.96 standard deviations of the mean, 5% are either above or below 1.96 standard deviations of the mean. By symmetry this implies there are 2.5% of the values that are above 1.96 standard deviations of the mean.
2. **A** – Since the population follows the standard normal model, it implies that that the mean is 0 and the standard deviation 1. Therefore 68% of the values will fall within one standard deviation of the mean, which will be the range from -1 to 1.
3. **A** – Since the mean is £5,000 and the standard deviation £500 we know that 68% of the values will be within one standard deviation of the mean i.e. between £4,500 and £5,500. This implies that 32% of the values will occur outside this range, and by symmetry this means that 16% of the values will be below £4,500 and the other 16% will be above £5,500.
4. **E** – We know that 95% of values will fall within 2 standard deviations of the mean, which translates to falling between 3 and 7 minutes.
5. **C** – Data following the normal model fall within 3 standard deviations more than 99.7% of the time, so they would fall outside 3 standard deviations of the mean 0.3% of the time. By symmetry, the values would fall above 3 standard deviations of the mean 0.15% of the time.
6. **C** – The smallest non-conformance rate is attained when the mean is 80 and the standard deviation is 4. In this situation, 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 1.24%.
7. **A** – The qq-plot that curves across the entire range of values exposes evident deviation from the normal model. If it followed the normal model, it should be a straight line.
8. **A** – 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 was according to the normal model.

Solutions to Problems

Answer to Question 1[1]

```
// USING THE STATISTICAL PROGRAM 'R' GET THE DATA FROM THE CSV FILE, THEN GET THE MEANS
tennis_full=read.csv("G:/Business Statistics/Tennis.csv")
mean(tennis_full)

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

// GET THE STANDARD DEVIATIONS
sd(tennis_full)

# Press.1       Press.2       Press.3       Press.4
# 8.305782      8.929725      12.138176     6.994067
```

The customer limits are between 70 and 100 limits. However we must standardise the limits with respect to the mean and standard deviation of each press. We find the standardised limits for each press using the following equation:

$$Z = \frac{y - \mu}{\sigma}$$

Z	=	Standard Normal Variable
y	=	Lower Limit (70) or Upper Limit (100)
μ	=	Mean value for Press
σ	=	Standard Deviation for Press

The standardised lower and upper limits for press 1 will therefore be given by:

$$SLL_1 = \frac{70 - 90.52151}{8.305782} \qquad SUL_1 = \frac{100 - 90.52151}{8.305782}$$

$$SLL_1 = -2.47 \qquad SUL_1 = 1.14$$

The standardised limits for the other presses are calculated similarly. Shown below in the table are the mean, standard deviation, and the standardised limits for all the presses.

Press No.	Mean	Standard Deviation	Standardised Lower Limit	Standardised Upper Limit
1	90.52151	8.305782	-2.47	1.14
2	88.28495	8.929725	-2.05	1.31
3	85.80645	12.138176	-1.30	1.17
4	76.80108	6.994067	-0.97	3.32

[1] Comments are preceded by "//" and are in green font. Outputs are preceded by "#" and are shown in red font.

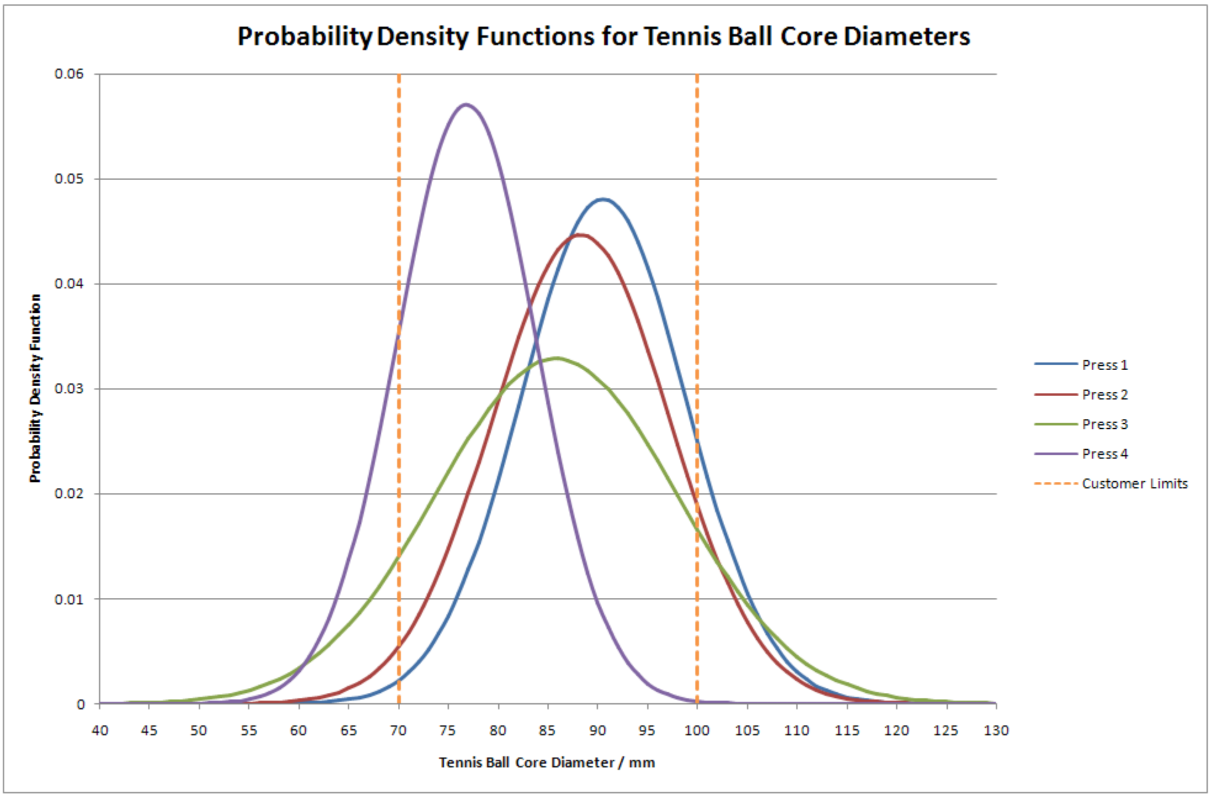
For each press we want to determine the non-conformance rate. This will correspond to the sum of the area under the curve to the left of the lower limit plus the area under the curve to the right of the upper limit. Since we know the standardised limits, we can use the pnorm() function in 'R'. For the first press the area under the curve from minus infinity to the lower limit is given by pnorm(-2.47). The area under the curve from minus infinity to the upper limit is given by pnorm(1.14). Therefore, since the area under the total curve from minus infinity to plus infinity is one, the total area outside the standardised limits of -2.47 and 1.14 is given by:

```
// FOR THE FIRST PRESS, GET THE NON-CONFORMANCE RATE
1 – pnorm(1.14) + pnorm(-2.47)

# 0.1338988
```

The non-conformance rates for the other presses are calculated similarly. Shown below in the table are the non-conformance rates for each machine, and the number of defects per 10,000 tennis balls that this is equivalent to.

Press No.	Non-Conformance Rate	Defects per 10,000 Tennis Balls
1	13.4%	1,340
2	11.5%	1,150
3	21.8%	2,180
4	16.6%	1,660



Answer to Question 2

The new means and standard deviations for the presses are shown below. We reduce the standard deviation of the third press by 20%. This means the standard deviation of press three is now:

(12.1)*(0.8) = 9.710541

Press No.	Mean	Standard Deviation
1	85	8.305782
2	85	8.929725
3	85	9.710541
4	85	6.664067

We use the following 'R' code to generate the new non-conformance rate for the first press:

```
// FOR THE FIRST PRESS, GET THE NEW NON-CONFORMANCE RATE
1 – pnorm(100, 85, 8.305782) + pnorm(70, 85, 8.305782)

# 0.0709229
```

The non-conformance rates for the other presses are calculated similarly. Shown below in the table are the new non-conformance rates for each machine, and the number of defects per 10,000 tennis balls that this is equivalent to.

Press No.	Non-Conformance Rate	Defects per 10,000 Tennis Balls
1	7.1%	710
2	9.3%	930
3	12.2%	1,220
4	2.4%	240

Since we are told that each press will still contribute 25% to total production, the total non-conformance rate is given by:

$$\frac{7.1\%}{4} + \frac{9.3\%}{4} + \frac{12.2\%}{4} + \frac{2.4\%}{4} = 7.75\%$$

As we can see the non-conformance rate is significantly improved over the other approach, which targeted a 10% non-conformance rate. One downside is that press 3 will now only produce 160 tennis balls per batch. Another problem is that press 3 is still expected to contribute to 25% of total production but will now only produce 160 tennis balls per batch, unlike the other presses which still produce 186 tennis balls per batch. For every batch produced by press 3, we are missing out on 26 tennis balls, compared to its siblings. So if we wanted press 3 to contribute to 25% of total production then it is important to note the following.

We look at the production rate during a single shift, and we want each machine to produce the same amount of tennis balls per shift. We define T_i as the time per shift that press i is operational and C_i is the cycle-time of press i (i.e. the time to produce one batch of tennis balls). If we divide T_i by C_i we obtain the number of batches produced per shift, and if we multiple the result by the batch size for the press we obtain the number of tennis balls each press produces during the shift. Therefore we want the following statement to be true:

$$\frac{186T_1}{C_1} = \frac{186T_2}{C_2} = \frac{160T_3}{C_3} = \frac{186T_4}{C_4}$$

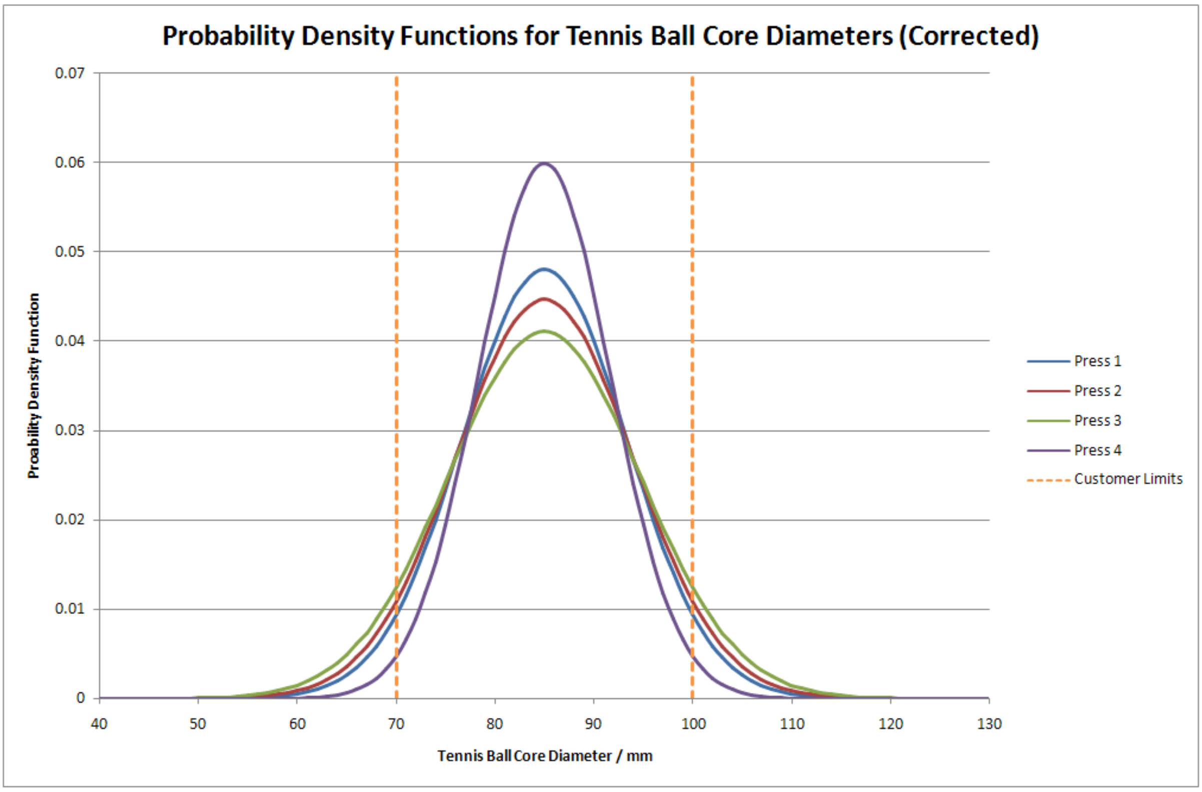
$$\frac{T_1}{C_1} = \frac{T_2}{C_2} = \frac{160T_3}{186C_3} = \frac{T_4}{C_4}$$

If we assume the cycle times for all the machines to be constant and non-adjustable, the equation reduces to:

$$T_1 = T_2 = \frac{160}{186}T_3 = T_4$$

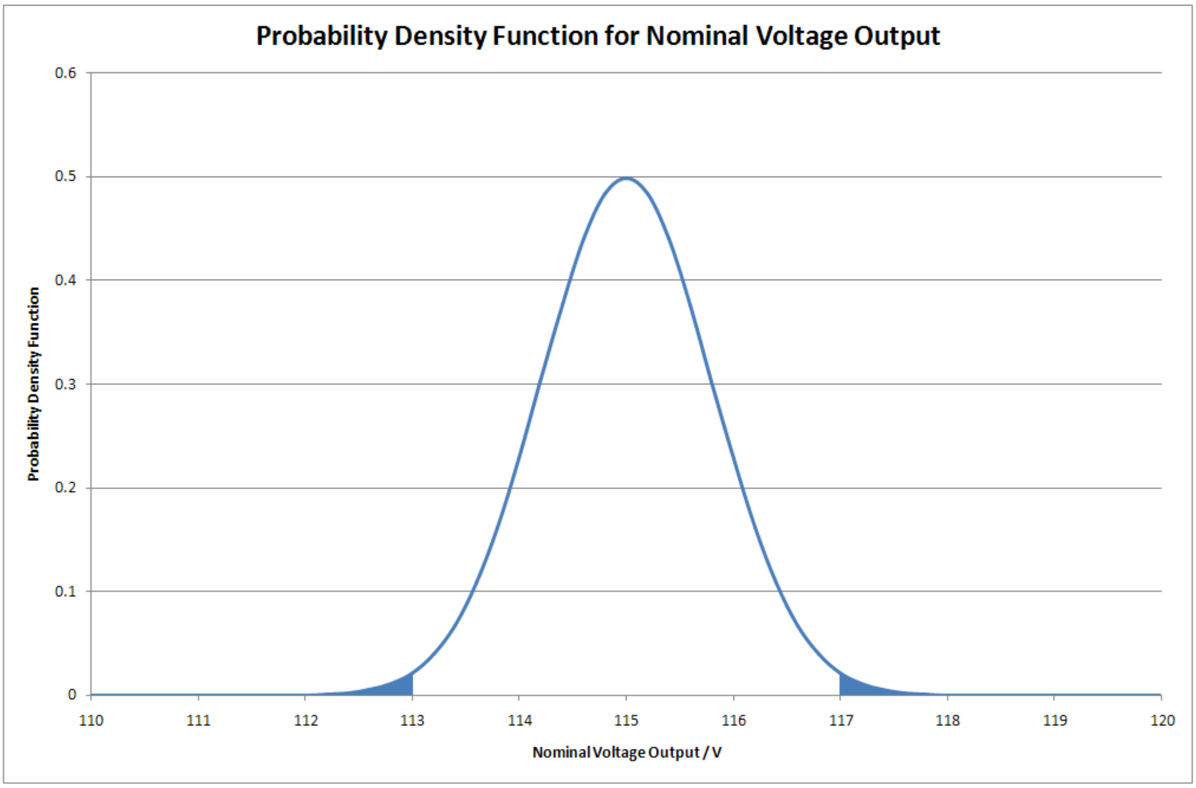
For this equation to hold true, the third press must spend 16.25% more time operational than the other machines. This means that if press 3 was operational for the entire shift, the operational time of the other presses must be capped at 83.75%, in order for press 3 to produce the same number of tennis balls as the other presses. Capping the operational time of presses 1, 2, and 4 and 83.75% is not a good idea, since it only serves to reduce throughput. Also, such a policy would not be implemented in practice since it would require extra monitoring of the allotted processing times for the presses. A more workable solution is to simple accept the fact that press 3 will contribute less to the total production than the other machines, and to allocate the same operational time to each press.

Another minor issue is that press 3 will now have to be treated differently to its siblings. All operators must be instructed to remember to not fill the outer perimeter of hollows for press 3 during loading.


Probability Density Functions for Tennis Ball Core Diameters (Corrected)
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Proability Density Function
40
50
60
70
80
90
100
110
120
130
Tennis Ball Core Diameter / mm
Press 1
Press 2
Press 3
Press 4
Customer Limits

Answer to Question 3

We are told that the quality level for the customer is required to be 99%. Hence the non-conformance rate for the customer is capped at 1%, and the customer limits are between 113 and 117 volts. Therefore looking at the below graph, we want the shaded area in blue to be less than 1%.



The standardised lower and upper limits for the voltage will therefore be given by:

$$SLL = \frac{113 - 115}{0.8} = -2.5 \qquad SUL = \frac{117 - 115}{0.8} = 2.5$$

Since the area of the "tails" in the above graph page are symmetric, the non-conformance rate will be given by the below equation where Z(2.5) is the value obtained from the Z-Table(given on the last page) for 2.5.

$$Area = 2[1 - Z(2.5)]$$

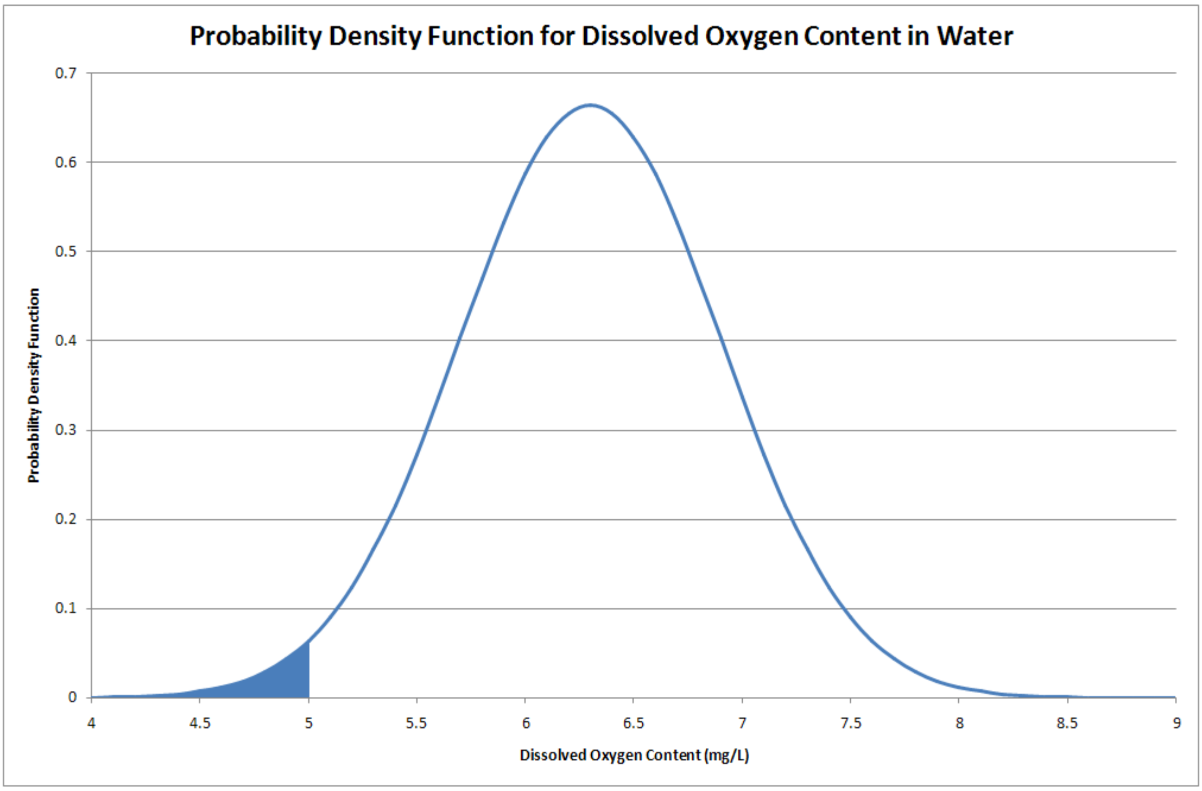
$$Area = 2[1 - 0.9938]$$

$$Area = 0.0124$$

Therefore the non-conformance rate 1.24%, higher than what the customer will allow.

Answer to Question 4

We are told that if the dissolved oxygen content is less than 5 mg/L is undesirable. This corresponds to the shaded area in the graph below.



The standardised lower limit for the dissolved oxygen content is given by:

$$SLL = \frac{5 - 6.3}{0.6} = -2.17$$

Due to symmetry, the area of the blue tail in the above graph is given by the following equation:

$$Area = 1 - Z(2.17)$$

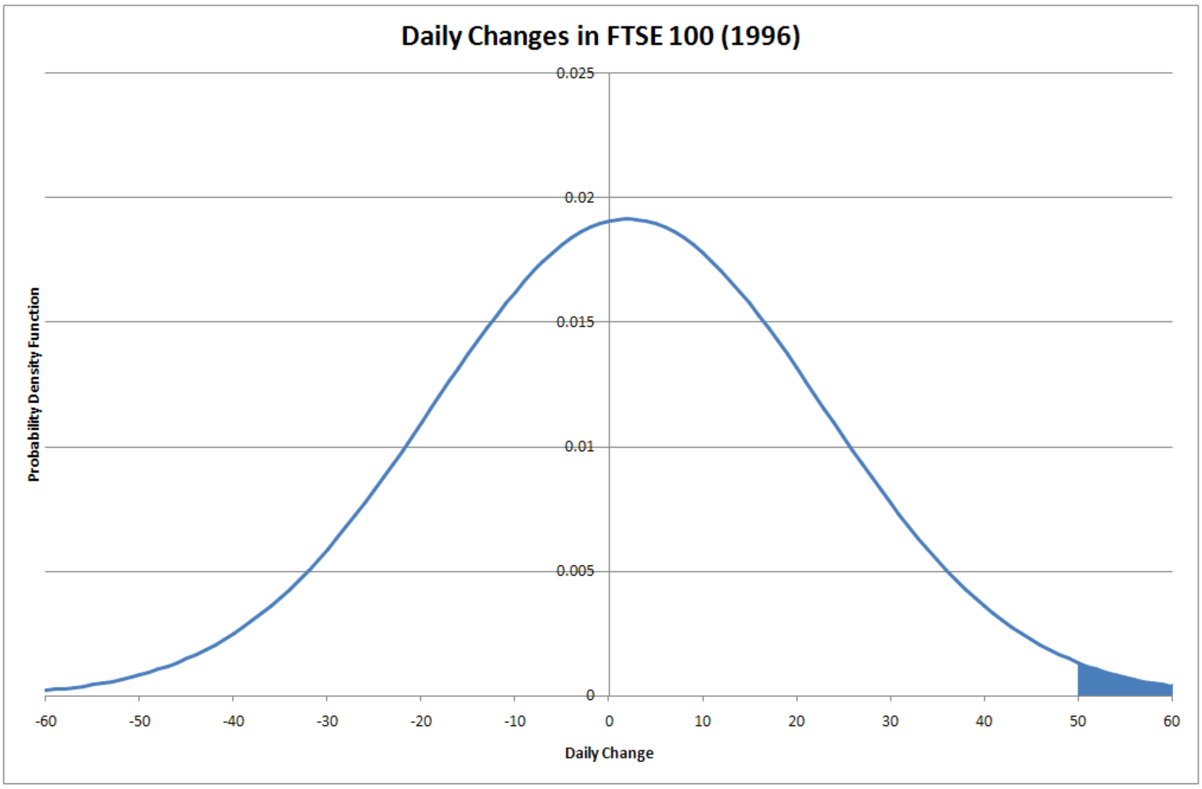
$$Area = 1 - 0.9850$$

$$Area = 0.015$$

Therefore 1.5% of the measurements will be undesirable.

Answer to Question 5

For the first part of the question, we want to know how many trading days we would expect with a daily change of more than 50. So firstly we want to find out the percentage of area under a curve with mean 2.00 and standard deviation 20.82 that is greater than 50.



The standardised upper limit for 50 days is given by:

$$SUL_{50} = \frac{50 - 2}{20.82} = 2.31$$

The area of the blue tail in the above graph is given by the following equation:

$$Area = 1 - Z(2.31)$$

$$Area = 1 - 0.9896$$

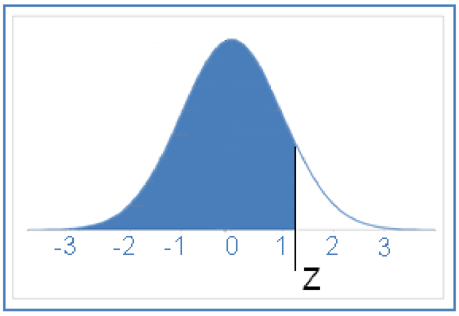
$$Area = 0.0104$$

This means the percentage of days where daily changes of 50 or more points in the FTSE is expected is 1.04%. Since there were 265 trading days in 1996, this translates to 2.756 days or rounding up, 3 days. We repeat this process for 1996 setting the upper limit to 75 and then 100.

In the second part of the question we look at the normal model for the second half of 1997 where the mean is 3 and the standard deviation 56. Also, in 1997 there were 130 trading days on the London Stock Exchange in the second half of 1997. Again we want to determine the number of days where we expect daily changes of greater than 50, 75, and 100 days. The results are shown in table format on the following page.

Year	No. Trading Days	Upper Limit	% Days Above Upper Limit	No. Days Above Upper Limit
1996	265	50	1.04%	3
1996	265	75	0.02%	0
1996	265	100	0%	0
1997	130	50	20%	26
1997	130	75	10%	13
1997	130	100	4%	5

We observe a dramatic change in the number of days with a high daily change that occurred in the second half of 1997. This can be attributed almost solely to the increase in standard deviation (volatility).



Z-Table (Area Under the Curve[2])

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997

[2] Values used in the questions are highlighted in yellow.