Question 1

A fabric manufacturer claims that its fabric's durability has a standard deviation σ of 3500 units. A quality control department evaluated the fabric by testing the durability of a random sample of size 25, and obtained a sample standard deviation of 4270 units. Assume that the durability measurements are normally distributed.

Set up the hypotheses to check if the true standard deviation is greater than the claimed value, and conduct a hypothesis test at the 10 % significance level.

(Hints: this can be done using a CI (think about which one), or using a p-value (think about which one). You can check your answer by doing the test using both methods. You may of course use Excel.)

. . . .

Answer: We are interested in testing the fabric manufacturer's claim on the standard deviation of the fabric's durability. As such, we set up the following null and alternate hypothesis:

$$H_0 : \sigma = 3500$$
$$H_1 : \sigma > 3500$$

To test this hypothesis, we compute the $100(1-\alpha)\%$ one-sided confidence interval for the variance:

$$\text{CI} = \left[\frac{(n-1)s_x^2}{\chi^2_{n-1,\alpha}}, \infty\right)$$

Given 25 samples, a sample standard deviation of 4270, and a significance level of 0.1:

$$\text{CI} = \left[\frac{(24)4270^2}{\chi^2_{24,0.1}}, \infty\right)$$

We can compute $\chi^2_{24,0.1}$ by using the Excel command `chisq.inv(0.9, 24)`, which gives:

$$\begin{aligned}\text{CI} &= \left[\frac{(24)4270^2}{33.19624429}, \infty\right) \\ &= [13181900.83, \infty)\end{aligned}$$

To get the confidence interval for the standard deviation, we can take the square root of the CI above. Hence the confidence interval for the standard deviation.

$$\begin{aligned}\text{CI} &= \left[\frac{(24)4270^2}{33.19624429}, \infty\right) \\ &= [3630.688754, \infty)\end{aligned}$$

Since $\sigma = 3500$ is outside this confidence interval, we can reject H_0 in favour of H_1. Thus the claim of the fabric manufacturer cannot be accepted.

Alternatively, we can also compute the p-value to perform the hypothesis testing. In this context, the p-value can be computed as follows:

$$\begin{aligned}
p &= \mathbb{P}(S^2 \geq 4270^2) \\
&= \mathbb{P}\left(\frac{(n-1)S^2}{\sigma_0^2} \geq \frac{(n-1)4270^2}{\sigma_0^2}\right) \\
&= \mathbb{P}\left(\chi^2_{n-1} \geq \frac{(n-1)4270^2}{\sigma_0^2}\right) \\
&= \mathbb{P}\left(\chi^2_{24} \geq \frac{(24)4270^2}{3500^2}\right) \\
&= \mathbb{P}\left(\chi^2_{24} \geq \frac{22326}{625}\right) = 1 - \mathbb{P}\left(\chi^2_{24} < \frac{22326}{625}\right)
\end{aligned}$$

By using the Excel command `1 - chisq.dist(22326/625, 24, 1)`, we can compute the p-value and obtain:

$$\boxed{p = 0.0583944577}$$

Since the p value is significantly lower than $\alpha = 0.1$, we can reject H_0 in favour of H_1. Our findings through the p-value confirm our findings using the confidence interval.

Question 2

In a series of cloud seeding experiments conducted from 1968–1972, 52 isolated clouds were observed, of which 26 were selected at random and injected with a chemical (silver-iodide), in the hope that these 'seeded' clouds would produce more rainfall than the unseeded clouds. The resulting rainfall data can be found in *Excel*. Let $\alpha = 0.05$.

(a) Write down the relevant hypotheses, and determine whether this is independent samples design or matched pairs design. It can be shown that the data itself is not normal, however, after taking the ln of each data value, the transformed data for each group is approximately normal, and the two transformed groups have comparable variances. (You do not need to verify these claims.)
(b) Carry out an appropriate hypothesis test for the transformed data, report the p-value, and state your conclusion.

(Hints: (a) define everything; (b) please show all relevant steps – attach an Excel screenshot if it helps, but no need to submit Excel.)

. . . .

(a) **Answer:** To begin, let U and S each denote the rainfall produced by unseeded and seeded clouds respectively. As we are interested to test if the rainfall from the seeded clouds significantly exceed that of the unseeded clouds, we utilize the Independent Samples Design. This is because the clouds were selected randomly and not paired one-to-one as one would expect in Matched Pairs Design. Hence our null and alternate hypotheses:

$$H_0 : \mu_S - \mu_U = 0$$
$$H_1 : \mu_S - \mu_U > 0$$

. . . .

(b) **Answer:** To perform the hypothesis testing, we need to transform the data. Let X and Y denote the natural logarithm of S and U respectively (so $X = \ln(S)$ and $Y = \ln(U)$). By using *Excel*, we can compute the value of X and Y. The results are shown below:

U	**S**	**Y**	**X**
1202.6	2745.6	7.09224115860547	7.91775490929364
830.1	1697.8	6.72154617546099	7.43708857430387
372.4	1656	5.91996854540291	7.4121603349452
345.5	978	5.84499264320772	6.88550967003482
321.2	703.4	5.77206398207261	6.55592572003988
244.3	489.1	5.49839697826369	6.19256696754531
163	430	5.09375020080676	6.06378520868761
147.8	334.1	4.9958600085141	5.81144034936271
95	302.8	4.55387689160054	5.71307252156329
87	274.7	4.46590811865458	5.61567959310123
81.2	274.7	4.39691524716763	5.61567959310123
68.5	255	4.22683374526818	5.54126354515843
47.3	242.5	3.85651029549789	5.49100171037754
41.1	200.7	3.71600812150219	5.30181125580229
36.6	198.6	3.60004824040732	5.29129275161107
29	129.6	3.36729582998647	4.86445278391817
28.6	119	3.35340671782581	4.77912349311153
26.3	118.3	3.26956893918372	4.77322377098434
26.1	115.3	3.26193531432865	4.74753742727501
24.4	92.4	3.19458313229916	4.52612697864764
21.7	40.6	3.07731226054641	3.70376806660769
17.3	32.7	2.85070650150373	3.48737507790321
11.5	31.4	2.4423470353692	3.44680789291421
4.9	17.5	1.58923520511658	2.86220088092947
4.9	7.7	1.58923520511658	2.04122032885964
1	4.1	0	1.41098697371026

We also update the null and alternate hypotheses into the following:

$$H_0 : \mu_X - \mu_Y = \delta_0$$
$$H_1 : \mu_X - \mu_Y > \delta_0$$
$$(\text{with } \delta_0 = 0)$$

These hypotheses are equivalent to the ones we set up in (a), as any significant difference between U and S would also reflect a significant difference between X and Y. Moving forward, we compute the sample mean and sample variance of X and Y by using the *Excel* commands `=AVERAGE()` and `=VAR.S()`.

$\bar{x}$	5.134186784
$\bar{y}$	3.990405634
s_x^2	2.55844379
s_y^2	2.69566314

Using these values, we compute the p-value:

$$\begin{aligned} p &= \mathbb{P}(T_{2n-2} \geq t), \text{with} \\ t &= \frac{\bar{x} - \bar{y} - \delta_0}{\sqrt{\frac{s_x^2 + s_y^2}{n}}} \\ &= \frac{5.134186784 - 3.990405634 - 0}{\sqrt{\frac{2.55844379 + 2.69566314}{26}}} \\ &= 2.544369349 \end{aligned}$$

Hence the p-value:

$$\begin{aligned} p &= \mathbb{P}(T_{2n-2} \geq 2.544369349) \\ &= 1 - \mathbb{P}(T_{2n-2} \leq 2.544369349) \end{aligned}$$

Using `t.dist()`, we compute $\mathbb{P}(T_{2n-2} \leq 2.544369349) = 0.992958671$. Thus the p-value becomes:

$$p = 1 - 0.992958671$$
$$\boxed{p = 0.007041329}$$

Since $p < \alpha$, we can reject H_0 in favour of H_1. As such, we conclude that the seeded clouds do yield a significantly higher amount of rainfall than that of the unseeded clouds.

Question 3

The Excel spreadsheet contains some data on the weight gains of 30 anorexic patients after being given a treatment (here, the aim of the treatment is to help the patients gain weight).

Let $\alpha = 0.01$; use the signed rank test to check whether the true mean weight gain is more than 0.

(Hint: some of the d_i's are equal in value, however Excel's `rank.avg` function will automatically assign the correct ranks to them. Please show all relevant steps – attach an Excel screenshot if it helps, but no need to submit Excel.)

....

Answer: The treatment is designed to help patients gain weight, and we are interested to test if it is really effective. To do so, let μ be the true mean weight gain. We can formulate the following null and alternative hypotheses to be tested:

$$H_0 : \mu = 0$$
$$H_1 : \mu > 0$$

This is a signed rank test. To proceed, we compute $\mathbb{E}(W^+)$ and $\text{Var}(W^+)$.

$$\begin{aligned}\mathbb{E}(W^+) &= \frac{n(n+1)}{4} \\ &= \frac{30(30+1)}{4} = 232.5 \quad (\text{since } n = 30) \\ \text{Var}(W^+) &= \frac{n(n+1)(2n+1)}{24} \\ &= \frac{30(30+1)(2 \cdot 30 + 1)}{24} = 2363.75\end{aligned}$$

Since n is large, W^+ is approximately normal with $\mathbb{E}(W^+) = 232.5$ and $\text{Var}(W^+) = 2363.75$ (i.e. $W^+ \sim \mathcal{N}(232.5, 2363.75)$). To complete the test, we calculate the p-value:

$$p = \mathbb{P}(W^+ \geq w_+)$$

Using *Excel*, we compute the absolute value of d_i (d_i is the signed weight difference in kilograms) and the sum of the positive ranks, w_+:

$\mathbf{d_i}$	$\mathbf{\lvert d_i \rvert}$	**rank**	**calculation of $\mathbf{w_+}$**
-1.9	1.9	19.5	0
-1.7	1.7	18	0
-1.3	1.3	15.5	0
-1.2	1.2	13	0
-1.2	1.2	13	0
-0.9	0.9	9	0
-0.9	0.9	9	0
-0.9	0.9	9	0
-0.5	0.5	5.5	0
-0.4	0.4	2.5	0
-0.4	0.4	2.5	0
-0.3	0.3	1	0
0.5	0.5	5.5	5.5
0.5	0.5	5.5	5.5
0.5	0.5	5.5	5.5
1	1	11	11
1.2	1.2	13	13
1.3	1.3	15.5	15.5
1.4	1.4	17	17
1.9	1.9	19.5	19.5
2.5	2.5	21	21
2.7	2.7	22	22
3.3	3.3	23	23
3.6	3.6	24	24
3.7	3.7	25	25
4.1	4.1	26	26
4.2	4.2	27	27
4.3	4.3	28.5	28.5
4.3	4.3	28.5	28.5
4.9	4.9	30	30

By summing all the calculations for w_+, we get $w_+ = 347.5$. Hence the p-value:

$$\begin{aligned} p &= \mathbb{P}(W^+ \geq 347.5) \\ &= \mathbb{P}\left(Z \geq \frac{347.5 - 232.5 - 0.5}{\sqrt{2363.75}}\right) \\ &= \mathbb{P}(Z \geq 2.35507) = \mathbb{P}(Z \leq -2.35507) \\ &\boxed{p = 0.00925} \end{aligned}$$

Since p is less than $\alpha = 0.01$, we can reject H_0 in favour of H_1.

Question 4

A credit card company is about to send out mail offers to a large number of people to test the market for a new credit card. The goal is to estimate the true proportion of people who will sign up for the card. A preliminary study suggests that at most 1% of the people receiving the offer will sign up.

To be within 0.1 percentage point of the true proportion with 95% confidence, how many mail offers should the company send out?

(Hint: there are 2 formulas for sample size in the slides... which is the most appropriate one to use here?)

. . . .

Answer: The goal here is to estimate the number of mail offers to be sent such that we can be within $E = 0.001$ of the true proportion of people who will sign up with 95% confidence. Let $\hat{p}$ be the estimated sample proportion of people who will sign up for the credit card. Thanks to the preliminary study, we know that $\hat{p} = 0.01$. We can now use this information with the equation for sample size given $\hat{p}$ provided in Week 10 Class 1:

$$\begin{aligned} n &= \left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{z_{0.025}}{E}\right)^2 \hat{p}(1-\hat{p}) \\ &= \left(\frac{1.96}{0.001}\right)^2 (0.01)(1-0.01) \\ &= 38031.84 \\ &\boxed{n = 38032} \end{aligned}$$

Thus, we now know the number of mail offers to send. Note that the number is rounded to the closest integer as one cannot send a fractional number of mail offers.

Question 5

In a computer game (Dota 2), each player controls a 'hero' that can be used to deal damage against other heroes. One hero possesses an attribute known as 'bash', namely, 17% of the time he can disable the target and deal extra damage.

During the course of one game, normal (non-bash) attacks and bashes are recorded and presented in Excel. Runs of normal attacks are indicated by integers, while each bash is indicated by a 'B'.

(a) Let p be the true proportion of bashes; we wish know whether the value claimed in the game (17%) is accurate. Test $H_0 : p = 0.17$ vs $H_1 : p \neq 0.17$ at the 5% significance level.

(b) Due to our lack of intuition for what is random, some of the game mechanics are not determined using truly random processes. Perform a runs test with $\alpha = 0.05$ to check whether the bashes occur randomly.

(Hints: you do not need to count the number of bashes or the number of runs manually, but can use Excel's count function cleverly; for both parts, compute a p-value.)

....

(a) **Answer:** We can perform this hypothesis testing by computing the p-value. Since n is large, we can apply the CLT on $\hat{p}$ to determine:

$$\begin{aligned}\hat{p} &\sim \mathcal{N}(p, p(1-p)/n)\\ \hat{p} &\sim \mathcal{N}(0.17, 0.17(1-0.17)/820)\\ \hat{p} &\sim \mathcal{N}(0.17, 0.000172)\end{aligned}$$

Using *Excel* functions, we find that $n(\text{bash}) = 130$, $n(\text{non-bash}) = 690$, and $n(\text{total}) = 820$. As such, we compute an estimate for p, $\hat{p} = \frac{130}{820} = 0.158536585$. Using this information, we compute the following test-statistic:

$$\begin{aligned}z &= \frac{\hat{p} - p'}{\sqrt{p'(1-p')/n}}\\ &= \frac{0.158536585 - 0.17}{\sqrt{(0.158536585)(1-0.158536585)/820}}\\ &= -0.8987492858\end{aligned}$$

Using the test-statistic, we compute the 2-sided p-value to complete the hypothesis testing:

$$\text{p-value} = \mathbb{P}(Z < -0.8987492858)$$

$$\boxed{\text{p-value} = 0.382}$$

As the computed p-value is not less than the significance level, $\alpha = 0.05$, we cannot reject the null hypothesis in favour of the alternate hypothesis. In other words, the given proportion of bashes, $p = 17\%$ is accurate.

....

(b) **Answer:** To perform the runs test, let n be the number of bash attacks, m be the number of non-bash attacks, and X be the total number of runs. Using *Excel*, we calculate the number of runs by comparing each entry with the previous entry. If an entry is different from the previous entry (e.g. an entry is 5 and the next entry is B), we consider it to be a run. However, if an entry is the same as its previous entry (e.g. B and B), then this is not considered a new run. We can perform this calculation by the formula `=IF(previous_cell_index=current_index,0,1)`. The observed total number of runs, x, is simply the sum of all these values. Thus, we find that $n = 130$, $m = 690$, and $x = 232$. See the screenshot below for more details. In a runs test, we know that X is approximately normal with mean μ and variance σ^2, which we can compute using the formula given in class:

$$\begin{aligned}\mu &= \frac{2mn}{m+n} + 1\\ &= \frac{2 \cdot 130 \cdot 690}{820} + 1\\ &= \frac{9011}{41} = 219.7804878\end{aligned}$$

$$\begin{aligned}\sigma^2 &= \frac{(\mu-1)(\mu-2)}{m+n-1}\\ &= \frac{(219.7804878-1)(219.7804878-2)}{130+690-1}\\ &= 58.17597235\end{aligned}$$

B3 =IF(A3=A2,0,1)

	A	B	C	D	E
1	dat	run?			
2	5	1			
3	B	1		For instance, '5 B 2 B 3' means:	
4	2	1		5 consecutive normal (non-bash) attacks, followed by a bash,	
5	B	1		followed by 2 consecutive normal attacks, followed by a bash,	
6	3	1		followed by 3 consecutive normal attacks.	
7	B	1			
8	9	1		**n(bash)**	130
9	B	1		**n(non-bash)**	690
10	4	1		**n(total)**	820
11	B	1			
12	3	1		**p hat**	0.158536585
13	B	1			
14	B	0		**number of runs**	232
15	6	1			
16	B	1			
17	3	1			
18	B	1			
19	4	1			
20	B	1			
21	4	1			
22	B	1			
23	1	1			

Figure 1: *Excel* results

B4 =IF(A4=A3,0,1)

	A	B	C	D	E
1	**data**	**run?**			
2	5	1			
3	B	=IF(A3=A2,0,1)		For instance, '5 B 2 B 3' means:	
4	2	=IF(A4=A3,0,1)		5 consecutive normal (non-bash) attacks, followed by a bash,	
5	B	=IF(A5=A4,0,1)		followed by 2 consecutive normal attacks, followed by a bash,	
6	3	=IF(A6=A5,0,1)		followed by 3 consecutive normal attacks.	
7	B	=IF(A7=A6,0,1)			
8	9	=IF(A8=A7,0,1)		**n(bash)**	=COUNTA(FILTER(A:A,A:A="B"))
9	B	=IF(A9=A8,0,1)		**n(non-bash)**	=SUM(A:A)
10	4	=IF(A10=A9,0,1)		**n(total)**	=SUM(E8:E9)
11	B	=IF(A11=A10,0,1)			
12	3	=IF(A12=A11,0,1)		**p hat**	=E8/E10
13	B	=IF(A13=A12,0,1)			
14	B	=IF(A14=A13,0,1)		**number of runs**	=SUM(B2:B247)
15	6	=IF(A15=A14,0,1)			
16	B	=IF(A16=A15,0,1)			
17	3	=IF(A17=A16,0,1)			
18	B	=IF(A18=A17,0,1)			
19	4	=IF(A19=A18,0,1)			
20	B	=IF(A20=A19,0,1)			
21	4	=IF(A21=A20,0,1)			
22	B	=IF(A22=A21,0,1)			
23	1	=IF(A23=A22,0,1)			

Figure 2: *Excel* formulas

Since $x > \mu$, we can compute the 2-sided p-value as follows (while remembering to perform continuity correction, as X is originally non-continuous):

$$\begin{aligned} p &= 2\mathbb{P}(Z > (x - \mu - 0.5)/\sigma) \\ &= 2\mathbb{P}(Z > (232 - 219.7804878 - 0.5)/58.17597235) \\ &= 2\mathbb{P}(Z > 1.536518204) \\ &= 2\mathbb{P}(Z < -1.536518204) \\ &\boxed{p = 0.124} \end{aligned}$$

Since the p-value is not less than the significance level $\alpha = 0.05$, we cannot reject the null hypothesis that the bashes occurred randomly. In other words, we conclude that the bashes did occur randomly.

Question 6

In Japan, there is a belief that blood type is predictive of personality. For instance, people with blood type O are seen as confident, strong-willed, and hence good leaders. To support this claim, it is known that a large proportion of Japanese prime ministers have blood type O. Indeed, among 35 recent prime ministers, 18 have blood type O, 10 have type A, 5 have type B, and 2 have type AB.

It is also known that in Japan, the population proportions of the above four blood types are, respectively, 0.3, 0.4, 0.2 and 0.1.

Conduct a chi-squared test, with $\alpha = 0.05$, to determine whether the prime ministers' blood types are significantly different from what one would expect from the population proportions.

(Hint: for simplicity, you do not need to treat any $e_i < 5$ differently for this question.)

Answer: To begin, we set up the following hypotheses:

$$\begin{aligned} &H_0 : (p_1 = 0.3) \cap (p_2 = 0.4) \cap (p_3 = 0.2) \cap (p_4 = 0.1) \\ &H_1 : \neg H_0 \end{aligned}$$

with p_1, p_2, p_3, p_4 each denoting the proportion of O, A, B, and AB blood types among the prime ministers of Japan. Under the null hypothesis, the distribution of blood types among the prime ministers of Japan is the same as that of regular people in Japan. To test if this hypothesis is correct, we need to compute the chi-squared statistic, $\mathcal{X}^2$, which is defined as:

$$\mathcal{X}^2 = \sum_{i=1}^{4} \frac{(n_i - e_i)^2}{e_i}$$

where $e_i = np_i'$ denote the expected counts of the i-th blood type. Given that $n = 35$, and that:

$$\begin{aligned} p_1' &= 0.3, & p_2' &= 0.4, & p_3' &= 0.2, & p_4' &= 0.1 \\ n_1 &= 18, & n_2 &= 10, & n_3 &= 5, & n_4 &= 2 \end{aligned}$$

we have:

$$e_1 = 10.5, \qquad e_2 = 14, \qquad e_3 = 7, \qquad e_4 = 3.5$$

and hence the chi-squared statistic:

$$\begin{aligned} \mathcal{X}^2 &= \frac{(18-10.5)^2}{10.5} + \frac{(10-14)^2}{14} + \frac{(5-7)^2}{7} + \frac{(2-3.5)^2}{3.5} \\ &= 7.71428571429 \end{aligned}$$

We compare this chi-squared statistic with the critical value $\mathcal{X}^2_{m-1,\alpha} = \mathcal{X}^2_{3,0.05}$. This critical value can be computed using the *Excel* function `=CHISQ.INV(1-0.05,3)`, which gives us:

$$\mathcal{X}^2_{3,0.05} = 7.814727903$$

Since $\mathcal{X}^2 < \mathcal{X}^2_{3,0.05}$, we cannot reject the null hypothesis. In other words, the distribution of blood types among the prime ministers of Japan do not differ significantly from that of regular people.

Question 7

In *Excel* you can find 800 data points, and the goal here is to use a chi-squared test to check whether a normal distribution fits the data well. To do so, we will first organize the raw data into 8 bins, where the bin ranges are chosen such that the expected count for each bin is $800/8 = 100$.

(a) Several of the bin ranges have been given (columns D and E in *Excel*); find the missing entries.

(b) Conduct the chi-squared test using $\alpha = 0.05$.

(Hints: (a) $\bar{x}$ and s_x have been computed from the data; H_0 says that the data is drawn from a $\mathcal{N}(\bar{x}, s_x^2)$ distribution; we are looking for critical values of that distribution, such that the probability between adjacent critical values is 1/8. (b) Think about the degree of freedom. Please show all relevant steps - attach an Excel screenshot if it helps, but no need to submit Excel.)

. . . .

(a) **Answer:** We are splitting the raw data into 8 bins, each with equal expected count. As such, there is an equal probability of 1/8 for a data point to fall into any of the bins.

Utilising this fact, and the fact that we are attempting to test whether a normal distribution fits the data well, we find the bin ranges by finding the z-values which divide the normal distribution into 8 parts of equal area / probability. With the help of *Excel* and the function `=NORM.INV()`, we identify these ranges as follows:

E8 =NORM.INV(C8/8,D$2,D$3)

	C	D	E	F	G
2	sample mean:	47.2452			
3	sample SD:	3.979			
4					
5	bin	*from:*	*to:*	observed	expected
6	*1*	-1000	42.668	114	100
7	*2*	42.668	44.561	99	100
8	*3*	44.561	45.9773	89	100
9	*4*	45.9773	47.2452	95	100
10	*5*	47.2452	48.5131	94	100
11	*6*	48.5131	49.929	84	100
12	*7*	49.929	51.8224	129	100
13	*8*	51.8224	1000	96	100
14					
15			total:	800	800

Figure 3: *Excel* working for the missing entries

. . . .

(b) **Answer:** To perform the chi-squared statistic, first we set up the null and alternate hypotheses:

$$H_0 : (p_1 = 1/8) \cap (p_2 = 1/8) \cap \cdots \cap (p_8 = 1/8)$$
$$H_1 : \neg H_0$$

To proceed, we compute the chi-squared statistic using *Excel*.

observed	**expected**	$\mathbf{\frac{(n_i - e_i)^2}{e_i}}$
114	100	1.96
99	100	0.01
89	100	1.21
95	100	0.25
94	100	0.36
84	100	2.56
129	100	8.41
96	100	0.16

This gives us the chi-squared statistic $\mathcal{X}^2 = 14.92$. Now, we compare this value with the critical value, which we can find by simply using the *Excel* function `CHISQ.INV(1-0.05,5)`. $1 - 0.05$ represent our confidence level, $100(1-\alpha)\%$, whereas 5 represent our degree of freedom. Our degree of freedom was initially $m - 1 = 7$ (with m being the number of bins, 8). However, since we estimated the true mean and variance of the population, we lose 2 degrees of freedom and hence 5 is our final degree of freedom. Thus we find:

$$\mathcal{X}_{5,0.05} = 11.07049769$$

Since $\mathcal{X}^2 > \mathcal{X}_{5,0.05}$, we have significant evidence to reject the null hypothesis in favour of the alternate hypothesis. In other words, at least one of the bins do not match the probability distribution of a normal random variable. We can therefore conclude that the normal distribution does not fit well with the datapoints.