Question 1

Due to federal cuts to the Government Health System, the researchers predicted that the average length of stay in hospital for Australians admitted for suspected myocardial infarction is now lower than the average of seven (7) days recorded in 2018. Conduct a one-sample *t*-test using the **LOS** variable to test this prediction.

Produce the relevant R output and write a one-sample *t*-test report.

A study was conducted to investigate if the length of stay in hospital for Australians admitted for suspected myocardial infraction is lower than the average of seven (7) days recorded in 2018. In a survey of 392 Australians, the average length of stay in the hospital was 7.22 days (s = 2.65 days). This is higher than the average [mean] length of stay in hospital of seven (7) days recorded in 2018, and a one-sample t-test shows that the difference in average length of stay is insignificant, t(391) = 1.638, $p < 0.001$. The 95% confidence interval indicates that since 2018, the average length of stay in hospital is between 6.96 days and 7.48 days.
Contrary to the expectation, the average length of stay in hospital for Australians suspected to have myocardial infractions has increased since 2018.

Include your R input and output for this question here.

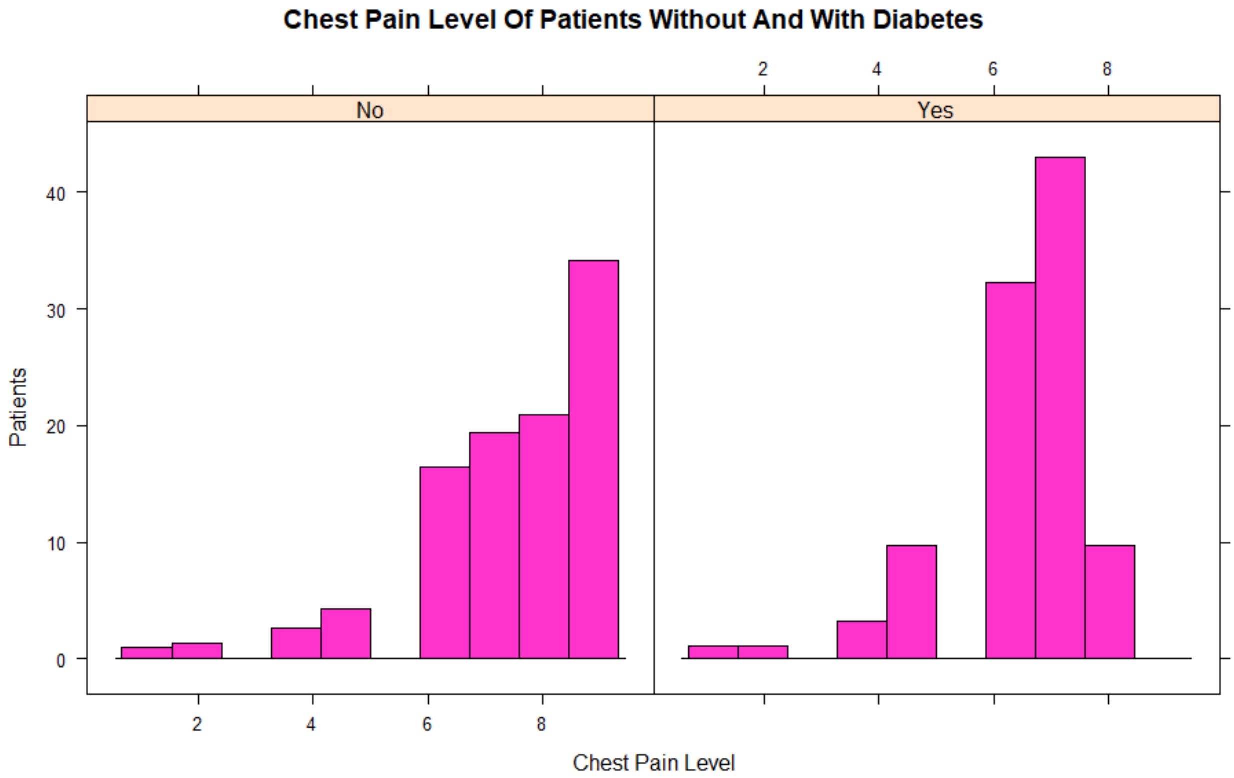
```
> rdm_smpl<-readRDS("sample.rds")
> stat.desc(rdm_smpl$los)
     nbr.val     nbr.null       nbr.na          min          max        range          sum
 392.0000000    0.0000000    8.0000000    4.0000000   18.0000000   14.0000000 2830.0000000
      median         mean      SE.mean CI.mean.0.95          var      std.dev     coef.var
   7.0000000    7.2193878    0.1339265    0.2633061    7.0310298    2.6516089    0.3672900
> t.test(rdm_smpl$los, alternative=c("less"), mu=7)

	One Sample t-test

data:  rdm_smpl$los
t = 1.6381, df = 391, p-value = 0.9489
alternative hypothesis: true mean is less than 7
95 percent confidence interval:
   -Inf 7.4402
sample estimates:
mean of x 
 7.219388 

> t.test(rdm_smpl$los, alternative=c("two.sided"), mu=7)

	One Sample t-test

data:  rdm_smpl$los
t = 1.6381, df = 391, p-value = 0.1022
alternative hypothesis: true mean is not equal to 7
95 percent confidence interval:
 6.956082 7.482694
sample estimates:
mean of x 
 7.219388 
```

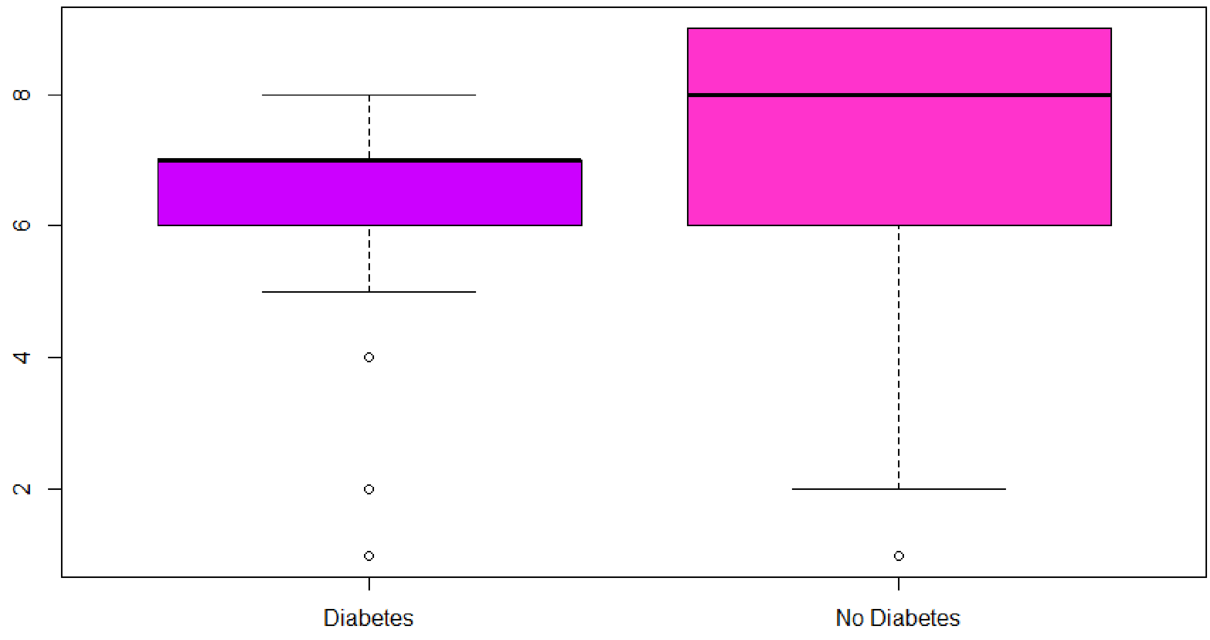
Question 2a

Based on previous research, the investigators have also predicted that the average chest pain level reported by patients admitted to hospital for suspected myocardial infarction is lower for those who have a history of diabetes than the average pain level for those who do not have a history of diabetes. The pain level was measured on a 0 to 10 points scale. Conduct an independent samples *t*-test using the **Pain1** and **Diabetes** variables to test this claim.

Produce the relevant output and write an independent samples *t*-test report.

It was hypothesised that patients who have a history of diabetes admitted to hospital for suspected myocardial infraction have lower average chest pain level than patients who do not have a history of diabetes. In the sample of 398 patients, the average chest pain level for patients with a history of diabetes ($\bar{x}$ = 6.37 points, s = 1.17 points, n = 93) was lower than patients without a history of diabetes ($\bar{x}$ = 7.44 points, s = 1.63 points, n = 305) and an independent sample t-test shows that this difference in mean chest pain level is significant, $t(396) = -5.92$, $p < 0.001$. The 95% confidence interval indicates that the mean chest pain level for patients with a history of diabetes is between 0.72 and 1.43 points lower than patients without a history of diabetes. As expected, patients with a history of diabetes have lower average chest pain level than patients without a history of diabetes.





Include your R input and output for this question here.

```
> Diabetes<-subset(rdm_smpl, diabetes=="Yes")
> NoDiabetes<-subset(rdm_smpl, diabetes=="No")
>
> stat.desc(Diabetes$pain1)
     nbr.val     nbr.null       nbr.na          min          max        range          sum
  93.0000000    0.0000000    0.0000000    1.0000000    8.0000000    7.0000000  592.0000000
      median         mean      SE.mean CI.mean.0.95          var      std.dev     coef.var
   7.0000000    6.3655914    0.1211455    0.2406055    1.3648901    1.1682851    0.1835313
> t.test(Diabetes$pain1, NoDiabetes$pain1, alternative=c("less"), var.equal=TRUE)

	Two Sample t-test

data:  Diabetes$pain1 and NoDiabetes$pain1
t = -5.9192, df = 396, p-value = 3.507e-09
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
       -Inf -0.7746751
sample estimates:
mean of x mean of y 
 6.365591  7.439344 

>
> t.test(NoDiabetes$pain1, Diabetes$pain1, var.equal=TRUE)

	Two Sample t-test

data:  NoDiabetes$pain1 and Diabetes$pain1
t = 5.9192, df = 396, p-value = 7.014e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.7171234 1.4303823
sample estimates:
mean of x mean of y 
 7.439344  6.365591 
```

```
> histogram(~pain1|diabetes, rdm_smpl, xlab="Chest Pain Level",
+           ylab="Patients",
+           main="Chest Pain Level Of Patients Without And With Diabetes",
+           col=c("#ff33cc"))
>
>
> boxplot(Diabetes$pain1, NoDiabetes$pain1, names=c("Diabetes", "No Diabetes"),
+         col=c("#cc00ff", "#ff33cc"))
>
```

```
> var.test(Diabetes$pain1, NoDiabetes$pain1)

        F test to compare two variances

data:  Diabetes$pain1 and NoDiabetes$pain1
F = 0.51664, num df = 92, denom df = 304, p-value = 0.0002615
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3764438 0.7304345
sample estimates:
ratio of variances
         0.5166383
```

Question 2b
Check and comment on the assumptions of the independent samples *t*-test.
The assumptions of the independent samples t-test are as follows. The dependent variable in the independent samples t-test which is the average chest pain level (points) are measured on a metric scale. The observations are independent whereby the sample selection must be unbiased. For the normality of samples, both samples are not from normally distributed populations therefore, the normality of samples assumption is violated. Finally, for the equal variances, the F test shows that the variance of the two samples is significant to be unequal, $p = 0.0002$. Thus, the samples come from populations with not equal variances and both groups does not have a similar spread. Therefore, the equal variances assumption is violated which means that there is a need to specify `var.equal=FALSE` for inequal variance between datasets and use the Welch approximation of degrees of freedom.
Include your R input and output for this question here.
Paste your R input and output in this space – the box will expand automatically.

Question 3

The researchers have also hypothesised that the average pain level reported by patients will be lower six hours after the admission to the hospital than the average pain level the patients reported on the admission to the hospital. Conduct a paired samples *t*-test using the **Pain1** and **Pain2** variables to test this claim.

Produce the relevant output and write a paired samples *t*-test report.

It was hypothesised that the average pain level reported by patients will be lower six hours after admission to the hospital than the average pain level the patients reported on the admission to the hospital. In a random sample of 398 patients, the average chest pain levels of patients admitted 6 hours later is ($\bar{x}$ = 7.190 points, s = 1.577 points) is higher than the average chest pain level of patients on the admission to the hospital ($\bar{x}$ = 7.188 points, s = 1.596 points) and a paired sample t-test shows that the difference in pain levels ($\bar{x}$ = 0.002 points, s = 0.166 points) is insignificant, t(397) = -0.301. The 95% confidence interval indicates that the patients' average chest pain level after 6 hours of admission to the hospital is between 0.014 points lower and 0.019 points higher than the average chest pain level of patients reported on the admission to the hospital. This study provides insufficient evidence to conclude that there is a difference in the average chest pain level of patients after 6 hours of admission and reported on admission to the hospital.

Include your R input and output for this question here.

```
> t.test(rdm_smpl$pain1, rdm_smpl$pain2, alternative=c("greater"), paired=TRUE)

        Paired t-test

data:  rdm_smpl$pain1 and rdm_smpl$pain2
t = -0.30117, df = 397, p-value = 0.6183
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.01626728         Inf
sample estimates:
mean of the differences
           -0.002512563

> t.test(rdm_smpl$pain2, rdm_smpl$pain1, paired=TRUE)

        Paired t-test

data:  rdm_smpl$pain2 and rdm_smpl$pain1
t = 0.30117, df = 397, p-value = 0.7634
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.01388896  0.01891408
sample estimates:
mean of the differences
            0.002512563
```

```
> stat.desc(rdm_smpl$pain1)
     nbr.val     nbr.null       nbr.na          min          max        range          sum
 398.0000000    0.0000000    2.0000000    1.0000000    9.0000000    8.0000000 2861.0000000
      median         mean      SE.mean CI.mean.0.95          var      std.dev     coef.var
   7.0000000    7.1884422    0.0799853    0.1572477    2.5462641    1.5957017    0.2219816
> stat.desc(rdm_smpl$pain2)
     nbr.val     nbr.null       nbr.na          min          max        range          sum
3.980000e+02 0.000000e+00 2.000000e+00 1.000000e+00 9.000000e+00 8.000000e+00 2.862000e+03
      median         mean      SE.mean CI.mean.0.95          var      std.dev     coef.var
7.000000e+00 7.190955e+00 7.905494e-02 1.554186e-01 2.487374e+00 1.577141e+00 2.193229e-01
>
> rdm_smpl$difference=rdm_smpl$pain2-rdm_smpl$pain1
> stat.desc(rdm_smpl$difference)
      nbr.val      nbr.null        nbr.na           min           max         range           sum
398.000000000 387.000000000   2.000000000  -1.000000000   1.000000000   2.000000000   1.000000000
       median          mean       SE.mean  CI.mean.0.95           var       std.dev      coef.var
  0.000000000   0.002512563   0.008342764   0.016401518   0.027701480   0.166437615  66.242170774
>
```

Question 4

A company that produces stop-watches claims that only 0.5% of stop-watches it produces are faulty but a rival manufacturer believes that the actual percentage of stop-watches that are faulty is more than 0.5%. The research department conducted a study where a random sample of stop-watches produced by the factory was obtained and the number of stop-watches that were faulty was recorded.

The appropriate statistical test was conducted and a *p*-value of 0.101 was obtained for the left-tailed test. Based on the results of this study, it was concluded that the percentage of stop-watches that are faulty is less than 0.5%. Is this conclusion valid or not? Explain why the researchers' conclusion is valid / not valid.

The conclusion is not valid because a Binomial test shows that the difference is insignificant, p = 0.101. Thus, there is insufficient evidence to suggest that the percentage of stopwatches which are faulty is less than 0.5%