Homework chapter #2- Math 447

Nhi Vu

9/16/2021

2.18 The breaking strength of a fiber is required to be at least 150 psi. Past experience has indicated that the standard deviation of breaking strength is σ = 3 psi. A random sample of four specimens is tested, and the results are y1 = 145, y2 = 153, y3 = 150, and y4 = 147.

(a) State the hypotheses that you think should be tested in this experiment.

$H_0: \mu \geq 150$ vs $H_a: \mu < 150$

(b) Test these hypotheses using α = 0.05. What are your conclusions?

Since σ^2 is known, then we use z-test, we have:

$$z_0 = \frac{\bar{y} - \mu_0}{\sigma\sqrt{(n)}}$$

```
mu_0=150
sigma=3
y=c(145,153,150,147)
y_bar=mean(y)
n=length(y)
z_0=(y_bar-mu_0)/(sigma/sqrt(n))
z_0

## [1] -0.8333333

z_cri=qnorm(0.05,lower.tail = F)
z_cri

## [1] 1.644854
```

We have $\alpha = 0.05$, then the rejection region for H_0 in this case is $z_0 < -1.645$, so we fail to reject H_0.

The sample data provides sufficient evident to conclude that the breaking strength of a fiber is required to be at least 150 psi.

(c) Find the P-value for the test in part (b).

Since $H_a: \mu < 150$,then

$$\text{p-value} = P(Z < z_0)$$

```r
p_value=pnorm(z_0,lower.tail = TRUE)
p_value

## [1] 0.2023284
```

(d) Construct a 95 percent confidence interval on the mean breaking strength.

Confidence Interval formula in this case is:

$$\bar{y} \pm z_{\alpha/2}\frac{\sigma}{\sqrt(n)}$$

```r
#use function from libarary(TeachingDemos)
confidendce_level=0.95
library(TeachingDemos)
z.test(y, sd=sigma,conf.level = confidendce_level)

##
##  One Sample z-test
##
## data:  y
## z = 99.167, n = 4.0, Std. Dev. = 3.0, Std. Dev. of the sample mean =
## 1.5, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  145.8101 151.6899
## sample estimates:
## mean of y
##    148.75

#calcute in R
z_interval=function(y,sigma,conf.level){
  y_bar=mean(y)
  n=length(y)
  z_halfApha=qnorm((1-conf.level)/2)
  margin_error=z_halfApha*sigma/sqrt(n)
  CI=c(y_bar-margin_error,y_bar+margin_error)
  return(CI)
}
z_interval(y,3,0.95)

## [1] 151.6899 145.8101
```

There is 95 % confidence interval on the mean breaking strength is (145.8101, 151.6899)

2.19 The viscosity of a liquid detergent is supposed to average 800 centistokes at 25∘C. A random sample of 16 batches of detergent is collected, and the average viscosity is 812. Suppose we know that the standard deviation of viscosity is σ = 25 centistokes.

(a) State the hypotheses that should be tested.

$H_0: \mu_0 = 800$ vs $H_a: \mu \neq 800$

(b) Test these hypotheses using α = 0.05. What are your conclusions?

1. Quantitative data, small sample, known σ,1-sample z-test.

2. Test statistic:

$$z_0 = \frac{\bar{y} - \mu_0}{\sigma\sqrt{(n)}}$$

```
mu_0=800
sigma=25
y_bar=812
n=16
z_0=(y_bar-mu_0)/(sigma/sqrt(n))
z_0

## [1] 1.92

z_cri=qnorm(0.05/2,lower.tail = F)
z_cri

## [1] 1.959964
```

Because of Two-tail tests, the rejected region is $z_0 > z_{\alpha/2}$, but $z_0 = 1.92 < z_{\alpha/2} = 1.96$, then we fail to reject H_0. Therefore, the sample data provides sufficient evient to conlude that the population mean of viscosity of a liquid detergent is 800 centistokes at 25∘C.

(c) What is the P-value for the test?

In this case, we have

$$\text{p\_value} = 2 \times P(z > z_0)$$

```
p_value=2*pnorm(z_0,lower.tail = FALSE)
p_value

## [1] 0.0548579
```

p_value=0.055 > $\alpha = 0.05$, then we fail to reject H_0 as well.

(d) Find a 95 percent confidence interval on the mean.

1. Using z-interval.

2. Confidence interval formula:

$$\bar{y} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{(n)}}$$

```
margin_error=z_cri*sigma/sqrt(n)
CI=c(y_bar-margin_error,y_bar+margin_error)
CI

## [1] 799.7502 824.2498
```

We can be 95% confident that the population mean of viscosity of a liquid detergent falls between 799.7502 and 824.2498 centistokes at 25∘C.

2.21 A normally distributed random variable has an unknown mean μ and a known variance $\sigma^2 = 9$. Find the sample size required to construct a 95 percent confidence interval on the mean that has a total length of 1.0.

1. Quantitative data, known variance, z-interval.

2. Confidence interval formula:

$$\bar{y} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{(n)}}$$

3. Then:

$$\text{Upper bound} - \text{Lower bound} = 2 \times z_{\alpha/2} \frac{\sigma}{\sqrt{(n)}}$$

$$n = \frac{(2 \times z_{\alpha/2} \times \sigma)^2}{(\text{Upper bound} - \text{Lower bound})^2}$$

```
difference=1
z_cri=qnorm((1-0.95)/2,lower.tail = F)
sigma=sqrt(9)
n=(2*z_cri*sigma)^2/(difference)^2
n

## [1] 138.2925
```

In this case, the required sample size is approximately to 139.

2.26 Two machines are used for filling plastic bottles with a net volume of 16.0 ounces. The filling processes can be assumed to be normal, with standard deviations of $\sigma_1 = 0.015$ and $\sigma_2 = 0.018$. The quality engineering department suspects that both machines fill to the same net volume, whether or not this volume is 16.0 ounces. An experiment is performed by taking a random sample from the output of each machine.

```
machine_1=c(16.03,16.01,16.04,15.96,16.05,15.98,16.05,16.02,16.02,15.99)
machine_2=c(16.02,16.03,15.97,16.04,15.96,16.02,16.01,16.01,15.99,16.00)
```

a) State the hypotheses that should be tested in this experiment.

$H_0: \mu_1 = \mu_2$ vs $H_a: \mu_1 \neq \mu_2$

(b) Test these hypotheses using α = 0.05. What are your conclusions?

1. Quantitative data, small sample n_1, n_2, known variances, z-test.

2. Test statistic

$$Z_0 = \frac{\bar{Y_1} - \bar{Y_2}}{\sqrt{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}}$$

3. Rejection region: Two-tail test: $Z_0 > Z_{\alpha/2}$

```
ybar_1=mean(machine_1)
ybar_2=mean(machine_2)
n1=length(machine_1)
n2=length(machine_2)
sigma_1 = 0.015
sigma_2= 0.018
z_0=(ybar_1-ybar_2)/(sqrt(sigma_1^2/n1+sigma_2^2/n2))
z_0

## [1] 1.349627

z_cri=qnorm(0.05/2,lower.tail = F)
z_cri

## [1] 1.959964
```

Since $z_0 = 1.349 < z_{\alpha/2} = 1.96$, we fail to reject H_0

4. The sample datas provide sufficient evidence to conclude that both machines fill to the same net volume

(c) Find the P-value for this test.

```
p_value=2*pnorm(z_0,lower.tail = F)
p_value
```

```
## [1] 0.1771356
```

(d) Find a 95 percent confidence interval on the difference in mean fill volume for the two machines.

Confidence Interval:

$$(\bar{y_1} - \bar{y_2}) \pm z_{\alpha/2}\sqrt{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}$$

```
d=ybar_1-ybar_2
marrign_error=z_cri*(sqrt(sigma_1^2/n1+sigma_2^2/n2))
CI=c(d-marrign_error,d+marrign_error)
CI

## [1] -0.004522262  0.024522262
```

We can be 95% confident that the difference of 2 population mean falls between -0.004522262 and 0.024522262.

2.28 The following are the burning times (in minutes) of chemical flares of two different formulations. The design engineers are interested in both the mean and variance of the burning times.

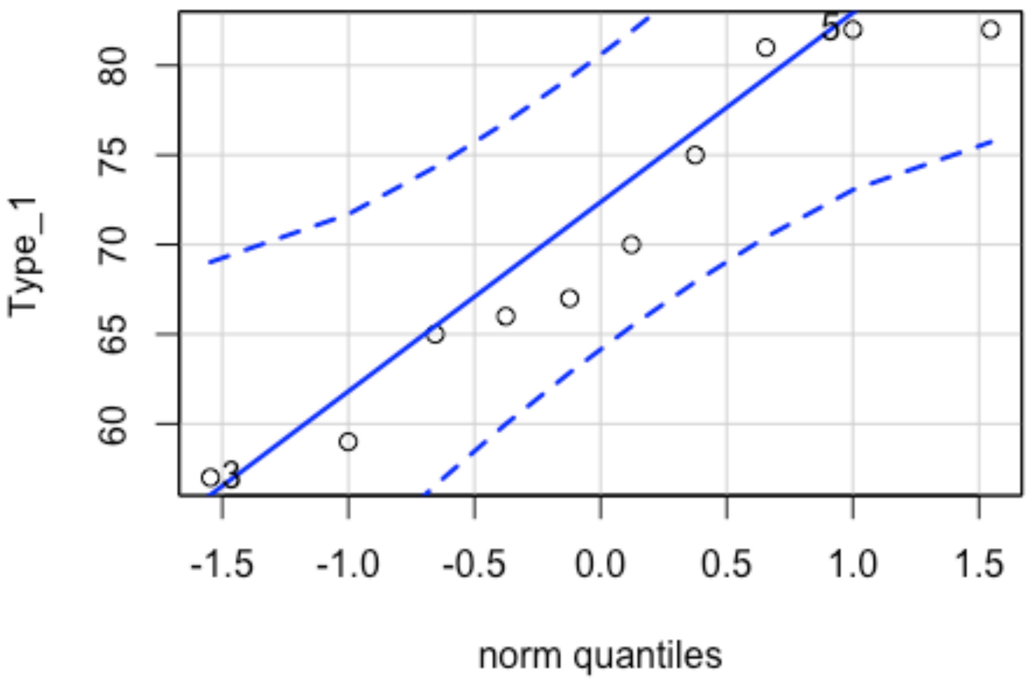
```
Type_1=c(65,81,57,66,82,82,67,59,75,70)
Type_2=c(56,69,74,82,79,64,71,83,59,65)
```

(a) Test the hypothesis that the two variances are equal. Use α = 0.05.

1. Quantitative data, small samples.

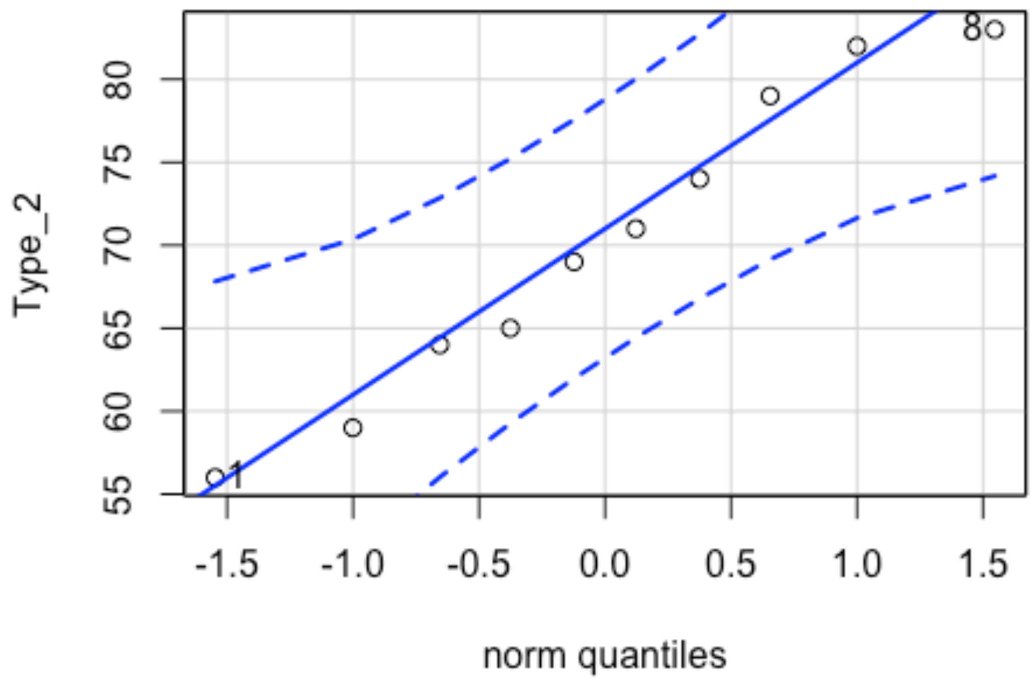
First, since data are small, we check normal assumption:

```
library(car)

## Loading required package: carData

qqPlot(Type_1)
```




```
## [1] 3 5
```

```
qqPlot(Type_2)
```




```
## [1] 1 8
```

From the graphs, we can conclude that sample datas are normally distribution.

2. $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_a: \sigma_1^2 \neq \sigma_2^2$
3. Using F-test:

$$F = \frac{s_1^2}{s_2^2}$$

```
var1=var(Type_1)
var2=var(Type_2)
F=var1/var2
n1=length(Type_1)
n2=length(Type_2)
p_value=2*pf(F,n1-1,n2-1,lower.tail = F)
p_value

## [1] 1.025634
```

Since p_value= 1.0256 > $\alpha = 0.05$ then we fail to reject H_0

4. The data samples provide sufficient evidence to conclude that the variance of the burning times (in minutes) of chemical flares of two different formulations are equal.

(b) Using the results of (a), test the hypothesis that the mean burning times are equal. Use α = 0.05. What is the P-value for this test?

From part (a), we know that the variance of 2 population are equal, then we use 2 sample pooled t test to test mean of 2 populations.

1. $H_0: \mu_1 = \mu_2$ vs $H_a: \mu_1 \neq \mu_2$
2. Test statistic:

$$t_0 = \frac{\bar{y_1} - \bar{y_2}}{S_p\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$S_p = \sqrt{(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2})}$$

```
ybar1=mean(Type_1)
ybar2=mean(Type_2)
sp=sqrt((n1-1)*var1+(n2-1)*var2)/sqrt(n1+n2-2)
t0=(ybar1-ybar2)/(sp*sqrt(1/n1+1/n2))
pvalue=2*pt(abs(t0),n1+n2-2,lower.tail = F)
pvalue

## [1] 0.9622388
```

3. Since pvalue= 0.9622388 > $\alpha = 0.05$, then we fail to reject H_0
4. The data samples provide sufficient evidence to conclude that the mean burning times of 2 different formulation are equal.

(c) Discuss the role of the normality assumption in this problem. Check the assumption of normality for both types of flares.

Because F-tests assume a normal distribution and will result in inaccurate results if the data differs significantly from this distribution. But from part (a), we check normal assumption by creating normal probability plot (qqplot). In both graphs, data points lie along a straight line that we can assume the data are normal distribution.

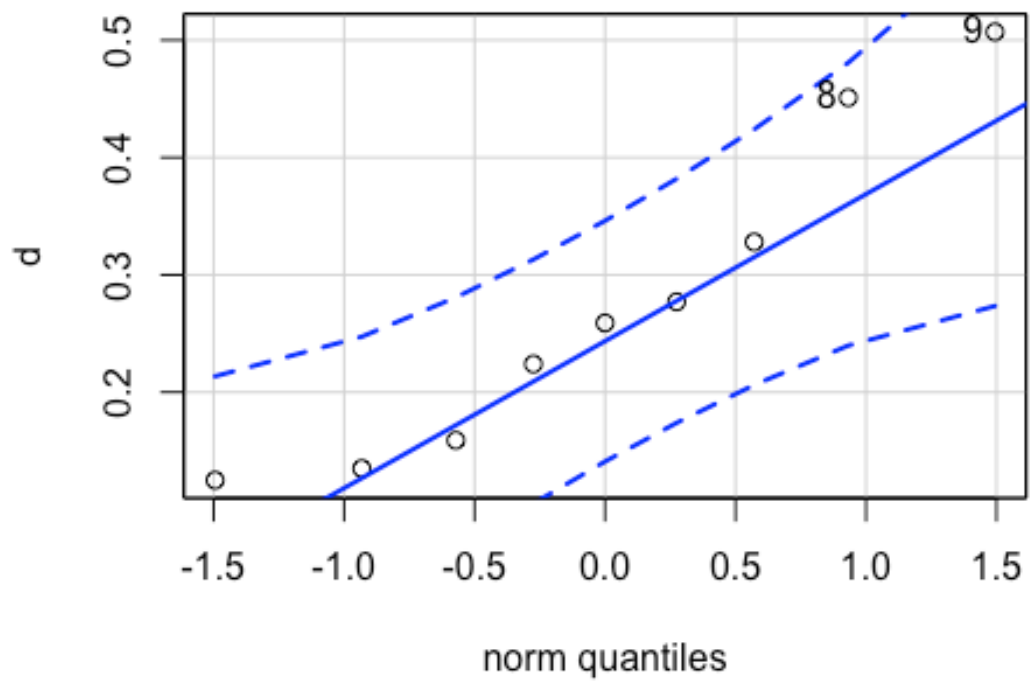
2.36 An article in the Journal of Strain Analysis (Vol. 18, no. 2, 1983) compares several procedures for predicting the shear strength for steel plate girders. Data for nine girders in the form of the ratio of predicted to observed load for two of these procedures, the Karlsruhe and Lehigh methods, are as follows:

```
Karlsruhe_Method=c(1.186,1.151,1.322,1.339,1.200,1.402,1.365,1.537,1.559)
Lehigh_Method=c(1.061,0.992,1.063,1.062,1.065,1.178,1.037,1.086,1.052)
d=Karlsruhe_Method-Lehigh_Method
```

(a) Is there any evidence to support a claim that there is a difference in mean performance between the two methods? Use α = 0.05.

$H_0: \mu_1 - \mu_2 = \mu_d = 0$ vs $H_a: \mu_d \neq 0$ 1. checking normality assumption for difference of 2 methods.

```
library(car)
qqPlot(d)
```




```
## [1] 9 8
```

All the points fall approximately along the (45-degree) reference line, for. So we can assume normality of the difference data of 2 methods.

2. Using paired t test: Test statistic:

$$t_0 = \frac{\bar{d}}{s_d/\sqrt(n)}$$

$$s_d^2 = \frac{\sum(d_i - \bar{d})^2}{n-1}$$

```
dbar=mean(d)
sd=sd(d)
n=length(d)
```

```
t0=dbar/(sd/sqrt(n))
tcri=qt(0.05/2,n-1,lower.tail = F)
```

3. construct reject region: |t0|>tcri, reject H_0.

4.There is sufficient evidence to conclude that there is a difference in mean performance between the two methods.

(b) What is the P-value for the test in part (a)?

```
pvalue=2*pt(abs(t0),n-1, lower.tail=F)
pvalue

## [1] 0.0002952955
```

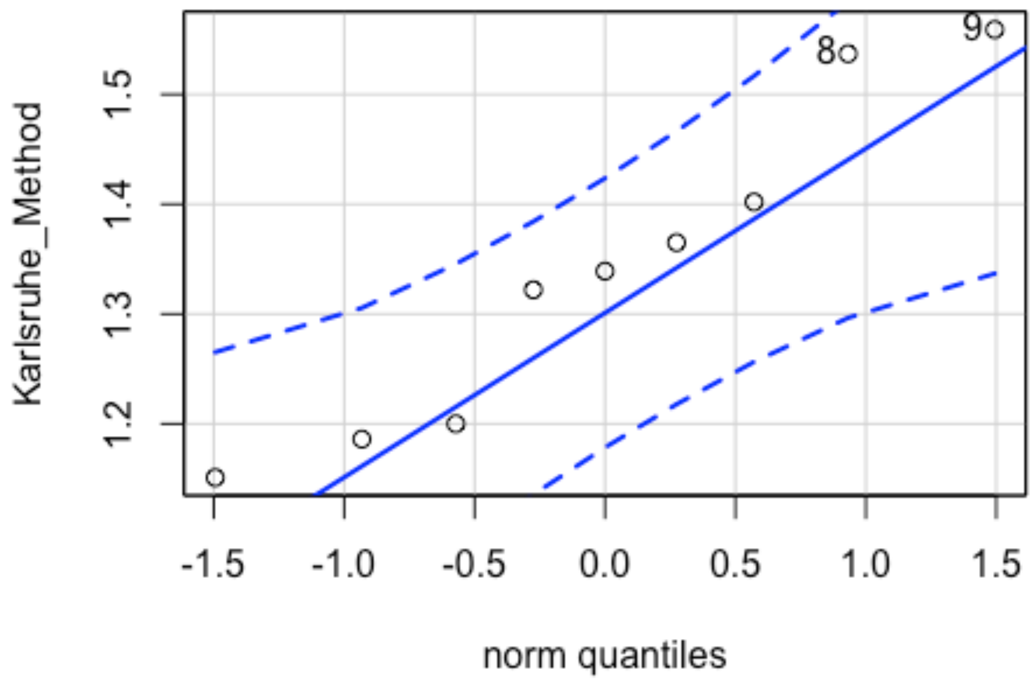
(c) Construct a 95 percent confidence interval for the difference in mean predicted to observed load.

```
t.test(Karlsruhe_Method,Lehigh_Method, alternative="two.sided", paired=T,
conf.level=0.95)

##
##  Paired t-test
##
## data:  Karlsruhe_Method and Lehigh_Method
## t = 6.0819, df = 8, p-value = 0.0002953
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1700423 0.3777355
## sample estimates:
## mean of the differences
##               0.2738889
```

We can be 95% confident to conclude that the difference in mean of 2 methods falls between 0.1700423 and 0.3777355

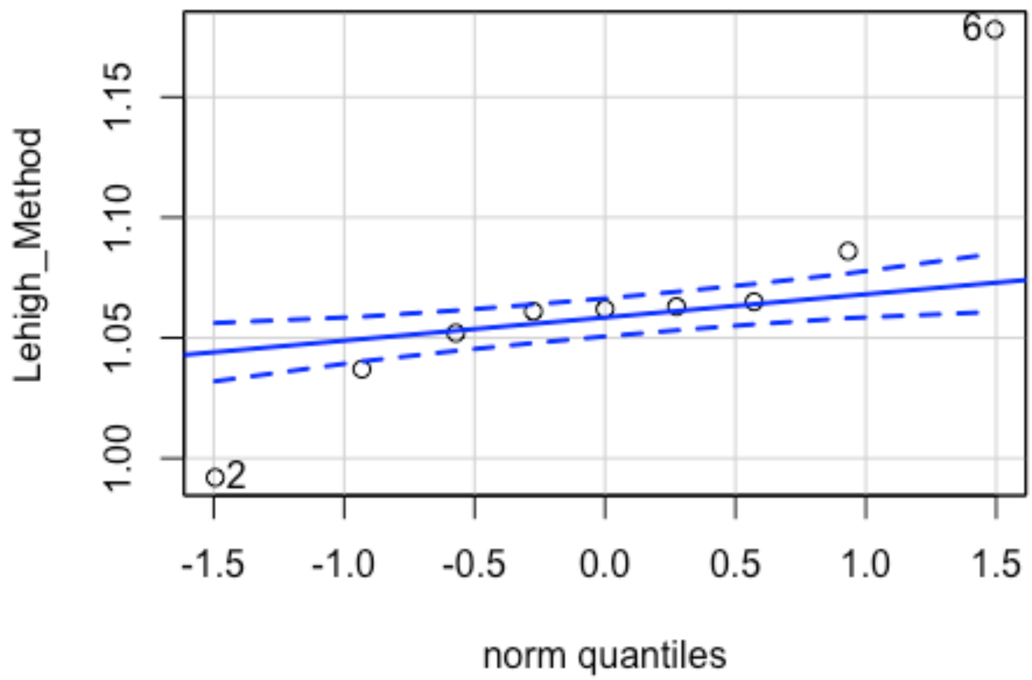
(d) Investigate the normality assumption for both samples.

```
library(car)
qqPlot(Karlsruhe_Method)
```




```
## [1] 9 8
```

```
qqPlot(Lehigh_Method)
```

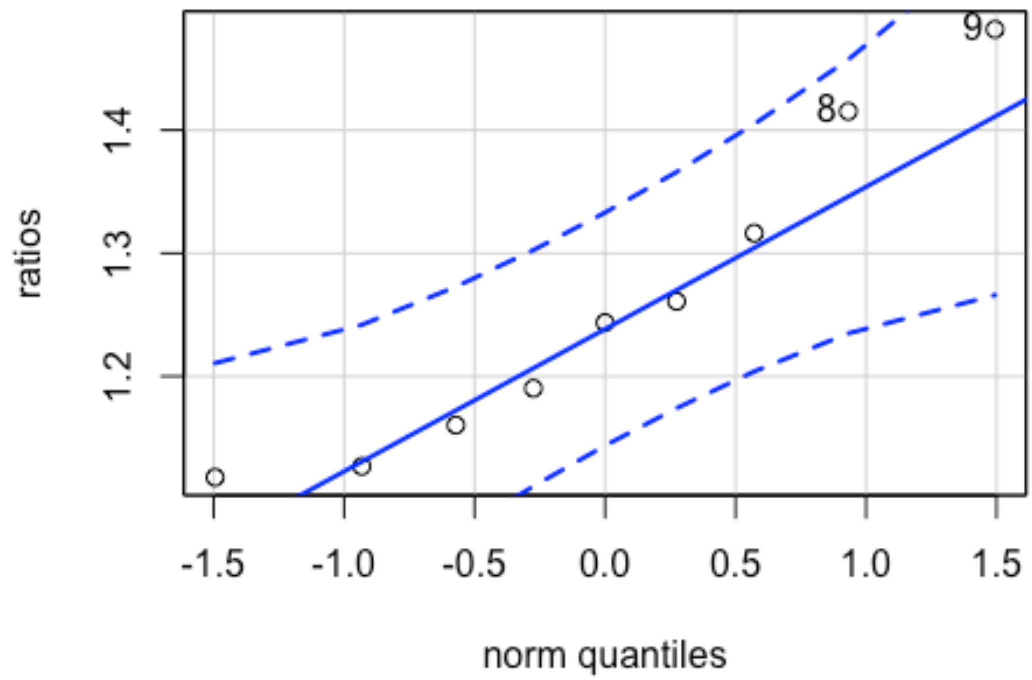



```
## [1] 6 2
```

As all the points fall approximately along this reference line, we can assume normality for both samples.

(e) Investigate the normality assumption for the difference in ratios for the two methods.

```
ratios=Karlsruhe_Method/Lehigh_Method
qqPlot(ratios)
```




```
## [1] 9 8
```

As all the points fall approximately along this reference line, we can assume normality for the difference in ratios for the two methods.

(f) Discuss the role of the normality assumption in the paired t-test.

The paired t–test assumes that the differences between pairs are normally distributed, not observations within each group are normal, so if the differences between pairs are not normally distributed, the result will be inaccurate.

2.49 Reconsider the bottle filling experiment described in Problem 2.26. Rework this problem assuming that the two population variances are unknown but equal.

```
machine_1=c(16.03,16.01,16.04,15.96,16.05,15.98,16.05,16.02,16.02,15.99)
machine_2=c(16.02,16.03,15.97,16.04,15.96,16.02,16.01,16.01,15.99,16.00)
```

1. $H_0: \mu_1 = \mu_2$ vs $H_a: \mu_1 \neq \mu_2$
2. Quantitative data, small sample n_1, n_2, unknown variances but equal, t-test. Test statistic:

$$t_0 = \frac{\bar{y_1} - \bar{y_2}}{S_p\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$S_p = \sqrt{(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2})}$$

```
ybar1=mean(machine_1)
ybar2=mean(machine_2)
n1=length(machine_1)
n2=length(machine_2)
var1=var(machine_1)
var2=var(machine_2)
sp=sqrt((n1-1)*var1+(n2-1)*var2)/sqrt(n1+n2-2)
t0=(ybar1-ybar2)/(sp*sqrt(1/n1+1/n2))
pvalue=2*pt(abs(t0),n1+n2-2,lower.tail = F)
pvalue

## [1] 0.4347438
```

3. Since p_value=0.4347 > 0.05, fail to reject H_0

4. There is a sufficient evidence that both machines fill to the same net volume.

```
t.test(machine_1,machine_2, alternative="two.sided", paired=F,
conf.level=0.95)

##
##  Paired t-test
##
## data:  machine_1 and machine_2
## t = 0.62284, df = 9, p-value = 0.5488
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02631997  0.04631997
## sample estimates:
## mean of the differences
##                    0.01
```