Stat 5309 HW 1

Tom Wilson

January 31, 2019

# 1

A new filtering device is installed ina chemical unit. Before its installation, a random sample yielded the following information about the percentage of impurity: , , and . After installation, a random sample yielded , ,

## a

Can you conclude that the two variances are equal? Use

The null hypothesis is that the two variances are equal, . The ratio of the sample variances is F-distributed with n-1 and n-2 degrees of freedom.

bar\_y\_1 <- 12.5  
s\_1\_squared <- 101.17  
n\_1 <- 8  
nu\_1 <- n\_1 - 1  
#After installation  
bar\_y\_2 <- 10.2  
s\_2\_squared <- 94.73  
n\_2 <- 9  
nu\_2 <- n\_2 - 1  
  
f\_stat <- s\_1\_squared / s\_2\_squared  
left\_tail <- pf(1/f\_stat,nu\_1,nu\_2)  
right\_tail <- 1 - pf(f\_stat,nu\_1,nu\_2)   
p\_value <- left\_tail + right\_tail  
p\_value

## [1] 0.9307519

Under the null hypothesis, we would expect a difference in means at least this extreme 93% of the time. At a confidence level of 5%, we would fail to reject the null hypothesis and conclude that the variances are equal.

## b

Has the filtering device reduced the percentage of impurity significantly? Use .

The null hypothesis is that the two averages are equal, . The distribution of differences is t-distributed with pooled variance and degrees of freedom.

mean\_difference <- bar\_y\_1 - bar\_y\_2  
pooled\_variance <- ((n\_1-1)\*s\_1\_squared + (n\_2-1)\*s\_2\_squared)/(n\_1+n\_2-2)  
t\_stat <- mean\_difference/(pooled\_variance\*sqrt(1/n\_1 + 1/n\_2))  
df <- n\_1 + n\_2 - 2  
  
alpha <- 0.05  
half\_alpha <- alpha/2  
t\_critical\_left <- qt(half\_alpha,df)  
t\_critical\_right <- qt(1-half\_alpha,df)   
p\_value <- pt(-abs(t\_stat),df) + 1-pt(abs(t\_stat),df)

Under the null hypothesis, we would expect a difference in means at least this extreme 96% of the time. At a confidence level of 5%, we fail to reject the null hypothesis and conclude that the filtering device has not significantly reduced impurity.

# 2

Twenty observations on etch uniformity on silicon wafers are taken during a qualification experiment for a plasma etcher. The data are as follows.

uniformity <- c(5.34,6.65,4.76,5.98,7.25,  
 6.00,7.55,5.54,5.62,6.21,  
 5.97,7.35,5.44,4.39,4.98,  
 5.25,6.35,4.61,6.00,5.32  
 )

## a

construct a 95% confidence interval estimate of .

Assuming that the population is normally distributed, will be distributed with n-1 degrees of freedom.

n <- length(uniformity)  
df <- n-1  
bar\_x <- mean(uniformity)  
s\_squared <- var(uniformity)  
s <- sqrt(s\_squared)  
sum\_of\_squares <- df\*s\_squared  
chi\_stat <- sum\_of\_squares/s\_squared  
alpha <- 0.05  
half\_alpha <- alpha/2  
chi\_critical\_left <- qchisq(half\_alpha,df)  
chi\_critical\_right <- qchisq(1-half\_alpha,df)  
  
right\_bound = sum\_of\_squares/chi\_critical\_left  
left\_bound = sum\_of\_squares/chi\_critical\_right  
  
paste(left\_bound,right\_bound)

## [1] "0.457152397856341 1.68623951130305"

The best estimate of is the sample variance of 0.79. If this experiment were repeated, we can expect 95% of the sample variances to be between 0.4571 and 1.6892

## b

test the hypothesis that . Use . What are your conclusions?

s\_0 <- 1.0  
chi\_stat\_right <- sum\_of\_squares/s\_0  
chi\_stat\_left <- s\_0/sum\_of\_squares  
alpha <- 0.05  
half\_alpha <- alpha/2  
left\_tail <- pchisq(chi\_stat\_left,df)  
right\_tail <- 1 - pchisq(chi\_stat\_right,df)  
p\_value <- left\_tail + right\_tail  
p\_value

## [1] 0.7214195

Given the null hypothesis that the true variance is 1.0, we would expect to observe a statistic that is at least this extreme 72% of the time. at a confidence level of 5%, we fail to reject the null hypothesis and conclude that the variance is not significantly different from 1.0.

## c

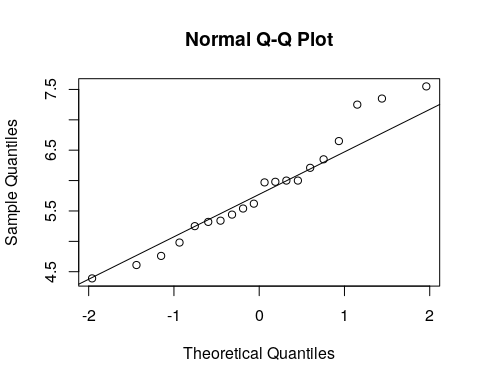
Discuss the normality assumption and its role in this problem.

In order to know the distribution of we must make some assumption about the distribution of the measurement itsself. The most reasonable assumption is that the population is normally distributed because uniformity is a function of many inputs.

## d

Check normality by constructing a normal probability plot. What are your conclusions?

qqnorm(uniformity)  
qqline(uniformity)



The quantiles of the uniformity measurements very nearly match that of a normal distribution. The three largest values are larger than expected, but not significantly.

# 3

The diameter of a ball bearing was measured by 12 inspectors, each using two different kinds of calipers. The results were:

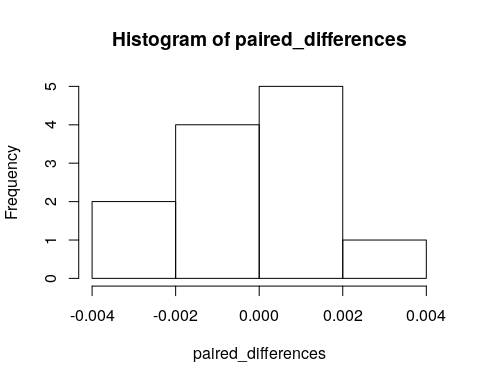
bearings <- data.frame(inspector = seq(1,12),  
 caliper1 = c(0.265,0.265,0.266,0.267,0.267,0.265,0.267,0.267,0.265,0.268,0.268,0.265),  
 caliper2 = c(0.264,0.265,0.264,0.266,0.267,0.268,0.264,0.265,0.265,0.267,0.268,0.269)  
 )

## a

Is there a significant difference between the means of the population of measurements from which the two samples were selected? Use .

Since each pair of measurements is taken by the same inspector a paired t-test is appropriate.

paired\_differences <- bearings$caliper1 - bearings$caliper2  
hist(paired\_differences)



The null hypothesis is that the paired difference is zero, . The distribution of differences is t-distributed with variance estimated by the variance of the differences and degrees of freedom.

bar\_d <- mean(paired\_differences)  
n <- length(bearings$caliper1)  
s\_d <- sd(paired\_differences)  
nu <- n-1  
SE <- s\_d/sqrt(n)  
t\_stat <- bar\_d/(SE)  
alpha <- 0.05  
half\_alpha <- alpha/2  
t\_critical\_left <- qt(half\_alpha,nu)  
t\_critical\_right <- qt(1-half\_alpha,nu)   
t\_stat

## [1] 0.4317878

paste(t\_critical\_left,t\_critical\_right)

## [1] "-2.20098516009164 2.20098516009164"

The t-statistic of 0.43 is between the critical t-statistics of . Therefore, we fail to reject the null hypothesis and conclude that the measurements from the two calipers are not significantly different.

## b

Find the p-value for the test in part a.

p\_value <- pt(-abs(t\_stat),nu) + 1-pt(abs(t\_stat),nu)  
p\_value

## [1] 0.6742372

The p-value is about 67%. Given that there is no difference, we can expect a difference at least this large 67% of the time.

## c

Construct a 95% confidence interval on the difference in mean diameter measurements for the two types of calipers.

t\_critical\_left <- qt(half\_alpha,nu)  
t\_critical\_right <- qt(1-half\_alpha,nu)   
left\_bound <- t\_critical\_left \* SE  
right\_bound <- t\_critical\_right \* SE  
paste(left\_bound,right\_bound)

## [1] "-0.00127434431630026 0.00127434431630026"

At a confidence level of 95%, a mean difference with magnitude larger than 0.0012 is statistically different from zero.