Act 8z/t test: Stat 463/563

Dr. Islam

Date: Nov 10, 2022

Last Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ First Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Test of equality of two population means and

Given two populations with unknown means and , the common practice while doing any statistical inference is to test against alternatives

1. ,
2. and

*Case I*

Under the assumption that the two population distributions are normal with known variances and , the test of aforementioned null hypothesis is known as the Z-test.

Of course, to implement this test we require two independent samples from two populations.

Given two independent samples of size from distribution and of size from distribution, the Z-test is implemented by the statistic

where and are means of the samples and .

**Conclusion:**

We reject the null hypothesis if of the Z-test is less than , the level of significance.

The -value of the -test in relation to the null hypothesis is computed as follows:

* -value when (i.e., left-tailed test)
* -value when (i.e., right-tailed test)
* -value when (i.e., two-tailed test)

where is a N(0,1) distribution and is the observed value of the test statistic.

**Remarks**

* We either accept or reject the null hypothesis.
* depends on the alternative hypotheses.

*Case II*

Under the assumption that the two population distributions are normal with unknown, but equal variances, i.e., , say, the test of aforementioned null hypothesis is known as the Pooled T-test given by

where is the pooled estimate of the common population variance given by

and and are the sample variance of the samples x and y.

**Conclusion:**

We reject the null hypothesis if of the T-test is less than , the level of significance.

The -value of the -test in relation to the null hypothesis is computed as follows:

* -value when (i.e., left-tailed test)
* -value when (i.e., right-tailed test)
* -value when (i.e., two-tailed test)

where refers to a -distribution with and and t is the observed value of the test statistic.

*Case III*

If the population sds and are assumed unknown and unequal, then the test statistic to test the null hypothesis is given by

which, under , approximately follows a -distribution with approximate given by

This test is called the **2-Sample Non-pooled T-Test**. To compute the , follow the same procedure as *Case II* with degrees of freedom of the -distribution is updated by rounding its value to the nearest integer value.

# Activity 8.1: Application of Z-test

A researcher believes that the Verizon customers pay more for their bills per month than the Sprint customers. To test the belief via TOH, the researcher takes two independent samples, one of 12 Verizon customers and the other of 15 Sprint customers. Payment amounts, rounded to the nearest $ amount, for two samples of customers are as follows:

verizon=129 135 145 143 133 137 161 147 146 138 170 165  
sprint=156 132 139 127 150 155 145 154 148 127 145 160 127 143 126

The monthly bills of Verizon and Sprint customers follow *normal distributions* with **standard deviations** 10 and 12, respectively.

Write your own function to print output for the following questions, where appropriate.

1. How would you specify hypotheses for this test? In other words, specify null and alternative hypothesis.
2. What test should you use for this test? Why? Write the form of the test statistic.
3. Does the value of the test statistic depend on the alternative?
4. Does the value of the test statistic depend on the null hypothesis?
5. Find the value of the test statistic given the two samples.
6. Find the p-value of the test. Does the p-value depend on the null or alternative hypothesis?
7. Find the critical-value of the test. Does the critical value depend on the null or alternative hypthesis? What else does the critical value depend on?
8. What is the critical value for the test at 5% and 10% level of significance?
9. What conclusion do you make about the acceptance or rejection of the null hypothesis at 5% level of significance?

# Program to implement a 2-sample z-test

zt.funct<-function(x,y,sigma1,sigma2, alt,alpha){  
mx=mean(x)  
my=mean(y)  
m=length(x);  
n=length(y);  
vars=(sigma1^2/m)+(sigma2^2/n)  
z=(mx-my)/sqrt(vars)  
p<-pnorm(z)  
if (alt== 'l') {pval<-p; cval<-qnorm(alpha, lower.tail = T)}  
else if (alt== 'g') {pval<-(1-p); cval<-qnorm(alpha, lower.tail = F)}  
else if (alt== 't') {pval<-2\*pnorm(abs(z),lower.tail=F);cval<- abs(qnorm(alpha/2, lower.tail = F))}  
conclusion=ifelse(pval>alpha,"acpt null", "rejt null")  
out=list(z=z,pvalue=pval,CritcalValue=cval,conclusion=conclusion)  
return(out)}  
  
x=c(129, 135, 145, 143, 133, 137, 161, 147, 146, 138, 170, 165)  
y=c(156, 132, 139, 127, 150, 155, 145, 154, 148, 127, 145, 160, 127, 143, 126)  
  
zt.funct(x,y,10,12,"l",0.05)

$z  
[1] 0.8225542  
  
$pvalue  
[1] 0.7946192  
  
$CritcalValue  
[1] -1.644854  
  
$conclusion  
[1] "acpt null"

zt.funct(x,y,10,12,"g",0.05)

$z  
[1] 0.8225542  
  
$pvalue  
[1] 0.2053808  
  
$CritcalValue  
[1] 1.644854  
  
$conclusion  
[1] "acpt null"

zt.funct(x,y,10,12,"t",0.05)

$z  
[1] 0.8225542  
  
$pvalue  
[1] 0.4107616  
  
$CritcalValue  
[1] 1.959964  
  
$conclusion  
[1] "acpt null"

# Activity 8.2: Application of T-test

A bank management is unsure if it will staff two of its branch offices with the same number of staff. It collects data for the number of transactions reported on each Monday for two months, from both branch offices. The data below refers to the number of transactions in two offices:

Office 1: 276, 323, 298, 256, 277, 309, 312, 265, 311

Office 2: 251, 320, 318, 247, 301, 279, 267, 315, 299

The management has information that the number of transactions in both offices follow *normal distributions* with **unknown but equal variances**. Given this information, the management is interested to test against

1. ,
2. and

at 5% level of significance.

1. What would be the appropriate test in this scenario? Why?
2. Write the test statistic of your choice.
3. Does the value of the test statistic depend the alternative?
4. Does the value of the test statistic depend on the null hypothesis?
5. Find the value of the test statistic given the two samples.
6. What alternative do you think is the most suitable for this problem?
7. Find the p-value of the test. What does the p-value depend on for this test?
8. What is your conclusion about the acceptance or rejection of the null hypothesis at 5% level of significance?

**Solution**

We should implement a 2-sample pooled T-test for this test. We define a function, callled t.funct, for this test.

t.funct<-function(x,y,alt,alpha){  
mx=mean(x)  
my=mean(y)  
n1=length(x);  
n2=length(y);  
df=n1+n2-2;  
var=((n1-1)\*var(x)+(n2-1)\*var(y))/df  
sp=sqrt(var)  
t=(mx-my)/(sp\*sqrt(1/n1+1/n2))  
p<-pt(t,df)  
if (alt== 'l') {pval<-p;cval<-qt(alpha,df, lower.tail = T)}  
else if (alt== 'g') {pval<-(1-p);cval<-qt(alpha,df, lower.tail = F)}  
else if (alt== 't') {pval<-2\*pt(abs(t),df,lower.tail=F);cval<-qt(alpha/2,df, lower.tail = F)}  
conclusion.pv=ifelse(pval>alpha,"acpt null", "rejt null")  
conclusion.cv=ifelse(abs(t)>abs(cval),"rejt null", "acpt null")  
  
out=list(t=t,df=df,pvalue=pval,CritcalValue=cval,conclusion.cval=conclusion.cv,"Conclusion.pval"=conclusion.pv)  
return(out)}  
x<-c(276, 323, 298, 256, 277, 309, 312, 265, 311)  
y<-c(251, 320, 318, 247, 301, 279, 267, 315, 299)  
t.funct(x,y,"l",0.05)

$t  
[1] 0.2688853  
  
$df  
[1] 16  
  
$pvalue  
[1] 0.6042729  
  
$CritcalValue  
[1] -1.745884  
  
$conclusion.cval  
[1] "acpt null"  
  
$Conclusion.pval  
[1] "acpt null"

t.funct(x,y,"g",0.05)

$t  
[1] 0.2688853  
  
$df  
[1] 16  
  
$pvalue  
[1] 0.3957271  
  
$CritcalValue  
[1] 1.745884  
  
$conclusion.cval  
[1] "acpt null"  
  
$Conclusion.pval  
[1] "acpt null"

t.funct(x,y,"t",0.05)

$t  
[1] 0.2688853  
  
$df  
[1] 16  
  
$pvalue  
[1] 0.7914543  
  
$CritcalValue  
[1] 2.119905  
  
$conclusion.cval  
[1] "acpt null"  
  
$Conclusion.pval  
[1] "acpt null"

#Test using existing R function t.test;  
t.test(x,y,alt="l",var.equal=T);

Two Sample t-test  
  
data: x and y  
t = 0.26889, df = 16, p-value = 0.6043  
alternative hypothesis: true difference in means is less than 0  
95 percent confidence interval:  
 -Inf 24.97681  
sample estimates:  
mean of x mean of y   
 291.8889 288.5556

t.test(x,y,alt="g",var.equal=T);

Two Sample t-test  
  
data: x and y  
t = 0.26889, df = 16, p-value = 0.3957  
alternative hypothesis: true difference in means is greater than 0  
95 percent confidence interval:  
 -18.31014 Inf  
sample estimates:  
mean of x mean of y   
 291.8889 288.5556

t.test(x,y,alt="t",var.equal=T)

Two Sample t-test  
  
data: x and y  
t = 0.26889, df = 16, p-value = 0.7915  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -22.94683 29.61350  
sample estimates:  
mean of x mean of y   
 291.8889 288.5556