

ADA HW2

Liangquan Zhou lz2377

Consider the data on Chicken Weights by Feed Type (`chickwts`) in R library `MASS`

1. Determine whether there is a significant difference in the mean weights of chicks fed meatmeal vs. those fed casein using each of the following procedures:

- a). A parametric procedure
- b). A non-parametric procedure
- c). A re-sampling procedure

Discuss the assumption underlying each of the analyses, their validity, and any remedial measures to be taken.

Solution

a). The hypothesis test:

$$H_0 : \mu_m - \mu_c = 0 \quad H_1 : \mu_m - \mu_c \neq 0$$

Assumptions:

- 1. The weights of chicks fed on meatmeal and casein are independent.
- 2. The meatmeal and casein data are normally distributed.
- 3. The standard deviation of them are the same.

Check assumptions: Use Shapiro-Wilk Normality Test to test the normality of two samples.

```
library(MASS)
data(chickwts)
meatmeal=chickwts[chickwts[,2]=="meatmeal",1]
casein=chickwts[chickwts[,2]=="casein",1]
```

```
shapiro.test(meatmeal)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  meatmeal
## W = 0.9791, p-value = 0.9612
```

```
shapiro.test(casein)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  casein
## W = 0.9166, p-value = 0.2592
```

And use F-test for standard deviations:

```
var.test(meatmeal,casein)
```

```
##
## F test to compare two variances
##
## data: meatmeal and casein
## F = 1.014, num df = 10, denom df = 11, p-value = 0.9739
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.2878 3.7182
## sample estimates:
## ratio of variances
## 1.015
```

Since these p -values are all greater than 0.05, we can think the assumption is reasonable.

And based on three assumptions above, we can use two sample t-test

```
t.test(meatmeal,casein,var.equal=T)
```

```
##
## Two Sample t-test
##
## data: meatmeal and casein
## t = -1.729, df = 21, p-value = 0.09842
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -102.801 9.453
## sample estimates:
## mean of x mean of y
## 276.9 323.6
```

From the result we can see that the p -value is greater than 0.05, so we do not reject H_0 .

b). For a non-parametric procedure, we use Wilcoxon Rank-Sum Test.

```
wilcox.test(meatmeal,casein)
```

```
##
## Wilcoxon rank sum test
##
## data: meatmeal and casein
## W = 38, p-value = 0.09084
## alternative hypothesis: true location shift is not equal to 0
```

The p -value is also greater than 0.05, so we do not reject H_0 .

c). Use bootstrap methods. Resampling `Meatmeal` and `Carsein` separately, and for each iteration we compare the statistic with the statistic from the observed dataset.

```

set.seed(1)
casein.bar= casein + mean(meatmeal) - mean(casein)
z.obs=(mean(meatmeal)-mean(casein))/sqrt(var(meatmeal)/length(meatmeal)+var(casein)/length(casein))
z.temp=c()
for(i in 1:1000){
  meatmeal.temp=sample(meatmeal,length(meatmeal),replace=T)
  casein.temp=sample(casein.bar,length(casein),replace=T)
  z.temp[i]=(mean(meatmeal.temp)-mean(casein.temp))/sqrt(var(meatmeal.temp)/length(meatmeal)+var(casein.temp)/length(casein.temp))
}
p.value=sum(abs(z.temp)>=abs(z.obs))/1000
p.value

```

```
## [1] 0.091
```

And the result is accord with the conclusion in (a) and (b). The p -value here is also greater than 0.05, then we should not reject H_0 .

2. Using the data for chicks fed soybean and those on sunflower, compute the following, based on a suitable bootstrap method:

- A 95% confidence interval for the difference in median weight for the two groups
- A 95% bootstrap confidence interval for the ratio of the variances of soybean fed to sunflower fed chicks
- A 95% confidence interval for the ratio of the variances of soybean fed to sunflower fed chicks

Discuss the assumption underlying each of the analyses, their validity, and any remedial measures to be taken

Solution

- The approach is to create a large number of samples from this pseudo-population using bootstrap. Then we use quantile function to find the 95% difference interval.

```

soybean=chickwts[chickwts[,2]=="soybean",1]
sunflower=chickwts[chickwts[,2]=="sunflower",1]
library(bootstrap)
median.soy=bootstrap(soybean,1000,median)
median.sun=bootstrap(sunflower,1000,median)
quantile(abs(median.soy$thetastar-median.sun$thetastar),c(0.025,0.975))

```

```
## 2.5% 97.5%
## 45.96 118.50
```

- Same method as (a).

```

var.soy= bootstrap(soybean,1000,var)
var.sun = bootstrap(sunflower,1000,var)
quantile(var.soy$thetastar/var.sun$thetastar,c(0.025,0.975))

```

```
## 2.5% 97.5%
## 0.3979 8.6492
```

c). First test the normality and then use `var.test` to get the 95% confidence interval.

```
shapiro.test(soybean)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  soybean  
## W = 0.9464, p-value = 0.5064
```

```
shapiro.test(sunflower)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  sunflower  
## W = 0.9281, p-value = 0.3603
```

```
var.test(soybean,sunflower)$conf.int
```

```
## [1] 0.3622 3.9281  
## attr(,"conf.level")  
## [1] 0.95
```

This means that the 95% confidence interval for the ratio of the variances of soybean fed to sunflower fed chicks is [0.3622039, 3.9281145].

3. Assume that if the weight of a chick is below 258, that chick is classified under “LOW WEIGHT”. For chicks fed meatmeal vs. those fed soybean,

a) Determine whether there is a significant difference in the proportions of the chicks classified under “LOW WEIGHT”.

b) Construct a 95% confidence interval for the difference in the proportions of the chicks classified under “LOW WEIGHT”.

Discuss the assumption underlying each of the analyses, their validity, and any remedial measures to be taken.

Solution

a) & b). The assumption is that the expected counts in LOW WEIGHT in both `meatmeal` and `soybean` are not too small (generally greater than 5).

```
n=length(meatmeal)+length(soybean) # total number  
low.m=sum(meatmeal<258)  
low.s=sum(soybean<258)  
p=(low.m+low.s)/n  
min(p*length(meatmeal),(1-p)*length(meatmeal),p*length(soybean),(1-p)*length(soybean))
```

```
## [1] 5.28
```

We can see that the expected counts of low weight or not low weight in `meatmeal` and `soybean` are all greater than 5. Then we can use `prop.test` to determine whether there is a significant difference in the proportions of the chicks classified under LOW WEIGHT and get its 95% confidence interval.

```
prop.test(c(low.m,low.s),c(length(meatmeal),length(soybean)))
```

```
##
## 2-sample test for equality of proportions with continuity
## correction
##
## data:  c(low.m, low.s) out of c(length(meatmeal), length(soybean))
## X-squared = 0.968, df = 1, p-value = 0.3252
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.7396  0.1812
## sample estimates:
## prop 1 prop 2
## 0.3636 0.6429
```

```
prop.test(c(low.m,low.s),c(length(meatmeal),length(soybean)))$conf.int
```

```
## [1] -0.7396  0.1812
## attr(,"conf.level")
## [1] 0.95
```

And from the result we can see that the p -value is greater than 0.05, thus we should not reject H_0 . The 95% confidence interval is $[-0.7396, 0.1812]$.