llorban / **stats101**

Branch: **master** **stats101** / live / **12-practice.r** Find file Copy path

llorban Create 12-practice.r bcd0acd 19 hours ago

1 contributor

121 lines (96 sloc) 2.42 KB

```r
x = read.csv(url("http://llorban.net/psyc2300/levente_weight.csv"))

## Practical issues and assumption checking

head(x)
length(x) ## number variables in the data
length(x$Date) ## number of records in the data
hist(x$Weight, main="Histogram of Weight", xlab="Weight", ylab="Frequency")
range = max(as.Date(x$Date))-min(as.Date(x$Date))
mydate = format(as.Date(x$Date),"%Y")
mydata = data.frame(mydate,x$Weight) ## we can do analysis with data

twelve = mydata[mydata$mydate==2012,]    ## "==" check if it's equal
seventeen = mydata[mydata$mydate==2017,]

## Assumptions checking
## voila!
## Normality
par(mfrow=c(1,2))
hist(twelve$x.Weight)
hist(seventeen$x.Weight)

## Homogeneity
## F max test
twelve_var=var(twelve$x.Weight)
seventeen_var = var(seventeen$x.Weight)
Fmax=seventeen_var/twelve_var

## Sample size
## Anything over 30/group is ideal
## Rough guide -- approx 12/group
n17 = length(seventeen$mydate)
n12 = length(twelve$mydate)

## Analysis
## independent samples t-test

## short name
w17 = seventeen$x.Weight
w12 = twelve$x.Weight

## degrees of freedom
df17 = n17-1
df12 = n12-1
df_total = df12+df17

## means and mean difference
```

```r
M17 = sum(w17)/n17
M12 = sum(w12)/n12
mdiff = M17-M12
# doublechecking
mean(w17)
mean(w12)

## centred data
c17=w17-M17
c12=w12-M12
# doublechecking
sum(c17)
sum(c12)

## sums of squares
ss17=sum(c17^2)
ss12=sum(c12^2)

## variance + pooled variance
v17=ss17/df17
v12=ss12/df12
vp=(ss17+ss12)/(df17+df12)
#doublechecking
var(w17)
var(w12)

## note that standard deviation was not necessary
## standard error -based on the pooled variance
se = sqrt(vp/n17 + vp/n12)
se

## execute t-test
## compute t-value
t = mdiff/se
t

## compute p-value
p = 2*pt(t,df_total,lower.tail=FALSE)
p

## state your conclusion
## For reference:
##  H0: M1 = M2; Weight between 2012 and 2017 did not change
##  H1: M1 != M2; There is a significant difference between weights in 2012 and 2017
We reject the null hypothesis.
There is a significant difference between
weights in 2012 and 2017.

#doublecheck
t.test(w17,w12,paired=FALSE,var.equal=TRUE)

## Effect size measures
## Cohen's d
d = mdiff/sqrt(vp)
d
This is a big effect.

## r^2 (variability explained)
r2 = t^2/(t^2+df_total)
r2
```

```
76% of the variability is explained by the year variable

## Confidence Interval
## 95%
criticalt = qt(0.025, df_total, lower=FALSE)

ci95upper = mdiff+(criticalt*se)
ci95lower = mdiff-(criticalt*se)
ci95lower;ci95upper

## Visualize the result
## Bar plot

```