

## HANDOUT IN CLASS 9-5

Researchers are interested if there is different sugar content in children's cereals than in adult cereals. They randomly collected the sugar content (percentage of weight) on 19 children's cereals and 26 adult cereals from a population of 600 children's cereals and 600 adult cereals in order to answer the question of interest.

1. Set up hypotheses to answer the question of interest.
2. What are the two types of tests you have learned so far to test these hypotheses?

Let's examine some plots to decide which type of test to use.

```
children;

## [1] 40.3 55.0 15.7 55.0 50.3 45.9 53.5 43.0 44.2 44.0 47.4 53.0 33.6 27.9
## [15] 48.8 50.4 50.8 56.1 46.6

summary(children)

##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      15.7   43.5   47.4   45.3   51.9   56.1

adult;

## [1] 20.0 30.2  2.2  7.5  4.4 22.2 16.6 14.5 21.4  3.3  6.6  7.8 10.6 16.2
## [15] 14.5  4.1 15.8  3.5  8.5 10.0  4.4  1.3  8.1  4.7 18.4

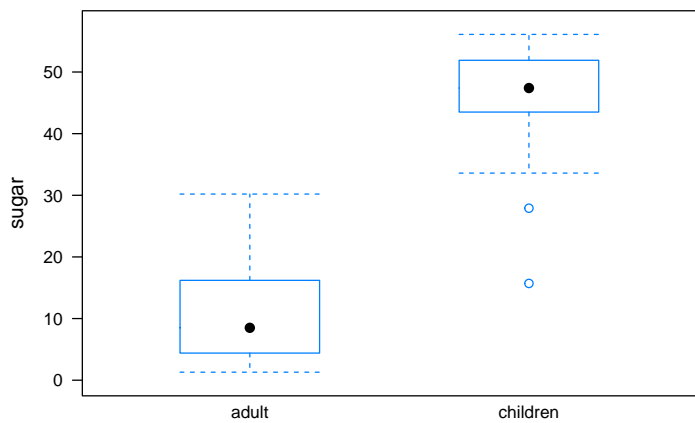
summary(adult);

##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##       1.3    4.4    8.5    11.1    16.2    30.2

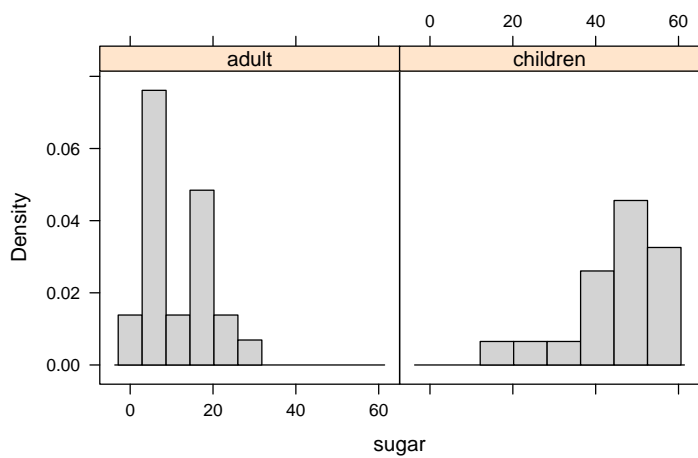
observed2<-compareMean(sugar~type,data=cereal, digits=1)
observed2

## [1] 34.27

bwplot(sugar~type,data=cereal)
```



```
histogram(~sugar|type,data=cereal, col = "lightgray")
```



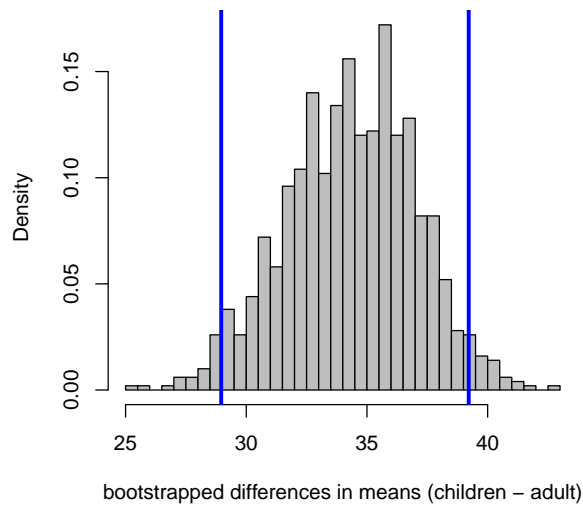
So it appears that the children cereals may be more left-skewed than we would expect from a normal distribution. Which type of test should we use?

Now... think back to Stat 216. Suppose we wanted to answer the question of interest, but we did not want to do a hypothesis test. What is another way to answer the question of interest?

## Bootstrapping

```
require(mosaic)
require(heplots)
set.seed(824862)
bootdist2<-do(1000)*compareMean(sugar~type,data=resample(cereal))
hist(bootdist2$result,col="gray",main="Histogram of bootstrapped Difference in Means", nclass = 50, freq = F, xlab = "bootstrapped Difference in Means")
abline(v=qdata(c(.025,0.975),bootdist2$result)$quantile,col="blue",lwd=3)
```

**Histogram of bootstrapped Difference in Means**



```
#pull off endpoints of the interval for 95% CI
qdata(c(.025,0.975),bootdist2$result)

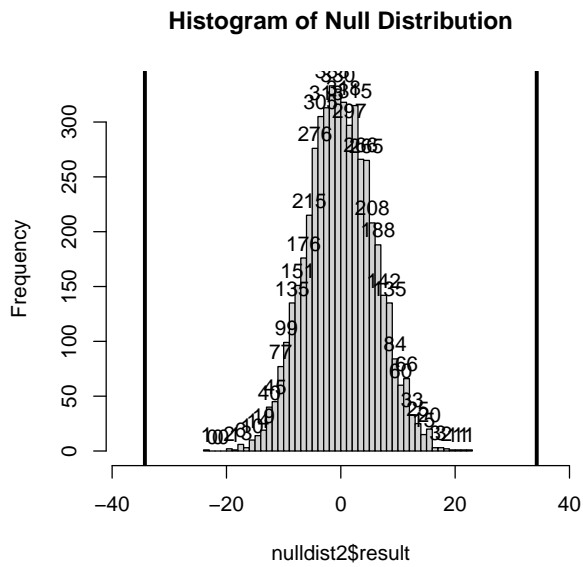
##      quantile      p
## 2.5%      28.96 0.025
## 97.5%      39.22 0.975

#99%CI
qdata(c(.005,0.995),bootdist2$result)

##      quantile      p
## 0.5%      27.42 0.005
## 99.5%      40.92 0.995

## RECALL THE PERMUTATION DISTRIBUTION

nulldist2<-do(5000)*compareMean(sugar~shuffle(type),data=cereal)
hist(nulldist2$result,col="lightgray",main="Histogram of Null Distribution",xlim = c(-38,38), nclass = 50, probability = F, labels = F)
abline(v=observed2,lwd=3)
abline(v=-observed2,lwd=3)
```



```
pdata(abs(observed2),abs(nulldist2$result),lower.tail=F)
```

```
## [1] 0
```