

PH525.1x: Week1

Shu Guo

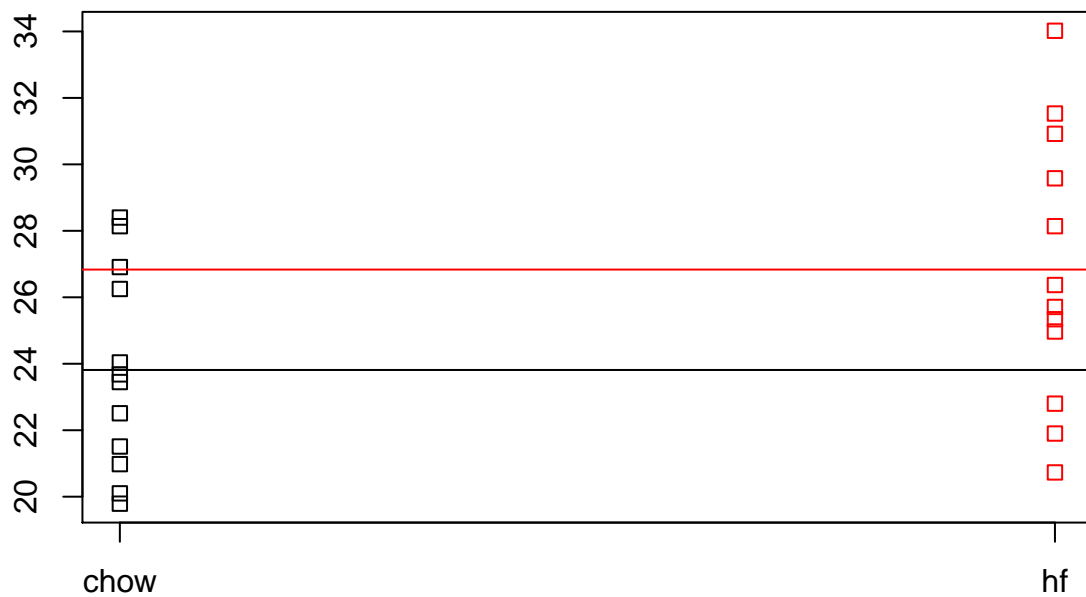
Sunday, February 01, 2015

Introduction to Random Variabbe I

```
## Load the data
dat = read.csv("femaleMiceWeights.csv")
## The observed difference between high fat diet and control was
mean(dat[13:24,2]) - mean(dat[1:12,2])
```

```
## [1] 3.020833
```

```
## A strip chart of the weights of these two groups
s = split(dat[,2], dat[,1])
stripchart(s, vertical=TRUE, col=1:2)
## Add the means to the plot
abline(h=sapply(s, mean), col=1:2)
```



```
## Question 1.1 How many of the high fat mice weigh less than the mean of the control mice (chow)?  
sum(s$hf < mean(s$chow))
```

```
## [1] 3
```

```
## Question 1.2 How many of the control mice weigh more than the mean of the high fat mice?  
sum(s$chow > mean(s$hf))
```

```
## [1] 3
```

```
## Question 1.3 What is the proportion of high fat diet mice over 30?  
sum(s$hf > 30)/length(s$hf)
```

```
## [1] 0.25
```

Introduction to Random Variables II

```
## Course example  
dat[1:12, 2]
```

```
## [1] 21.51 28.14 24.04 23.45 23.68 19.79 28.40 20.98 22.51 20.10 26.91  
## [12] 26.25
```

```
mean(dat[13:24, 2] - mean(dat[1:12, 2]))
```

```
## [1] 3.020833
```

```
population <- read.csv("femaleControlsPopulation.csv")
```

```
n <- 10000  
null <- vector("numeric", n)  
for (i in 1:n){  
  control <- sample(population[, 1], 12)  
  treatment <- sample(population[, 1], 12)  
  null[i] <- mean(treatment) - mean(control)  
}
```

```
diff <- mean(dat[13:24, 2]) - mean(dat[1:12, 2])  
#what percent are bigger than `diff`?  
mean(null > diff)
```

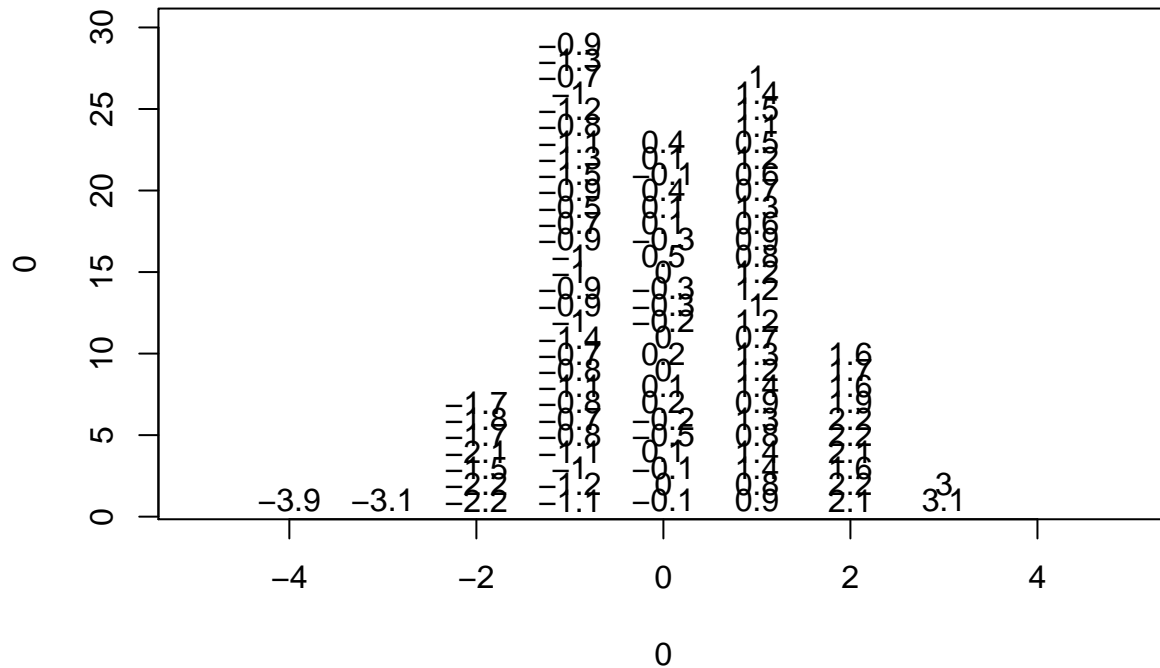
```
## [1] 0.0141
```

```
## Illustration of the null distribution Repeat the loop above but this time add a point to the figure every  
time re-run the experiment
```

```

# Read the population
n <- 100
plot(0, 0, xlim = c(-5, 5), ylim = c(1, 30), type = "n")
totals <- vector("numeric", 11)
for (i in 1:n){
  control <- sample(population[, 1], 12)
  treatment <- sample(population[, 1], 12)
  nulldiff <- mean(treatment) - mean(control)
  j <- pmax(pmin(round(nulldiff) + 6, 11), 1)
  totals[j] <- totals[j] + 1
  text(j - 6, totals[j], pch = 15, round(nulldiff, 1))
  ##if(i < 15) scan() ## add this line to interactively see values appear
}

```



Recreate the vector of differences between means of random samples from the control population.

```

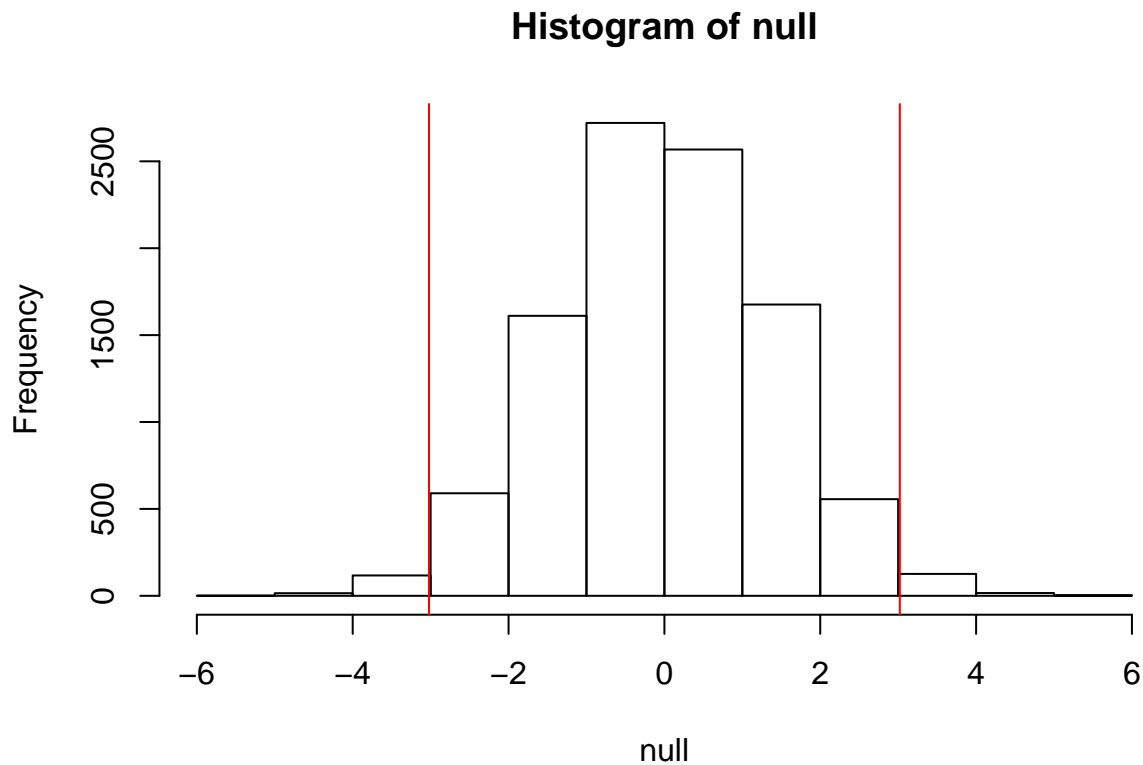
null <- replicate(10000, mean(sample(population[, 1], 12)) -
                          mean(sample(population[, 1], 12)))

# Make a histogram
hist(null)

# The original difference observed between the mice fed high fat diets and control mice:
diff = mean(dat[13:24,2]) - mean(dat[1:12,2])
# Add this difference to the histogram:
abline(v=diff, col="red")
# Also add the negative of the difference:

```

```
abline(v=-diff, col="red")
```



If we look for the number of null distribution values to the right of the (right) red line, we would say “we calculated the probability of observing a larger difference from the null distribution”. This is sometimes called a “one-tailed” probability, because we only look at one “tail” of the histogram (the left and right sides where the bars become short).

By looking at the tails on both sides of the histogram, we can say “we calculated the probability of observing as extreme a difference from the null distribution”. This is sometimes called a “two-tailed” probability. This probability is commonly referred to as a p-value.

Question 3.1: What is the one-tailed probability of seeing as big a difference as we observed, calculated from your null distribution? (0.0141)

Question 3.2: What is the two-tailed probability of seeing as big a difference as we observed, calculated from your null distribution? (0.0266)