Functions in Action for Probability!

Is a mean different from 0?

Recall last week that we calculated p-values assuming we knew a population's standard deviation.

Often we want to know if a sample mean is different from 0. We know that an estimated mean from a large sample size is normally distributed, so...

Enter the Z-Test

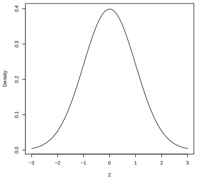
$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}}$$

The Z-score compares a sample mean to an assumed population mean.

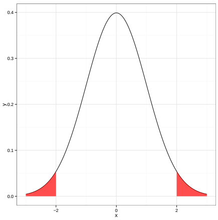
We call it a Z-Score because we correct by the standard error of the mean to compare to a standard deviation of the sample mean (SE).

Standard Normal (Z) Distribution

$$\text{Z-score} = \frac{Y_i - \bar{Y}}{\sigma}$$



Z-Test



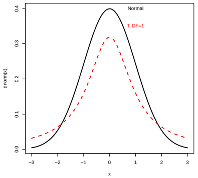
Does Z fall into these tails?

A Simple P(z) Function

```
pz <- function(sample, mu = 0) {
    z <- (mean(sample) - mu)/(sd(sample)/sqrt(length(sample)))

    return(2 * pnorm(abs(z), lower.tail = F))
}
```

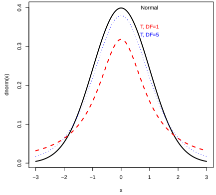
```
set.seed(697)
pz(rnorm(5000))

## [1] 0.6087

pz(rnorm(5, mean = 1))

## [1] 0.0703
```

Sample's Aren't Normal: Student's T



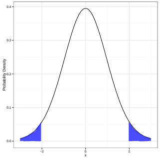
DF = Degrees of Freedom
= Sample Size - # of Estiamted Parameters
For t, this is n-1

Sample's Aren't Normal: Student's T



Exercise: Gimme a T (function)

$$T = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}}$$



Write a oneSampleT function. To make it more readable, write a SE function, too.
Challenge: make it return a list with additional information of your choice.

Exercise: Gimme a T (function)

```
# SE first
se <- function(sample) sd(sample)/(sqrt(length(sample)))

# now T
oneSampleT <- function(sample, mu = 0) {
    t <- (mean(sample) - mu)/se(sample)

    2 * pt(abs(t), df = length(sample) - 1, lower.tail = F)
}
```

Exercise: Gimme a T (function)

```
samp <- rnorm(50, 1)
t.test(samp)

##
##  One Sample t-test
##
## data:  samp
## t = 9.405, df = 49, p-value = 1.48e-12
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.9509 1.4676
## sample estimates:
## mean of x
##     1.209

oneSampleT(samp)

## [1] 1.48e-12
```

Consider the χ^2

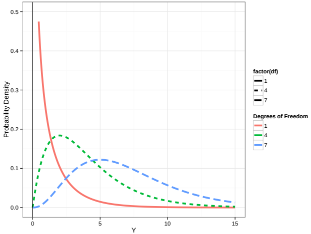
If the difference between an observation and it's expectation follows a Z-Distribution, the square should follow a chi-square distribution!

Think of a normal distribution squared yielding a χ^2 distribution defined by degrees of freetom = n-1

$$\chi^2 = \frac{\sum (O - E)^2}{E}$$

Consider the χ^2

$$\chi^2 = \frac{\sum (O - E)^2}{E}$$



Assumptions of the χ^2

1. No expected values less that 1
2. 80% of the expected values must be >5

What happens if I violate the assumptions? Combine categories or use a different test.

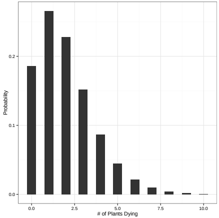
Example of the χ^2

The number of tomato plants that failed due to disease in a farm should follow a negative binomial distribution in any given year with 2 plants dying and an overdispersion parameter of 5. Famer Dale seeded tomato plants in 100 different plots. Looking across his field, he finds the following pattern of mortality.
0 plants died: 9 plots
1 plants died: 23 plots
2 plants died: 22 plots
3 plants died: 18 plots
4 plants died: 19 plots
>4 plants died: 9 plots

Was this an anomolously bad year?

What are our Expectations?

Negative Binomial with a μ of 2 and dispersion parameter of 5.



Example of the χ^2

```
100*dnbinom(0:4, mu=2, size=5)

## [1] 18.593 26.562 22.767 15.178  8.673

100*(1 - sum(dnbinom(0:4, mu=2, size=5)))

## [1] 8.225
```

0 plants died: 9 plots, 18.59 expected
1 plants died: 23 plots, 26,56 expected
2 plants died: 22 plots, 22.77 expected
3 plants died: 18 plots, 15.17 expected
4 plants died: 19 plots, 8.67 expected
$>$4 plants died: 9 plots, 8.23 expected

Example of the χ^2

```
observed <- c(9,23,22,18,19,9)

expected<- c(100*dnbinom(0:4, mu=2, size=5),
             100*(1-sum(dnbinom(0:4, mu=2, size=5))))

chisq<-sum((observed-expected)^2/expected)

chisq

## [1] 18.35


pchisq(chisq, df=6-1, lower.tail=FALSE)

## [1] 0.002543
```

What if we ONLY new the distribution and dispersion parameter?