Day 3

Lucas Anderton

1/4/2020

Statistics with R

The Root-Mean-Square (RMS)

RMS describes average *magnitude* of variable's values. The RMS takes each values,

1. squares it,
2. takes the mean of these squares, and then
3. takes the square root.

Why take the square, then square root?

Standard Deviation (SD)

SD describes the spread of a variable. SD: the RMS of the deviations from the average. Tells us "How far from the average is a typical value of the variable?" Calculate by taking each observation's difference from the average, then take the RMS of those differences. Use `sd` to calculate in R.

Variance (SD^2)

Variance: SD squared. Variance: average of the squared deviations from the mean. Use `var` to calculate it in R.

The z-score

For variable X, the z-score of observation xi, tells how far it is from average, in units of the standard deviation.

Inteperetation: z-score does not depend on units we measure in (as long as linnear transformation). So, for example, we could compare currencies.

Practice with two respondents.

```
resp1_inc <- (65000 - 50000) / 15000
resp1_inc
```

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

```r
resp2_inc <- (20000 - 50000) / 15000
resp2_inc
```

```
## [1] -2
```

QQ plot

Visually compare the quantiles of two distributions.

Probability

Foundations:

1. **Experiment**: a process that yields a probailistic/stochastic outcome. Not nec. entirely random, but with soem random component.
2. **Outcome space/Sample space**: The set of all possible outcomes of an experiment. Usually denoted by Omega Ω.
3. **Event**: A subset of Omega. Usually denoted A, B, etc. The probability of A happening is P(A).
4. **Complement**:

Examples:

1. Experiment: a voter will vote Dem, vote Rep, vot other, or abstain.
2. Outcome space: Omega Ω = {Dem, Rep, other, abstainn}
3. Event:
   - A = abstains. Assuming all equally likely, what is P(A)?
   - B = supports a major party candidate. Assuminng all equally likely, what is P(B)?
4. Complement:
   - A^C is the probability they vote for a Democrat, Republican, or third party candidate, so 75%.
   - B^C is the probability they don't vote or don't vote for a major party.

Probability of either of 2 events

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Law of total probability

We can decompose probability of A into two components: A happening when B also happens, and A happening when "not B" happens.

Combinations: Counting selected sets

How many ways to **select** k things from a set of n things?

Indpendence

Events are **independent** if they don't provide any innformation about each other. Knnowinng A happened doesn't change the probability of B happening. Know B happening doesn't change the probability of A happening.

If A and B are independent, then:
P(A and B) = P(A)P(B|A)
= P(A)P(B)

A probability example

For a given Conngress,

1. Take a random sample of 10 bills.
2. Calculate p, the prop in your sample that passed.
3. Calculate the Standard Error around the p
4. Find the critical value for an 80% interval
5. Calculate an 80% confidence innterval aroundn p

```
samp <- rbinom(10, 1, .35)
samp
```

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

```
phat <- mean(samp)
se <- sqrt(phat * (1 - phat) / 10)
phat
```

```
## [1] 0.6
```

```
se
```

```
## [1] 0.1549193
```

```
critval <- qnorm(.9)
lower <- phat - critval * se
upper <- phat + critval * se
lower
```

```
## [1] 0.4014629
```

```
upper
```

```
## [1] 0.7985371
```