binomial_test

October 29, 2020

The Binomial Test and Categorical Data

1 1. Categorical Data

- Variables representing group members
- Examples:
    - Political party affiliation
    - City of origin
    - Gender
    - Ethnicity

2 2. The Binomial Test

- Probably the most basic example of a hypothesis tests (and very useful)
- Used to compare distribution of observations in two categories against theoretical distribution
- Essentially, we use the binomial test when we have a problem that can be expressed in terms of "successes" and "failures"

2.1 2.1 Binomial Test Examples

Example questions we can answer: - Given N tosses of a coin, $X_1, X_2, ..., X_n$, where $X_i = 1$ denotes heads and $X_i = 0$ is tails, is this a fair coin? - Given the counts of females and males in a particular class, are there significantly more females than males? - Suppose we are doing quality control on a medical device known to have a 0.001% failure rate. Given the number of failures in a specific batch and the batch size, does this batch have significantly more failures than we expect?

2.1.1 2.1.1 Review Binomial Distribution

1. Discrete probability distribution
2. Has two parameter

- n: number of "trials"
- p: probability of "success" for a given trial

[1.] Image source: wikipedia.org

2.2 2.2 Binomial Test

2.3 2.3 Binomial Test: Coin Toss Example

Suppose we have the following data after tossing a coin several times:

[H, T, T, T, H, H, T, H, T, T, H, T, T, T, T]

Is this a fair coin?

2.3.1 2.3.1 Data Generation

In [2]:
```
# create variable to store data
coin_tosses <- c("H", "T", "T", "T", "H", "H", "T", "H", "T", "T", "H",
                 "T", "T", "T", "T")

# get number of tosses
n_tosses <- length(coin_tosses)

# get number of heads
n_heads <- sum(coin_tosses == "H")

# print variables we created to check sanity
print(n_tosses)
print(n_heads)
```

```
[1] 15
[1] 5
```

2.3.2 2.3.2 Using `binom.test()`

In [4]:
```
# run binomial test on coin toss data

bin_test1 <- binom.test(n_heads, n_tosses)

print(bin_test1)
```

```

	Exact binomial test

data:  n_heads and n_tosses
number of successes = 5, number of trials = 15, p-value = 0.3018
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.1182411 0.6161963
sample estimates:
probability of success
             0.3333333
```

2.4 2.4 Binomial Test: Device Defects Examples

Suppose we are doing quality control for a medical device known to have a 0.0001% failure rate. We are given a batch of 250000 to be tested. Of these, we find 17 defective devices. Does this batch have a significantly higher failure rate than our known failure rate?

```
In [8]: # specify our inputs

        p_failure <- 0.0001      # a-priori known failure rate

        n_trials <- 250000        # number of devices produced

        n_defectives <- 17        # number of defective devices
```

2.4.1 2.4.1 Device Defects Example (cont.)

```
In [9]: # run binomial test on medical device data

        test2 <- binom.test(n_defectives, n_trials, p = p_failure, alternative = "greater")

        print(test2)


        Exact binomial test

data:  n_defectives and n_trials
number of successes = 17, number of trials = 250000, p-value =
1.557e-09
alternative hypothesis: true probability of success is greater than 1e-05
95 percent confidence interval:
 4.332901e-05 1.000000e+00
sample estimates:
probability of success
               6.8e-05
```