Independent-sample t-test

A/B TESTING IN R



Lauryn Burleigh
Data Scientist

datacamp

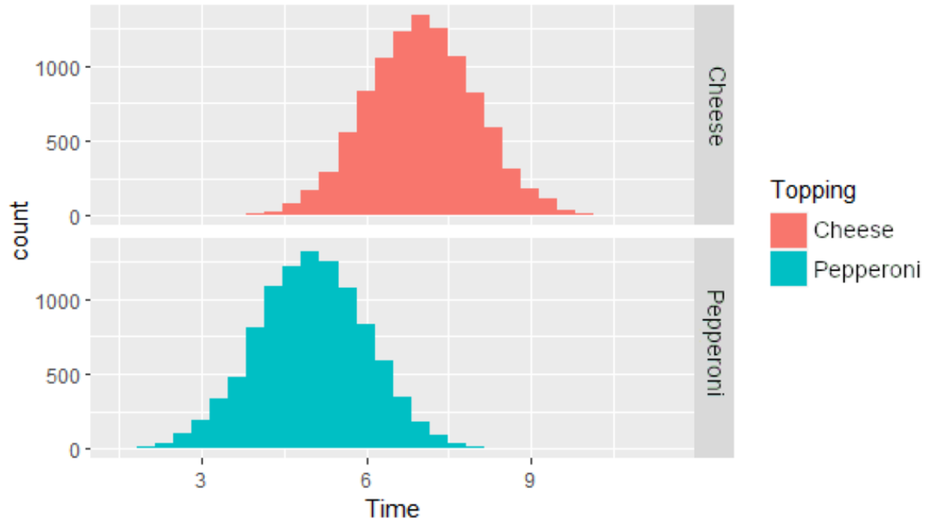
Independent t-test in A/B design

- Significance in difference of means
- Null hypothesis:
  - Cheese and Pepperoni pizza are eaten in the same amount of time (no difference)

Assumptions

- Dependent variable:
  - Interval or ratio
  - Equal intervals
- Random samples
- Normal distribution
- Similar group variances

```
library(ggplot2)
ggplot(pizza, aes(x = Time,
                  fill = Topping)) +
       geom_histogram() +
       facet_grid(Topping~.)
```



Sample size

```
library(pwr)
pwr.t.test(d = 0.73, power = 0.80,
           sig.level = 0.05,
           type = "two.sample",
           alternative = "two.sided")
```

```
     Two-sample t test power calculation

              n = 30.44799
              d = 0.73
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Assessing variances

- Group variances equal
- Levene's test

Not significant (p > 0.05) = variances equal

If significant (p < 0.05) = variances not equal

```
library(car)
leveneTest(Time ~ Topping,
           data = Pizza)
```

```
Levene's Test for Homogeneity of Variance
          Df F value Pr(>F)
group   1  0.1457 0.7031
```

Test

```
t.test(Time ~ Topping, data = Pizza,
       paired = FALSE,
       alternative = "two.sided",
       var.equal = TRUE)
```

```
    Two Sample t-test
data:  Time by Topping
t = 2.3811, df = 198, p-value = 0.01821
alternative hypothesis: true difference
in means between group Pepperoni and
group Cheese is not equal to 0
95 percent confidence interval:
 0.0599370 0.6377601
```

Cohen's d

- Cohen's d: t-test effect size
  - Standardized measure of differences between means

- Small: 0.2
- Medium: 0.5
- Large: 0.8

```
library(effectsize)
cohens_d(Time ~ Topping, data = Pizza)
```

```
Cohen's d |        95% CI
<-------------------------
0.34      | [0.06, 0.62]
```

Power

```
library(pwr)

pwr.t.test(n = 1000,
           sig.level = 0.0182,
           d = 0.34,
           type = "two.sample")
```

```
  Two-sample t test power calculation
                n = 100
                d = 0.34
        sig.level = 0.0182
            power = 0.510256
      alternative = two.sided
NOTE: n is number in *each* group
```

- Ideal power to accept results: 0.8
  - Probability of error: 20%
  - 100 - 80 = 20

Let's practice!

A/B TESTING IN R

Mann-Whitney U test

A/B TESTING IN R

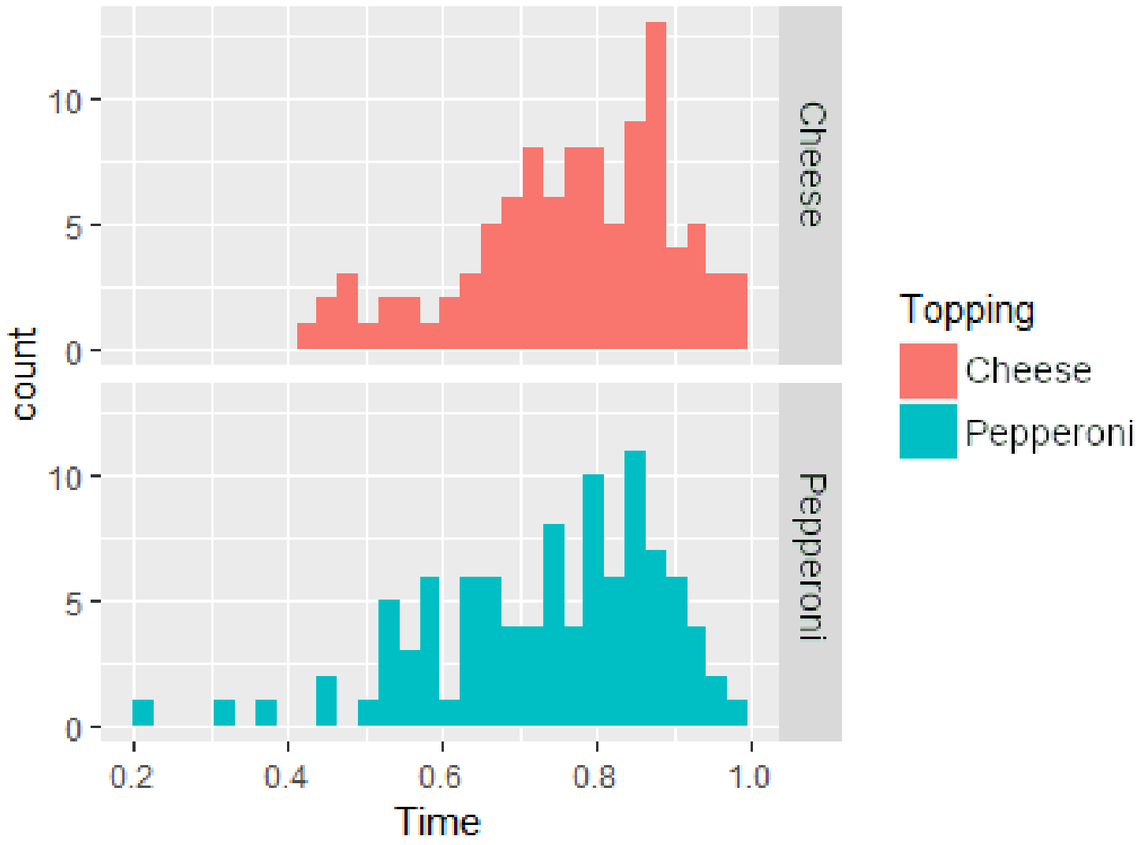


Lauryn Burleigh
Data Scientist

datacamp

Mann-Whitney U

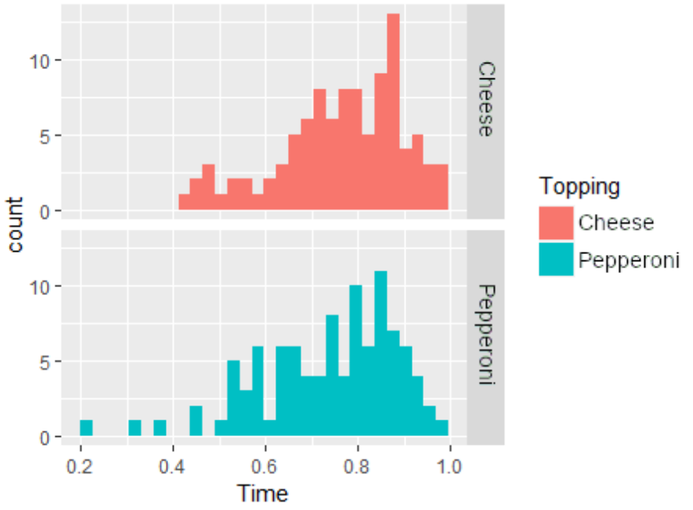
- Time to eat Cheese versus Pepperoni pizza
- Not normal distribution
- Non-parametric
  - Distribution shape not assumed
  - Mann-Whitney U test



Assumptions

- Same distribution shapes
- Assesses difference in medians
- Normal distribution: mean = median
- Non-normal distribution: median more appropriate
- Extra assumptions: more powerful test
- Null hypothesis: no difference in the median time to eat cheese and pepperoni pizza

```
library(ggplot2)
ggplot(pizza, aes(x = Time,
                  fill = Topping)) +
       geom_histogram() +
       facet_grid(Topping~.)
```



Sample Size

```
library(pwr)
pwr.2p2n.test(h = 0.40,
              sig.level = 0.05,
              power = 0.8, n1 = 100)
```

```
    difference of proportion power
 calculation for binomial distribution
              h = 0.4
             n1 = 100
             n2 = 96.29156
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
NOTE: different sample sizes
```

```
pwr.2p2n.test(h = 0.40,
              sig.level = 0.05,
              power = 0.8, n1 = 110)
```

```
    difference of proportion power
 calculation for binomial distribution
              h = 0.4
             n1 = 110
             n2 = 88.54092
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
NOTE: different sample sizes
```

- Expected effect size `h` : rank-biserial correlation r
  - See how subjects in groups rank

Test

```
wilcox.test(Time ~ Topping,
            data = Pizza)
```

- y ~ x
  - y: data
  - x: group

```
    Wilcoxon rank sum test with
    continuity correction
data:  Enjoyment by Topping
W = 6051, p-value = 0.01026
alternative hypothesis: true location
shift is not equal to 0
```

Effect size and power

Effect size

```
library(effectsize)
rank_biserial(Time ~ Topping,
              data = pizza)
```

```
r (rank biserial) |         95% CI
<---------------------------------
0.21              | [0.05, 0.36]
```

- Small: 0.1
- Medium: 0.3
- Large: 0.5

1 - 0.14 = 0.86 probability of Type II error

Power Analysis

```
library(pwr)
pwr.2p2n.test(h = 0.21, sig.level = 0.01,
              n1 = 100, n2 = 100)
```

```
     difference of proportion power calculation for binomi

              h = 0.21
             n1 = 100
             n2 = 100
      sig.level = 0.01
          power = 0.1376818
    alternative = two.sided

NOTE: different sample sizes
```

Let's practice!

A/B TESTING IN R

Chi-Square test of independence

A/B TESTING IN R



Lauryn Burleigh
Data Scientist

datacamp

Frequency tables

- Whether nominal variables are related
  - Eating the pizza topping again
- Frequency table - number of times each of the named conditions co-occurs
  - One-topping pizza: variable 1
  - Would eat again: variable 2

```
table(Pizza$Topping, Pizza$EatAgain)
```

```
             Maybe  No Yes
  Cheese       207  90 703
  Pepperoni    147 101 752
```

Hypotheses

- Null hypothesis: Topping and interest in eating again have no relationship
- Alternative hypothesis: Topping and interest in eating again have a relationship

Cannot conclude cause



Sample size

```
library(pwr)
pwr.chisq.test(w = 0.1, df = 2,
               power = 0.80,
               sig.level = 0.05)
```

```
     Chi squared power calculation
              w = 0.1
              N = 963.4689
             df = 2
      sig.level = 0.05
          power = 0.8
NOTE: N is the number of observations
```

- w:
  - Small: 0.10
  - Medium: 0.30
  - Large: 0.50
- Degrees of freedom:
  - (Conditions-of-variable-1 - 1) * (Conditions-of-variable-2 - 1)
  - (2-1) * (3-1) = 2

Minimum of five values at each variable combination

Chi-square test

```
freqtbl <- table(Pizza$Topping,
                 Pizza$EatAgain)
chisq.test(freqtbl)
```

```
    Pearson's Chi-squared test


data:  freqtbl
X-squared = 12.453, df = 2,
p-value = 0.001976
```

Large sample sizes: over 1000

Effect size

- Effect size: w
- Probabilities

```
probtbl <- freqtbl/sum(freqtbl)


library(pwr)
ES.w2(probtbl)
```

- Small: 0.10
- Medium: 0.30
- Large: 0.50

```
[1] 0.07890872
```

Power

```
pwr.chisq.test(w = 0.079, df = 2,
               N = 1000,
               sig.level = 0.002)
```

```
     Chi squared power calculation

              w = 0.079
              N = 1000
             df = 2
      sig.level = 0.002
          power = 1

NOTE: N is the number of observations
```

Let's practice!

A/B TESTING IN R

Fisher's Exact test

A/B TESTING IN R



Lauryn Burleigh
Data Scientist

Specifications

- Non-parametric of categorical data
  - Whether Cheese or Pepperoni pizza eaten again
- Sample < 1000
- Frequency < 5 for at least 20%
- Use Fisher's Exact test for exact p-value
- 2×2 frequency table

Hypergeometric Distribution

- Likelihood of frequency table

```
table(Pizza$Topping, Pizza$EatAgain)
```

```
           No Yes
Cheese     140 360
Pepperoni   80 420
```

Hypotheses

- Null hypothesis: No relationship between eating pizza again and topping group



Sample

```
library(pwr)
pwr.chisq.test(w = 0.1, df = 2,
               power = 0.80,
               sig.level = 0.05)
```

```
     Chi squared power calculation
              w = 0.1
              N = 963.4689
             df = 2
      sig.level = 0.05
          power = 0.8
NOTE: N is the number of observations
```

```
table(Pizza$Topping, Pizza$EatAgain)
```

```
                No Yes
  Cheese       140 360
  Pepperoni     80 420
```

Test

```
freqtbl <- table(Pizza$Topping,
                 Pizza$Time)
fisher.test(freqtbl)
```

```
    Fisher's Exact Test for Count Data
data:  freqtbl
p-value = 0.001809
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.4569013 0.8417208
sample estimates:
odds ratio
 0.6209987
```

Effect size

```
freqtbl <- table(Pizza$Topping,
                 Pizza$Time)
fisher.test(freqtbl)
```

- Small: 1.5
- Medium: 2.5
- Large: 4
- Pepperoni pizza x 0.62 than Cheese pizza

```
Fisher's Exact Test for Count Data
data:  freqtbl
p-value = 0.001809
alternative hypothesis: true odds
ratio is not equal to 1
95 percent confidence interval:
 0.4569013 0.8417208
sample estimates:
odds ratio
 0.6209987
```

Power

```
library(pwr)
pwr.2p.test(h = 0.62, n = 500,
            sig.level = 0.002)
```

```
    Difference of proportion power
 calculation for binomial distribution

              h = 0.62
              n = 500
      sig.level = 0.002
          power = 1
    alternative = two.sided

NOTE: same sample sizes
```

Let's practice!

A/B TESTING IN R