Introduction to R for Epidemiologists

Jenna Krall, PhD

Thursday, Feburary 12, 2015

Outline

1. One sample T-tests
2. Two sample T-tests
3. Tests of proportion
4. Chi-squared tests
5. Relative risk
6. Odds ratio

One sample T tests in R

Review

- One sample Z and T tests are used for determining whether the mean in a population is different than a hypothesized value
- Examples
  - Is the average concentration of particulate matter air pollution in Atlanta different than 12 μ g/m^3?
  - Is the average gestational age for infants born with very low birthweight less than 39 weeks?

Assumptions for Z and T tests

- Large sample size or data are approximately normal if sample size is small

Assumptions for Z test

- Population standard deviation is known

One sample T-tests in R

Is average gestational age in the population different than 39 weeks (use α=0.05)?

- Null hypothesis $H_0: \mu = 39$
- Alternative hypothesis $H_1: \mu \neq 39$

One sample T-tests in R

```r
t_age <- t.test(x = vlbw$gest, mu = 39)
t_age
```

```
##
##  One Sample t-test
##
## data:  vlbw$gest
## t = -54.2729, df = 173, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 39
## 95 percent confidence interval:
##  28.92818 29.63504
## sample estimates:
## mean of x
##  29.28161
```

We reject the null hypothesis that the average gestational age of infants born with very low birthweight is significantly different than 39 weeks at α=0.05.

Two sample T-tests

Unpaired two sample t-tests

Recall that a two sample t-test tests the hypothesis that the means in two populations are the same:

- Is the average concentration of particulate matter air pollution in Atlanta different than the average air pollution concentration in Birmingham?
- Does the average gestational age of infants born with very low birthweight differ between males and females?

So we are testing whether the means of a continuous variable differ between two groups:

- Null hypothesis $H_0: \ \mu_1 = \mu_2$
- Alternative hypothesis $H_1: \ \mu_1 \neq \mu_2$

Two sample T-tests

Paired two sample t-tests

- If the data are paired, use paired tests
  - e.g. Is the mean BMI the same after enrollment in an exercise program?
  - Paired tests account for the fact that we expect pairs to be more similar than we would expect if the data were unpaired.

Two sample T-tests

Does mean gestational age differ between male and female low birthweight infants?

```
age_female <- vlbw$gest[vlbw$sex == "female"]
age_male <- vlbw$gest[vlbw$sex == "male"]
t.test(age_female, age_male, alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  age_female and age_male
## t = -0.2063, df = 170.313, p-value = 0.4184
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf 0.5191429
## sample estimates:
## mean of x mean of y
##  29.24419  29.31818
```

Tests of proportion

We can also test proportions in R.

Is the proportion of those with pneumothorax different than 6.3%?

- One sample test of proportion
  - Null hypothesis $H_0: \; p_1 = 0.063$
  - Alternative hypothesis $H_1: \; p_1 \neq 0.063$

Is the proportion of those with pneumothorax different between multiple and singleton births?

- Two sample test of proportion
  - Null hypothesis $H_0: \; p_1 = p_2$
  - Alternative hypothesis $H_1: \; p_1 \neq p_2$

Tests of proportion

Is the proportion of pneumothorax different than 6.3%?

```
table_pneumo <- table(vlbw$pneumo)
table_pneumo
```

```
##
##   0   1
## 151  23
```

Tests of proportion

Is the proportion of pneumothorax different than 6.3%?

```
table(vlbw$pneumo)
```

```
##
##   0   1
## 151  23
```

```
table_pneumo <- matrix(c(23, 151), ncol = 2)
prop.test(table_pneumo, p = 0.063)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  table_pneumo, null probability 0.063
## X-squared = 12.9608, df = 1, p-value = 0.0003181
## alternative hypothesis: true p is not equal to 0.063
## 95 percent confidence interval:
##  0.08735651 0.19378764
## sample estimates:
##         p
## 0.1321839
```

Tests of proportion

Is the proportion of pneumothorax different between multiple and singleton births?

```r
table(twin = vlbw$twn, pneumo = vlbw$pneumo)
```

```
##      pneumo
## twin    0   1
##     0 115  17
##     1  36   6
```

```r
table_pneumo <- matrix(c(17, 6, 115, 36), ncol = 2)
colnames(table_pneumo) <- c("Pneumo", "No pneumo")
rownames(table_pneumo) <- c("Not twin", "Twin")
```

Tests of proportion

Is the proportion of pneumothorax different between multiple and singleton births?

```r
prop.test(table_pneumo)
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  table_pneumo
## X-squared = 0, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.1484085  0.1202700
## sample estimates:
##    prop 1    prop 2
## 0.1287879 0.1428571
```

Chi-squared test

Are two cateogorical variables independent?

- Is HIV infection associated with MRSA infection?
- Is sex associated with being a twin in very low birthweight infants?

Hypothesis:

- Null hypothesis: Sex is independent of being a twin
- Alternative hypothesis: Sex is not independent of being a twin

Assumptions:

- If 2x2 table, no cell counts < 5
- If rxc table, no more than 20% cells < 5

Chi-squared test

```
chsq_surgery <- chisq.test(vlbw$sex, vlbw$twn)
chsq_surgery
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  vlbw$sex and vlbw$twn
## X-squared = 0.3807, df = 1, p-value = 0.5372
```

```
names(chsq_surgery)
```

```
## [1] "statistic" "parameter" "p.value"   "method"    "data.name" "observed"
## [7] "expected"  "residuals" "stdres"
```

```
chsq_surgery$p.value
```

```
## [1] 0.5372067
```

Relative risk and odds ratio

Relative risk (RR)

- Ratio of risks: p_1/p_2
- Is the risk of disease the same in the exposed and unexposed groups?
- Often interested in testing $H_0: \; RR = 1$ vs. $H_1: \; RR \neq 1$
- Can only be calculated in prospective studies

Relative risk and odds ratio

Odds ratio (OR)

- Ratio of odds
- Is the odds of disease the same in the exposed and unexposed groups?
- Odds is **NOT** the same as risk
- Odds: p/(1-p) or p/q
- OR $= (p_1/(1 - p_1)) / (p_2/(1 - p_2)) = (p_1/q_1) / (p_2 / q_2)$
- Often interested in testing $H_0: \; OR = 1$ vs. $H_1: \; OR \neq 1$
- Useful in retrospective studies

Relative risk and odds ratio

Remember: Reference groups are first row and first column

- We need to reverse the columns using the rev argument
- We want to compare the odds of pneumothorax in twins compared to not twins
- If we don't reverse the columns, we are comparing the odds of not having pneumothorax in twins vs. not twins

Relative risk and odds ratio

```r
library(epitools)
epitab(table_pneumo, method = "oddsratio", rev = "columns")
```

```
## $tab
##          No pneumo        p0 Pneumo        p1 oddsratio    lower    upper
## Not twin       115 0.7615894     17 0.7391304  1.000000       NA       NA
## Twin            36 0.2384106      6 0.2608696  1.127451 0.413459 3.074418
##            p.value
## Not twin        NA
## Twin     0.7972418
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "fisher.exact"
```

Relative risk and odds ratio

Remember: Reference groups are first row and first column

- We need to reverse the columns using the rev argument

```
epi_pneumo <- epitab(table_pneumo, method = "riskratio", rev = "columns")
epi_pneumo
```

```
## $tab
##            No pneumo        p0 Pneumo        p1 riskratio     lower    upper
## Not twin         115 0.8712121     17 0.1287879  1.000000        NA       NA
## Twin              36 0.8571429      6 0.1428571  1.109244 0.4677461 2.630533
##              p.value
## Not twin          NA
## Twin       0.7972418
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "fisher.exact"
```

Relative risk and odds ratio

```
names(epi_pneumo)
```

```
## [1] "tab"          "measure"    "conf.level" "pvalue"
```

```
epi_pneumo_out <- epi_pneumo$tab
colnames(epi_pneumo_out)
```

```
## [1] "No pneumo" "p0"          "Pneumo"     "p1"          "riskratio" "lower"
## [7] "upper"      "p.value"
```

Sample size calculations in R

How many observations would we need to test whether two means are different if

- The difference in means is 0.1
- The standard deviation is 1
- We want 90% power

```
power.t.test(delta = 0.1, power = 0.9, type = "two.sample",
  alternative = "two.sided")
```

```
##
##      Two-sample t test power calculation
##
##               n = 2102.445
##           delta = 0.1
##              sd = 1
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```