

# NOMINAL TESTS



UNIVERSITÄT  
LEIPZIG

Erik Kusch

[erik.kusch@uni-leipzig.de](mailto:erik.kusch@uni-leipzig.de)

Behavioural Ecology Research Group  
University of Leipzig

??/??/20??

## 1 Background

## 2 Analyses

- Binomial Test
- McNemar
- Cochran's Q
- Chi-Squared

## 3 Our Data

- Choice Of Variables
- Research Questions

# Introduction

These approaches only allow for the use of **categorical** (nominal) **variables**!

Prominent nominal tests include:

- **Binomial Test**
- **McNemar**
- Fisher's Exact
- **Cochran's Q**
- **Chi-Squared**
- ...

# The `table()` function

In reality, you will need to convert your data to fit the various nominal test specifications. To do so, you may wish to enlist the help of the `table()` function of base R which converts nominal records into count data.

```
Samples <- c("A", "B")
set.seed(42)
counts <- sample(Samples, size = 1000, replace = TRUE)
table(counts)
```

```
## counts
##      A      B
## 527 473
```

# Purpose And Assumptions

## Binomial Test

`binom.test()` in base R

*Purpose:*

To test whether the observed distribution of data values of a binomial variable differ from what was expected.

$H_0$

*The observed binomial data proportions do not differ significantly from the expected proportions.*

*Assumptions:*

- Variable values are binomial.
- The population is significantly larger than the sample.
- The sample accurately represents the population.
- Sampled values are independent (one value does not influence another).

# Minimal Working Example

We feed the `binom.test()` function a 1/4 (`c(200, 800)`) data set whilst expecting the distribution to be 1 to 1 ( $p = 0.5$ ).

```
binom.test(c(200, 800), p = 0.5)
```

```
##  
## Exact binomial test  
##  
## data: c(200, 800)  
## number of successes = 200, number of trials = 1000,  
## p-value <2e-16  
## alternative hypothesis: true probability of success is not equal to 0.5  
## 95 percent confidence interval:  
##  0.18 0.23  
## sample estimates:  
## probability of success  
##                0.2
```

The result is significant ( $p \approx 0$ ).

# Purpose And Assumptions

## McNemar

`mcnemar.test()` in base R

*Purpose:* To test whether there is a change in proportion of paired data.

$H_0$  *The observed binomial data proportions do not differ significantly between treatments/paired sets.*

*Assumptions:*

- Variable values are binomial.
- The population is significantly larger than the sample.
- The sample accurately represents the population.

# Minimal Working Example

We feed the `mcnemar.test()` function a 1 to 1 (`c(500, 500)`) as well as a 1 to 4 (`c(200, 800)`) data set for the paired data sets.

```
Performance <- matrix(c(500, 500, 200, 800), nrow = 2)
Performance
```

```
##           [,1] [,2]
## [1,]      500  200
## [2,]      500  800
```

```
mcnemar.test(Performance)
```

```
##
## McNemar's Chi-squared test with continuity
## correction
##
## data: Performance
## McNemar's chi-squared = 100, df = 1, p-value <2e-16
```

With a p-value of  $\approx 0$  the test concludes significantly.



# Purpose And Assumptions

## Cochran's Q

`cochran.q()` in the `nonpar` package

*Purpose:*

To test whether there are differences in matched sets of three or more frequencies or proportions.

$H_0$

*The observed proportions of data values in treatments is equal among the matched sets.*

*Assumptions:*

- The first (dependent/response) variable is binomial.
- The second (independent/predictor) variable is nominal/categorical with three values.
- The population is significantly larger than the sample.
- The sample accurately represents the population.

# Minimal Working Example

We feed the `cochrans.q()` function a 6 by 4 matrix of binomial values. The first column represents our dependent, binomial variable. The remaining columns represent our independent variable on three levels expressed as binomial values.

```
CochranMatrix <- matrix(c(1, 1, 1, 1, 1, 1, 1, 1, 0,  
  1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1), 6,  
  4)  
cochrans.q(CochranMatrix)
```

```
##  
## Cochran's Q Test  
##  
## H0: There is no difference in the effectiveness of treatments.  
## HA: There is a difference in the effectiveness of treatments.  
##  
## Q = 9.31578947368421  
##  
## Degrees of Freedom = 3  
##  
## Significance Level = 0.05  
## The p-value is 0.0253739987887868  
## There is enough evidence to conclude that the effectiveness of at least two treatments differ.  
##
```

# Purpose And Assumptions

## Chi-Squared

`chisq.test()` in base R

*Purpose:*

To tests whether distributions of categorical variables differ from one another thus identifying whether they are related.

$H_0$

*The distributions of nominal variables are equal.*

*Assumptions:*

- Variable values are nominal/categorical.
- The population is significantly larger than the sample.
- The sample accurately represents the population.
- Sampled values are independent (one value does not influence another).

# Minimal Working Example - One Sample Situation I

We feed the `chisq.test()` function an unbiased nominal set of three levels:

```
set.seed(42)
ChiMat1 <- table(sample(c("A", "B", "C"), 1000, replace = TRUE))
ChiMat1
```

```
##
##   A    B    C
## 346 329 325
```

```
chisq.test(ChiMat1)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  ChiMat1
## X-squared = 0.7, df = 2, p-value = 0.7
```

Obviously, the observed distribution does not differ from our expectation of equally distributed proportions and so the test concludes non-significantly.

# Minimal Working Example - One Sample Situation II

We feed the `chisq.test()` function a skewed (towards "A") nominal set of three levels:

```
set.seed(42)
ChiMat2 <- table(sample(c("A", "B", "C"), 1000, replace = TRUE,
  prob = c(0.8, 0.1, 0.1)))
ChiMat2
```

```
##
##   A    B    C
## 804 111  85
```

```
chisq.test(ChiMat2)
```

```
##
##   Chi-squared test for given probabilities
##
## data:  ChiMat2
## X-squared = 1000, df = 2, p-value <2e-16
```

Obviously, the observed distribution does differ from our expectation of equally distributed proportions and so the test concludes significantly.

# Minimal Working Example - Two Sample Situation

We feed the `chisq.test()` our unbiased as well as our skewed (towards "A") nominal set of three levels to see whether their distributions differ significantly.

```
ChiMatrix <- cbind(ChiMat1, ChiMat2)
ChiMatrix
```

```
##      ChiMat1 ChiMat2
## A      346      804
## B      329      111
## C      325       85
```

```
chisq.test(ChiMatrix)
```

```
##
## Pearson's Chi-squared test
##
## data:  ChiMatrix
## X-squared = 400, df = 2, p-value <2e-16
```

Clearly, they do differ significantly.

# Variables We Can Use

Which variables in our *Passer domesticus* data set are nominal?

- Site Index
- Climate
- Population Status
- Colour
- Sex
- Nesting Site
- Flock
- Home Range
- Predator Presence
- Predator Type

All of these are nominal but some are binomial.

# Research Questions And Hypotheses

So which of our major research questions (seminar 6) can we answer?

## Binomial Test

- *Sexual Dimorphism*: Are the sexes represented in equal proportions?
- *Predation*: Are our sites dominated by predators or not?

## McNemar

- *Sexual Dimorphism*: Compare sex ratio over time (we need to generate some new data for this using the sample function).

## Cochran's Q

- *Sexual Dimorphism*: Are colours related to sex?
- *Predation*: Are colours related to predators?

## Chi-Squared

- *Sexual Dimorphism*: Are colours related to sex?
- *Predation*: Are colours or nesting sites related to predators?