A Handbook of Statistical Analyses Using **R** — 3rd Edition

Torsten Hothorn and Brian S. Everitt

Conditional Inference: Guessing Lengths, Suicides, Gastrointestinal Damage, and Newborn Infants

4.1 Introduction

4.2 Conditional Test Procedures

4.3 Analysis Using **R**

4.3.1 Estimating the Width of a Room Revised

The unconditional analysis of the room width estimated by two groups of students in Chapter 3 led to the conclusion that the estimates in meters are slightly larger than the estimates in feet. Here, we reanalyze these data in a conditional framework. First, we convert meters into feet and store the vector of observations in a variable y:

```
R> data("roomwidth", package = "HSAUR3")
R> convert <- ifelse(roomwidth$unit == "feet", 1, 3.28)
R> feet <- roomwidth$unit == "feet"
R> meter <- !feet
R> y <- roomwidth$width * convert
```

The test statistic is simply the difference in means

```
R> T <- mean(y[feet]) - mean(y[meter])
R> T
```

```
[1] -8.86
```

In order to approximate the conditional distribution of the test statistic T we compute 9999 test statistics for shuffled y values. A permutation of the y vector can be obtained from the `sample` function.

```
R> meandiffs <- double(9999)
R> for (i in 1:length(meandiffs)) {
+     sy <- sample(y)
+     meandiffs[i] <- mean(sy[feet]) - mean(sy[meter])
+ }
```

The distribution of the test statistic T under the null hypothesis of independence of room width estimates and groups is depicted in Figure 4.1. Now, the value of the test statistic T for the original unshuffled data can be compared

```
R> hist(meandiffs)
R> abline(v = T, lty = 2)
R> abline(v = -T, lty = 2)
```

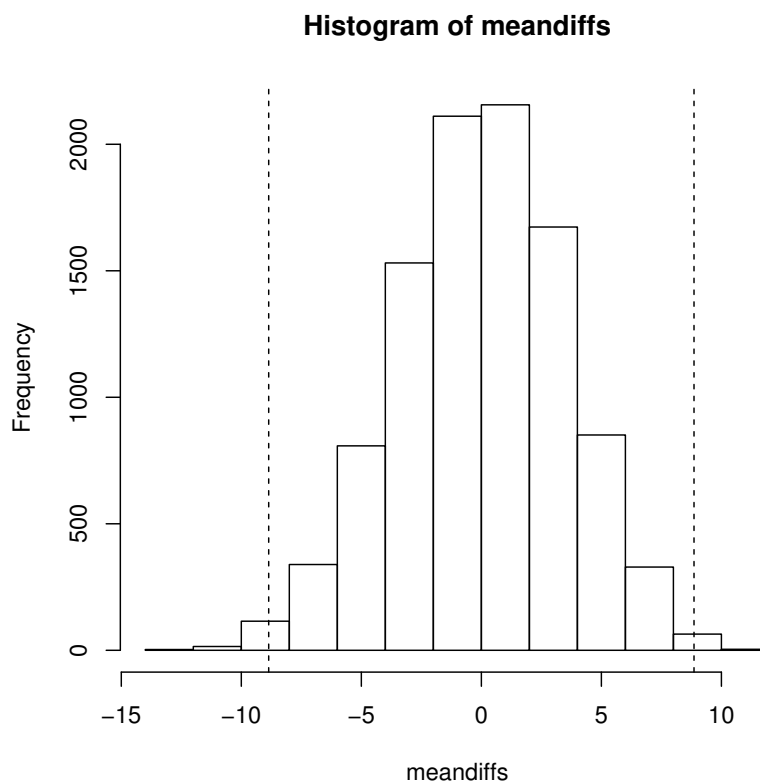**Histogram of meandiffs**

Figure 4.1   An approximation for the conditional distribution of the difference of mean `roomwidth` estimates in the feet and meters group under the null hypothesis. The vertical lines show the negative and positive absolute value of the test statistic T obtained from the original data.

with the distribution of T under the null hypothesis (the vertical lines in Figure 4.1). The p-value, i.e., the proportion of test statistics T larger than 8.859 or smaller than -8.859, is

```
R> greater <- abs(meandiffs) > abs(T)
R> mean(greater)
```
[1] 0.008

with a confidence interval of

```
R> binom.test(sum(greater), length(greater))$conf.int
```

```
[1] 0.00635 0.00995
attr(,"conf.level")
[1] 0.95
```

Note that the approximated conditional *p*-value is roughly the same as the *p*-value reported by the *t*-test in Chapter 3.

```
R> library("coin")
R> independence_test(y ~ unit, data = roomwidth,
+                     distribution = exact())

        Exact General Independence Test

data:  y by unit (feet, metres)
Z = -3, p-value = 0.008
alternative hypothesis: two.sided
```

Figure 4.2   R output of the exact permutation test applied to the `roomwidth` data.

```
R> wilcox_test(y ~ unit, data = roomwidth,
+              distribution = exact())

        Exact Wilcoxon-Mann-Whitney Test

data:  y by unit (feet, metres)
Z = -2, p-value = 0.03
alternative hypothesis: true mu is not equal to 0
```

Figure 4.3   R output of the exact conditional Wilcoxon rank sum test applied to the `roomwidth` data.

4.3.2 Crowds and Threatened Suicide

```
R> data("suicides", package = "HSAUR3")
R> fisher.test(suicides)

        Fisher's Exact Test for Count Data

data:  suicides
p-value = 0.08
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  0.731 91.029
sample estimates:
odds ratio
      6.3
```

Figure 4.4   R output of Fisher's exact test for the `suicides` data.

4.3.3 Gastrointestinal Damage

Here we are interested in the comparison of two groups of patients, where one group received a placebo and the other one Misoprostol. In the trials shown here, the response variable is measured on an ordered scale – see Table **??**. Data from four clinical studies are available and thus the observations are naturally grouped together. From the *data.frame* `Lanza` we can construct a three-way table as follows:

```
R> data("Lanza", package = "HSAUR3")
R> xtabs(~ treatment + classification + study, data = Lanza)
, , study = I

            classification
treatment       1  2  3  4  5
  Misoprostol  21  2  4  2  0
  Placebo       2  2  4  9 13

, , study = II

            classification
treatment       1  2  3  4  5
  Misoprostol  20  4  6  0  0
  Placebo       8  4  9  4  5

, , study = III

            classification
treatment       1  2  3  4  5
  Misoprostol  20  4  3  1  2
  Placebo       0  2  5  5 17

, , study = IV

            classification
treatment       1  2  3  4  5
  Misoprostol   1  4  5  0  0
  Placebo       0  0  0  4  6
```

For the first study, the null hypothesis of independence of treatment and gastrointestinal damage, i.e., of no treatment effect of Misoprostol, is tested by

```
R> library("coin")
R> cmh_test(classification ~ treatment, data = Lanza,
+           scores = list(classification = c(0, 1, 6, 17, 30)),
+           subset = Lanza$study == "I")

       Asymptotic Linear-by-Linear Association Test

data:  classification (ordered) by treatment (Misoprostol, Placebo)
```

```
Z = -5, p-value = 8e-08
alternative hypothesis: two.sided
```

and, by default, the conditional distribution is approximated by the corresponding limiting distribution. The p-value indicates a strong treatment effect. For the second study, the asymptotic p-value is a little bit larger:

```
R> cmh_test(classification ~ treatment, data = Lanza,
+           scores = list(classification = c(0, 1, 6, 17, 30)),
+           subset = Lanza$study == "II")

        Asymptotic Linear-by-Linear Association Test

data: classification (ordered) by treatment (Misoprostol, Placebo)
Z = -3, p-value = 5e-04
alternative hypothesis: two.sided
```

and we make sure that the implied decision is correct by calculating a confidence interval for the exact p-value:

```
R> p <- cmh_test(classification ~ treatment, data = Lanza,
+           scores = list(classification = c(0, 1, 6, 17, 30)),
+           subset = Lanza$study == "II", distribution =
+           approximate(B = 19999))
R> pvalue(p)

[1] 5e-05
99 percent confidence interval:
 2.51e-07 3.71e-04
```

The third and fourth study indicate a strong treatment effect as well:

```
R> cmh_test(classification ~ treatment, data = Lanza,
+           scores = list(classification = c(0, 1, 6, 17, 30)),
+           subset = Lanza$study == "III")

        Asymptotic Linear-by-Linear Association Test

data: classification (ordered) by treatment (Misoprostol, Placebo)
Z = -5, p-value = 1e-07
alternative hypothesis: two.sided

R> cmh_test(classification ~ treatment, data = Lanza,
+           scores = list(classification = c(0, 1, 6, 17, 30)),
+           subset = Lanza$study == "IV")

        Asymptotic Linear-by-Linear Association Test

data: classification (ordered) by treatment (Misoprostol, Placebo)
Z = -4, p-value = 7e-05
alternative hypothesis: two.sided
```

At the end, a separate analysis for each study is unsatisfactory. Because the design of the four studies is the same, we can use study as a block variable and perform a global linear-association test investigating the treatment effect

of Misoprostol in all four studies. The block variable can be incorporated into
the *formula* by the | symbol.

```
R> cmh_test(classification ~ treatment | study, data = Lanza,
+           scores = list(classification = c(0, 1, 6, 17, 30)))

        Asymptotic Linear-by-Linear Association Test

data:  classification (ordered) by
        treatment (Misoprostol, Placebo)
        stratified by study
Z = -9, p-value <2e-16
alternative hypothesis: two.sided
```

Based on this result, a strong treatment effect can be established.

4.3.4 Teratogenesis

In this example, the medical doctor (MD) and the research assistant (RA)
assessed the number of anomalies $(0, 1, 2$ or $3)$ for each of 395 babies:

```
R> anomalies <- c(235, 23, 3, 0, 41, 35, 8, 0,
+                 20, 11, 11, 1, 2, 1, 3, 1)
R> anomalies <- as.table(matrix(anomalies,
+     ncol = 4, dimnames = list(MD = 0:3, RA = 0:3)))
R> anomalies
   RA
MD    0    1    2    3
  0 235   41   20    2
  1  23   35   11    1
  2   3    8   11    3
  3   0    0    1    1
```

We are interested in testing whether the number of anomalies assessed by the
medical doctor differs structurally from the number reported by the research
assistant. Because we compare *paired* observations, i.e., one pair of measure-
ments for each newborn, a test of marginal homogeneity (a generalization of
McNemar's test, Chapter 3) needs to be applied:

```
R> mh_test(anomalies)

        Asymptotic Marginal Homogeneity Test

data:  response by
        conditions (MD, RA)
        stratified by block
chi-squared = 20, df = 3, p-value = 9e-05
```

The p-value indicates a deviation from the null hypothesis. However, the levels
of the response are not treated as ordered. Similar to the analysis of the
gastrointestinal damage data above, we can take this information into account
by the definition of an appropriate score. Here, the number of anomalies is a
natural choice:

```
R> mh_test(anomalies, scores = list(response = c(0, 1, 2, 3)))

        Asymptotic Marginal Homogeneity Test for Ordered Data

data:  response (ordered) by
        conditions (MD, RA)
        stratified by block
Z = -5, p-value = 5e-06
alternative hypothesis: two.sided
```

In our case, one can conclude that the assessment of the number of anomalies differs between the medical doctor and the research assistant.