

# Medical Statistics – Lab 4

R version

Welcome to lab 4 in the medical statistics course. In this lab, we will focus on the analysis of categorical data and the comparison of proportions between groups. We will also perform several statistical tests for the analysis of paired data.

We will use the following R packages: `haven`, `ggplot2`, and `rcompanion`. The first two packages were also used in the previous labs and therefore already installed on your computer. For `rcompanion` (used to perform pairwise chi-square tests) you may have to download and install it first by running `install.packages("rcompanion")`.

## Part 1: Inference for categorical data

### Smoking and post-surgical complications

A study was conducted to investigate whether smoking is associated with an increased risk of post-surgical complications. The relationship between smoking status (smoker or non-smoker) and the occurrence of complications following surgery was examined. The outcome of interest was whether or not a complication occurred (yes or no), with smoking status serving as the explanatory variable to compare complication rates between the two groups.

The data from the study are summarized in the following 2x2 contingency table:

	Complication	No Complication	Total
Smokers	8	12	20
Non-smokers	10	50	60
Total	18	62	80

## Confidence intervals and hypothesis testing for the difference in proportions using the normal approximation

### Question 1

Using the data provided in the table, calculate an approximate 95% confidence interval for the difference in proportions of post-surgical complications between smokers and non-smokers.

### Question 2

Based on the 95% confidence interval, can we conclude that there is a statistically significant difference in the proportion of post-surgical complications between smokers and non-smokers?

We can also use the normal approximation to test the hypothesis that the proportion of complications is the same for smokers and non-smokers. This test is known as the two-sample Z test for equality of proportions.

As explained in the syllabus, the two-sample Z test uses the pooled population proportion  $\hat{p}$ , which is calculated as the total number of events divided by the total sample size. This pooled proportion is used under the null hypothesis, which assumes that the two groups share the same underlying proportion. The standard error of the difference in proportions is then calculated as  $\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}$ , where  $n_1$  and  $n_2$  are the sample sizes in the two groups.

In contrast, the 95% confidence interval for the difference in proportions does not rely on the pooled proportion. Instead, it calculates the standard error separately for each group using the observed proportions, resulting in an unpooled standard error:  $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$ , where  $p_1$  and  $p_2$  are the sample proportions for each group. This approach provides an interval that better reflects the variability in the observed data, independent of the null hypothesis assumption.

To conduct the two-sample Z test, we start by creating a contingency table:

```
# Create a contingency table
complications <- matrix(c(8, 12, 10, 50), nrow = 2, byrow = TRUE)
colnames(complications) <- c("Complication", "No Complication")
rownames(complications) <- c("Smokers", "Non-smokers")
complications <- as.table(complications)
complications
```

Next, we supply this table to the `prop.test()` function to calculate the test statistic and p-value:

```
prop.test(complications)
```

### Question 3

Based on the results of the test, can we conclude that there is a statistically significant difference in the proportion of post-surgical complications between smokers and non-smokers?

### Question 4

In addition to the p-value, output of the `prop.test()` function also provides an approximate 95% confidence interval for the difference in proportions. How does this confidence interval compare to the one you calculated manually?

## Checking of assumptions

For the use of the normal approximation to be valid, the expected number of events and non-events in each group should be at least 5.

### Exercise

Check this assumption by calculating the expected counts for each cell in the contingency table.

### Question 5

Is it reasonable to use the normal approximation in this case?

## Fisher's exact test

When the expected cell counts are small, the normal approximation may not be appropriate. In such cases, Fisher's exact test is recommended for testing the association between two categorical variables.

In R, we can perform Fisher's exact test using the `fisher.test()` function. The test is based on the hypergeometric distribution and provides an exact p-value for the association between the two variables:

```
fisher.test(complications)
```

### Question 6

Based on the results of Fisher's exact test, can we conclude that there is a statistically significant difference in the proportion of post-surgical complications between smokers and non-smokers?

## Vaccine side effects across age groups

A study was conducted to investigate whether the occurrence of vaccine side effects differs across age groups. Researchers categorized side effects into three types: none, mild, and severe. The study participants were divided into three age groups: 18–39, 40–59, and 60+, and data was collected on the type of side effect experienced by individuals in each group.

The research objective was to determine whether the distribution of side effects is consistent across these age groups.

The data is summarized in the following contingency table:

Age Group	None	Mild	Severe	Total
18–39	50	30	10	90
40–59	40	40	20	100
60+	30	50	40	120
Total	120	120	70	310

## Chi-square test of homogeneity

To conduct a chi-square test of homogeneity, we start by creating a contingency table:

```
# Create a contingency table
side_effects <- matrix(c(50, 30, 10, 40, 40, 20, 30, 50, 40), nrow = 3, byrow = TRUE)
colnames(side_effects) <- c("None", "Mild", "Severe")
rownames(side_effects) <- c("18-39", "40-59", "60+")
side_effects <- as.table(side_effects)
side_effects
```

Next, we supply this table to the `chisq.test()` function to calculate the test statistic and p-value:

```
chisq_test_overall <- chisq.test(side_effects)
print(chisq_test_overall)
```

### Question 7

Based on the results of the chi-square test, can we conclude that the distribution of vaccine side effects is consistent across the three age groups?

### Checking of assumptions

To use the chi-square test, the expected cell counts should be at least 5 for most cells. To check this assumption, we retrieve the table of expected counts from the output of the `chisq.test()` function, which we stored in the `chisq_test_overall` object:

```
# Retrieve the table of expected counts
chisq_test_overall$expected
```

### Question 8

Are the expected cell counts greater than 5 for the different cells in the contingency table?

### Post-hoc pairwise comparisons

Finally, we are interested in determining which age groups have significantly different distributions of side effects. For a  $n \times 2$  table, this can be achieved easily using the `pairwise.prop.test()` function, which performs pairwise comparisons of proportions between groups with adjustments for multiple testing. However, in our case, we are working with a  $3 \times 3$  contingency table, where the outcomes have more than two categories (None, Mild, Severe).

For contingency tables with more than two outcome categories, we can use the `pairwiseNominalIndependence()` function from the `rcompanion` package. This function performs pairwise chi-square tests between all pairs of rows (or columns) and applies a correction for multiple testing:

```
library(rcompanion)
pairwiseNominalIndependence(side_effects,
                             compare = "row",
                             fisher = FALSE,
                             method = "bonferroni")
```

The output shows the results of pairwise comparisons between all age groups. The key columns to focus on are:

- **Comparison:** The pair of age groups being compared
- **p.Chisq:** The unadjusted p-value from the chi-square test

- **p.adj.Chisq**: The Bonferroni-adjusted p-value for multiple comparisons

We use the adjusted p-values (**p.adj.Chisq**) to determine statistical significance, as these account for the increased risk of Type I errors when performing multiple comparisons. At a significance level of 0.05, we can see that the distribution of side effects differs significantly between the 18–39 and 60+ age groups (adjusted  $p < 0.001$ ), while the other pairwise comparisons are not statistically significant after correction.

#### Note

The output also includes p-values from the G-test (**p.Gtest** and **p.adj.Gtest**), which is an alternative to the chi-square test based on the likelihood ratio. In most cases, the chi-square and G-test give similar results.

Alternatively, we can perform the pairwise comparisons manually to better understand the underlying process. We start by comparing the first two age groups (18–39 and 40–59). First, we set up the contingency table for these two groups:

```
table_12 <- side_effects[c("18-39", "40-59"), ]
table_12
```

Next, we perform the chi-square test for this subset of the data and adjust the p-value by applying the Bonferroni correction:

```
# Perform chi-square test for the subset of data
chisq_test_12 <- chisq.test(table_12)
print(chisq_test_12)

# Adjust the p-value for multiple testing
p_adjusted_12 <- 3*chisq_test_12$p.value
p_adjusted_12
```

#### Exercise

Perform the pairwise comparison between the other two pairs of age groups (40–59 and 60+, 18–39 and 60+) using the same approach.

#### Question 9

Based on the results of the pairwise comparisons, which age groups have significantly different distributions of side effects?

## Part 2: Analysis of paired continuous data

### Introduction

In this part of the lab, we will analyze paired data on pocket depth before and after an intervention. Pocket depth refers to the depth of the gum pockets around teeth, measured using a periodontal probe. It is an important indicator of periodontal health. Healthy gums typically have pocket depths less than 3 mm, while deeper pockets may indicate conditions such as gingivitis or periodontitis.

The dataset `pockets_paired.sav`, available from the Datasets menu, contains the following columns:

- `subjectID`: Unique identifier for each participant
- `pocket_depth_before`: Average pocket depth (in mm) measured before the intervention
- `pocket_depth_after`: Average pocket depth (in mm) measured after the intervention

The objective is to determine whether the intervention significantly reduces pocket depth. We will apply three statistical methods to analyze the paired data:

1. Paired t-test
2. Sign test
3. Wilcoxon signed-rank test

### Paired t-test

First, we load the data from the provided SPSS file `pockets_paired.sav`:

```
library(haven)
pockets <- read_sav("pockets_paired.sav")
head(pockets)
```

Next, we perform a paired t-test to compare the average pocket depth before and after the intervention:

```
# Perform paired t-test
t_test <- t.test(pockets$pocket_depth_before,
                 pockets$pocket_depth_after,
                 paired = TRUE)
print(t_test)
```

The `t.test()` function with the argument `paired = TRUE` conducts a paired t-test; setting `paired = FALSE` (default) would perform an independent samples t-test. The output includes the test statistic, degrees of freedom, and the p-value.

#### Question 10

Based on the results of the paired t-test, can we conclude that the intervention significantly reduces pocket depth?

### Checking of assumptions

To determine whether it is appropriate to apply the paired t-test to these data, we need to verify that the differences in pocket depth before and after the intervention are normally distributed. We can visually inspect the distribution of differences using a histogram:

```
# Calculate the differences in pocket depth
pockets$diff <- pockets$pocket_depth_after -
               pockets$pocket_depth_before

# Create a histogram of the differences
library(ggplot2)
ggplot(pockets, aes(x = diff)) +
  geom_histogram(binwidth = 0.05,
                 fill = "skyblue",
                 color = "black") +
  labs(title = "Distribution of differences in pocket depth",
       x = "Difference in pocket depth (mm)",
       y = "Frequency")
```

#### Question 11

Based on the histogram, do the differences in pocket depth appear to be approximately normally distributed?

### Sign test

The sign test is a non-parametric test used to compare two related samples. It is based on the signs of the differences between the pairs of observations. We will apply the sign test to the pocket depth data to determine whether the intervention has a significant effect.

First, we calculate the total number of positive and negative signs in the differences:



```
# Calculate the number of positive and negative signs
signs_positive <- sum(pockets$diff > 0)
signs_negative <- sum(pockets$diff < 0)
```

Next, we perform the sign test using the `binom.test()` function, which calculates the exact p-value for the sign test:

```
# Perform the sign test
sign_test <- binom.test(signs_positive,
                        signs_positive + signs_negative,
                        p = 0.5)
print(sign_test)
```

#### Question 12

Based on the results of the sign test, can we conclude that the intervention significantly reduces pocket depth?

### Wilcoxon signed-rank test

The Wilcoxon signed-rank test is another non-parametric test used to compare two related samples. It is based on the ranks of the absolute differences between the pairs of observations. In this case, the sign test is more appropriate because the Wilcoxon signed-rank test requires the assumption of symmetry in the distribution of differences, whereas the previously constructed histogram suggests that the distribution of these differences is left-skewed.

To perform the Wilcoxon signed-rank test, we use the `wilcox.test()` function:

```
# Perform the Wilcoxon signed-rank test
wilcox_test <- wilcox.test(pockets$pocket_depth_before,
                           pockets$pocket_depth_after,
                           paired = TRUE)
print(wilcox_test)
```

#### Question 13

Based on the results of the Wilcoxon signed-rank test, can we conclude that the intervention significantly reduces pocket depth?

## Part 3: Analysis of paired dichotomous data

In this part of the lab, we will analyze paired dichotomous data from a study investigating the skin response to two substances: **dinitrochlorobenzene (DNCB)**, a contact allergen, and **croton oil**, a skin irritant. The objective is to determine whether the proportion of patients with a negative response to DNCB is the same as the proportion with a negative response to croton oil.

The data, summarized in the following table, represents the results of simultaneous skin reaction tests on 173 patients with skin cancer:

	DNCB +ve	DNCB -ve	Total
Croton oil +ve	81	23	104
Croton oil -ve	48	21	69
Total	129	44	173

Because the data are paired, we will use the McNemar test to compare the proportions of positive and negative responses to DNCB and croton oil.

First, we create a 2x2 contingency table from the data:

```
# Create the table for DNCB and Croton Oil responses
skin_response_table <- matrix(c(81, 23, 48, 21), nrow = 2, byrow = TRUE)
colnames(skin_response_table) <- c("DNCB +ve", "DNCB -ve")
rownames(skin_response_table) <- c("Croton Oil +ve", "Croton Oil -ve")
skin_response_table <- as.table(skin_response_table)

# Print the table
skin_response_table
```

Next, we perform the McNemar test using the `mcnemar.test()` function:

```
# Perform the McNemar test
mcnemar_test <- mcnemar.test(skin_response_table)
print(mcnemar_test)
```

### Question 14

Based on the results of the McNemar test, can we conclude that there is a significant difference in the proportions of patients with a negative response to DNCB and croton oil? If so, can you determine which substance is associated with a higher proportion of

negative responses?