Medical Statistics – Lab 3

SPSS version

Welcome to lab 3 in the medical statistics course. For today's exercises, we will continue exploring the `lowbwt.sav` dataset, which you can downloaded from the Dataset section of the menu.

As a reminder, the dataset includes the following variables (see the previous lab for more details):

Variable	Abbreviation
Identification Code	ID
Low Birth Weight (0 = Birth Weight 2500g, 1 = Birth Weight < 2500g)	low
Age of the Mother in Years	age
Weight in Pounds at the Last Menstrual Period	lwt
Ethnicity of the mother (1 = Caucasian, 2 = Afro-American, 3 = Asian)	ethnicity
Smoking Status During Pregnancy (1 = Yes, 0 = No)	smoke
History of Premature Labor (0 = None, 1 = One, etc.)	ptl
History of Hypertension (1 = Yes, 0 = No)	ht
Presence of Uterine Irritability (1 = Yes, 0 = No)	urirr
Number of Physician Visits During the First Trimester (0 = None, 1 = One, 2 = Two, etc.)	pvft
Birth Weight in Grams	bwt

Part 1: Independent Samples t-test and Mann-Whitney U Test

In this part of the lab, we will examine the effect of smoking during pregnancy on birth weight.

Exploratory data analysis

To get a sense of the data and the extent of a possible effect of smoking, we start by creating a boxplot. In SPSS, this can be achieved by following these steps:

1. **Create Boxplot**: Go to `Graphs` > `Chart Builder`.
2. **Select Boxplot**: In the `Chart Builder`, select `Boxplot` from the gallery.
3. **Select Variables**: Drag `bwt` to the y-axis and `smoke` to the x-axis to create a boxplot comparing the birth weights for mothers who smoked and those who did not.
4. **Run the Graph**: Click `OK` to generate the boxplot.

> Question 1
>
> Based on the boxplot, do you expect the smoking status to have an effect on birth weight?

Independent samples t-test

Next, we will perform an independent samples t-test to compare the mean birth weights between mothers who smoked and those who did not. To conduct this analysis in SPSS, follow these steps:

1. **Perform the Test**: Go to `Analyze` > `Compare Means` > `Independent-Samples t Test`.
2. **Select Test Variable and Grouping Variable**: Move `bwt` to the `Test Variable` box and `smoke` to the `Grouping Variable` box.
3. **Define Groups**: Click on `Define Groups` and specify `0` and `1` for mothers who did not smoke and mothers who smoked, respectively.
4. **Run the Test**: Click `OK` to run the analysis.

SPSS will generate several tables. Here we focus on the table **Independent Samples Test**, which includes:

- **Levene's Test for Equality of Variances**: Tests the assumption of equal variances between the two groups. The **F** and **Sig.** (p-value) columns represent the results of this test. If the **p-value** is greater than 0.05, the assumption of equal variances holds, and you should use the row labeled "Equal variances assumed" for interpreting the t-test results. If the **p-value** is less than or equal to 0.05, the assumption of equal variances is violated,

and you should use the row labeled "Equal variances not assumed" for interpreting the t-test results.

- **t-test for Equality of Means**: Provides the t-test results, including the t-value, degrees of freedom, and p-value. It also includes the **Mean Difference** and **95% Confidence Interval of the Difference**. It is important to choose the appropriate row ("Equal variances assumed" or "Equal variances not assumed") based on the result of Levene's test.

> Question 2
>
> Based on the Levene test, does the assumption of equal variances hold?

> Question 3
>
> Is there a significant difference in birth weights between mothers who smoked and those who did not?

Checking of assumptions

To assess whether the assumption of normality holds for the outcome variable in both groups, follow these steps to create histograms:

1. **Create the Graph**: Go to `Graphs` > `Chart Builder`.
2. **Create Histogram**: In the `Chart Builder`, select `Histogram` from the gallery.
3. **Select Variables**: Drag `bwt` into the x-axis box.
4. **Split by Group**: In the `Chart Builder`, click on the `Groups/Points ID` tab. Check the `Rows panel variable` checkbox to add a panel field to the chart preview. Then, drag the `smoke` variable into this field to split the histograms by group.
5. **Run the Graph**: Click `OK` to generate the histograms.

> Question 4
>
> Do the histograms indicate that the birth weight data are approximately normally distributed for both groups?

95% Confidence Interval for the mean difference

In addition to performing hypothesis tests, it is often informative to estimate the effect size and its uncertainty. One way to do this is by calculating a confidence interval for the mean difference in birth weight between the two groups. The 95% confidence interval is included in the output provided by the independent samples t-test procedure, so in principle, we could

extract it from there. As an exercise, we are also going to calculate it manually based on the formulas provided in the lecture/course syllabus.

To calculate the summary statistics required for the manual calculation, follow these steps:

1. **Navigate to Explore Analysis**:
    - Go to `Analyze` > `Descriptive Statistics` > `Explore`.
2. **Select Variables**:
    - Move `bwt` to the `Dependent List` box.
    - Move `smoke` to the `Factor List` box.
3. **Set Display Options**:
    - Check the option `Statistics` to prevent SPSS from cluttering the output with a range of plots.
4. **Run the Analysis**:
    - Click `OK` to run the analysis. The output will provide group-wise means, standard deviations, and sample sizes, which can be used for manual calculations.

Exercise

Based on these summary statistics, calculate the pooled standard deviation and the standard error of the mean difference. Then compute the 95% confidence interval for the mean difference in birth weight between mothers who smoked and those who did not. You may simplify the calculation by using the 97.5th percentile from the standard normal distribution (1.96) rather than the corresponding percentile from the t-distribution.

Question 5

Does your manually calculated 95% confidence interval for the mean difference in birth weight between the two groups agree with the one provided in the output of the independent samples t-test procedure?

Mann-Whitney U Test

In case the assumption of normality is violated, we can use the Mann-Whitney U test as a non-parametric alternative to the independent samples t-test. To perform this test in SPSS, follow these steps:

1. **Navigate to the Mann-Whitney U Test**:

- Go to `Analyze` > `Nonparametric Tests` > `Legacy Dialogs` > `2 Independent Samples`.

2. **Select Variables**:
   - Move `bwt` to the `Test Variable List`.
   - Move `smoke` to the `Grouping Variable`.
   - Click `Define Groups` and specify the values of the `smoke` variable (e.g., 0 = No, 1 = Yes). Click `Continue`.
3. **Select the Mann-Whitney U Test**:
   - Under `Test Type`, ensure that **Mann-Whitney U** is selected.
4. **Run the Test**:
   - Click `OK` to run the analysis. The output will include the test statistic and associated p-value.

> Question 6
>
> What are the null and alternative hypotheses for the Mann-Whitney U test, and what does the p-value indicate about the difference in birth weight between mothers who smoked and those who did not?

Part 2: One-Way ANOVA and Kruskal-Wallis Test

In this part of the lab, we are going to examine the effect of ethnicity on birth weight.

Exploratory data analysis

To get a sense of the differences in birth weights across the different ethnic groups, we will create a boxplot. In SPSS, follow these steps:

1. **Create Boxplot**:
   - Go to `Graphs` > `Chart Builder`.
   - Select `Boxplot` from the gallery.
2. **Assign Variables**:
   - Drag `bwt` to the y-axis.
   - Drag `ethnicity` to the x-axis.
3. **Run the Graph**:
   - Click `OK` to generate the boxplot.

Question 7

What does the boxplot suggest about the distribution of birth weights across different ethnic groups?

One-way ANOVA

To test whether the mean birth weights differ across ethnic groups, we will perform a one-way ANOVA. Follow these steps:

1. **Perform the Test**:
   - Go to `Analyze` > `Compare Means` > `One-Way ANOVA`.
2. **Select Variables**:
   - Set **Dependent Variable**: Birth Weight (`bwt`).
   - Set **Factor**: Ethnicity (`ethnicity`).
3. **Run the Test**:
   - Click `OK` to run the analysis.

SPSS will generate the following key output:

- **ANOVA Table**: Contains the F-statistic and the associated p-value. These values indicate whether there is a statistically significant difference in mean birth weights across ethnic groups.

Question 8

What conclusions can be drawn from the results of the one-way ANOVA?

Post-hoc tests

If the one-way ANOVA indicates a statistically significant difference, we will perform post-hoc tests to determine which specific groups differ from each other. To do this in SPSS:

1. **Enable Post-Hoc Tests**:
   - In the one-way ANOVA dialog, click on the `Post Hoc` button.
   - Select **Bonferroni** as the method.
2. **Run the Test**:
   - Click `Continue` and then `OK` to run the analysis.

The post-hoc output will include pairwise comparisons of the means between all groups, with adjusted p-values.

> Question 9
>
> Which groups differ significantly, and what do the results suggest about the differences in birth weights across ethnic groups?

Checking of assumptions

To assess whether the results of the one-way ANOVA are valid, we need to check the assumptions of normality and homogeneity of variances. This step is analogous to the previous examples, and is left as an exercise.

Kruskal-Wallis Test

If the assumptions of the one-way ANOVA are not satisfied, we can use the Kruskal-Wallis test as a non-parametric alternative. To perform this test in SPSS, follow these steps:

1. **Navigate to the Kruskal-Wallis Test**:
   - Go to `Analyze` > `Nonparametric Tests` > `Legacy Dialogs` > `K Independent Samples`.
2. **Select Variables**:
   - Move `bwt` to the `Test Variable List`.
   - Move `ethnicity` to the `Grouping Variable`.
   - Click `Define Range` and specify the range of the grouping variable (e.g., 1 to 3 for the ethnicity variable). Click `Continue`.
3. **Select the Kruskal-Wallis Test**:
   - Under `Test Type`, ensure that **Kruskal-Wallis H** is selected.
4. **Run the Test**:
   - Click `OK` to run the analysis. The output will include the test statistic and associated p-value.

> Question 10
>
> Are the results of the Kruskal-Wallis test consistent with the one-way ANOVA results?

Part 3: Unguided exercises

Effect of hypertension on birth weight

Examine the effect of history of hypertension on birth weight by performing the following steps:

- Create a boxplot to visualize the distribution of birth weights by history of hypertension
- Perform an independent samples t-test to compare the mean birth weights between mothers with and without a history of hypertension
- Check the assumptions of the t-test, including normality and homogeneity of variances
- If the assumptions of the t-test are violated, perform a Mann-Whitney U test as a non-parametric alternative

Comparing red cell folate levels across ventilation strategies in cardiac bypass patients

Twenty-two patients undergoing cardiac bypass surgery were randomized to one of three ventilation groups:

- **Group I**: Received a 50% nitrous oxide and 50% oxygen mixture continuously for 24 hours
- **Group II**: Received a 50% nitrous oxide and 50% oxygen mixture only during the operation
- **Group III**: Received no nitrous oxide and a 35-50% oxygen mixture continuously for 24 hours

The data file `ex5_6.sav` contains the red cell folate levels for the three groups after 24 hours of ventilation. The aim of this study is to compare the three groups and test whether they have the same red cell folate levels.

Tasks

1. **Exploratory data analysis** Create a boxplot to visualize the distribution of red cell folate levels by ventilation group. Based on this plot:
    - What are your first conclusions regarding the means and variances of the different groups?
2. **Perform a one-way ANOVA**:
    - Interpret the results
    - Are the assumptions satisfied?
3. **Try a log transformation on the data**:

- Perform another one-way ANOVA
- Are the assumptions satisfied after the transformation?

4. **Determine which means differ**:

   - Which means do you think differ?
   - Explain your reasoning.

5. **Try a non-parametric approach**:

   - What are your conclusions from this method?