

## Homework #1 Solutions

### Submit your homework along with your SAS code in Canvas.

For each problem, state the methods and the results sections by specifically providing the following:

#### Methods:

1. Name or describe the tests used to answer the research question
2. Write down the statistical hypothesis
3. State the assumptions that are necessary for the validity of the statistical tests and the statistical techniques used to assess these assumptions.

#### Results:

1. Provide some descriptive statistics of the sample *in a univariate or bivariate analysis* as appropriate, and briefly summarize the sample characteristics.
2. Answer the research question by providing appropriate results (e.g. test statistic, degrees of freedom, p-value...).

#### Appendix:

1. Submit your SAS code

### Problem 1

In a study to explore the possibility of hormonal alteration in asthma, Weinstein et al. (A-54) collected data on 22 postmenopausal women with asthma and 22 age-matched, postmenopausal, women without asthma. The following are the dehydroepiandrosterone sulfate (DHEAS) values collected by the investigators. Do these data provide sufficient evidence to indicate a significant difference in mean DHEAS between the asthma and the non-asthma groups? Data are provided in both Excel and SAS in Blackboard as well as in the class folder in SAS Studio and save as **HW1\_Problem1**.

Without Asthma	With Asthma	Without Asthma	With Asthma
20.59	87.50	15.90	166.02
37.81	111.52	49.77	129.01
76.95	143.75	25.86	31.02
77.54	25.16	55.27	47.66
19.30	68.16	33.83	171.88
35.00	136.13	56.45	241.88
146.09	89.26	19.91	235.16
166.02	96.88	24.92	25.16
96.58	144.34	76.37	78.71
24.57	97.46	6.64	111.52
53.52	82.81	115.04	54.69

Source: Dr. Robert E. Weinstein. Used with permission.

### Problem 1 - Solutions

#### Methods:

1. **A two-sample pooled t-test** was used to test the research question of whether the data provide sufficient evidence to indicate a significant difference in mean DHEAS between the asthma and the non-asthma groups.
2. The **statistical hypotheses** associated with the test above are:  $H_0: \mu_A - \mu_{NA} = 0$  versus  $H_a: \mu_A - \mu_{NA} \neq 0$ , where  $\mu_A$  and  $\mu_{NA}$  are the population mean DHEAS for the asthma and the non-asthma groups, respectively.
3. The **assumptions** that are necessary for the validity of the statistical tests and the statistical techniques used to assess these assumptions:
  - a. Each group/population represents an **independent random sample**.
  - b. Each group distribution is **normally distributed** as verified with:
    - I. The Kolmogorov-Smirnov test (p-value > 0.15 for the asthma group and p-value = 0.0680 for the non-asthma group).
    - II. **Note 1:** The use of histograms and the probability plots to confirm normality revealed that the assumption is not met for the non-asthma group. The histogram shows a right-skewed pattern, and the data points on the probability plot deviate from the diagonal line (See appendix). If you choose these graphs to prove normality, then you should speculate about the potential reasons behind the observed departure from normality such as the relatively small sample size in each group.
    - III. **Note 2:** You can't use the Central Limit Theorem to prove normality in this context, given that both groups have a sample size of less than 30 (n = 22 in each group).
  - c. **Equal Variance** Assumption: F-test = 1.95, p-value = 0.1342, variances are equal. We assumed equal variances since the p-value for testing the equality of variances is 0.1342 failing to reject the null hypothesis of non-equal variances.

#### Results:

The study sample consisted of 44 postmenopausal women of which 50% (n=22) are asthmatic and half are non-asthmatic. The descriptive statistics of the sample are provided in **Table 1**. The mean (SD) DHEAS for the asthma group was 107.99 (60.01) while the mean (SD) DHEAS for the non-asthma group was 56.09 (42.98). The difference in means (SD) DHEAS between the two groups was 51.90 (52.19). The 95% confidence interval for the difference ranged from 20.14 to 83.65. This difference is statistically significant as shown by the results of the pooled two-sample t-test in **Table 1**: T= 3.30, df=42, and p-value = 0.0020. Since the p-value is small and less than 0.05, we rejected the null hypothesis and concluded that the data provided sufficient evidence to indicate a significant difference in mean DHEAS between the asthma and the non-asthma groups among postmenopausal women.

More specifically, the mean DHEAS among the asthmatic postmenopausal women is significantly higher than that of the non-asthmatic postmenopausal women.

**Table 1. Sample Characteristics by Asthma Status and Results of the Two-sample t test**

	With Asthma (N = 22)	Without Asthma (N = 22)	DHEAS Mean Difference	Pooled t value	Degrees of Freedom (df)	P- value
DHEAS (Mean $\pm$ SD)	107.99 $\pm$ 60.01	56.09 $\pm$ 42.98	51.90 $\pm$ 52.19	3.3	42	0.002

SD = Standard deviation

### Appendix:

#### 1. Submit your SAS code

```
***** Problem 1 Solutions: Two-Sample t-test *****;

Libname HW1 '~/MPHO712/Module 1'; *Created a library reference HW1 to my own
folder;

*Importing Problem 1 data;
FILENAME REFFILE '~/MPHO712/Module 1/HW1_Problem1.csv';

PROC IMPORT DATAFILE=REFFILE
    DBMS=CSV
    OUT=HW1.Problem1;
    GETNAMES=YES;
RUN;

*Showing the content of the SAS data set;
PROC CONTENTS DATA=HW1.Problem1; RUN;

*It is always better to keep a copy of the original data file. So, we will
save the original data into a temporary data set called "Problem1";

data Problem1;
    set HW1.Problem1;
run;

*Descriptive Statistics of the variables in the dataset;

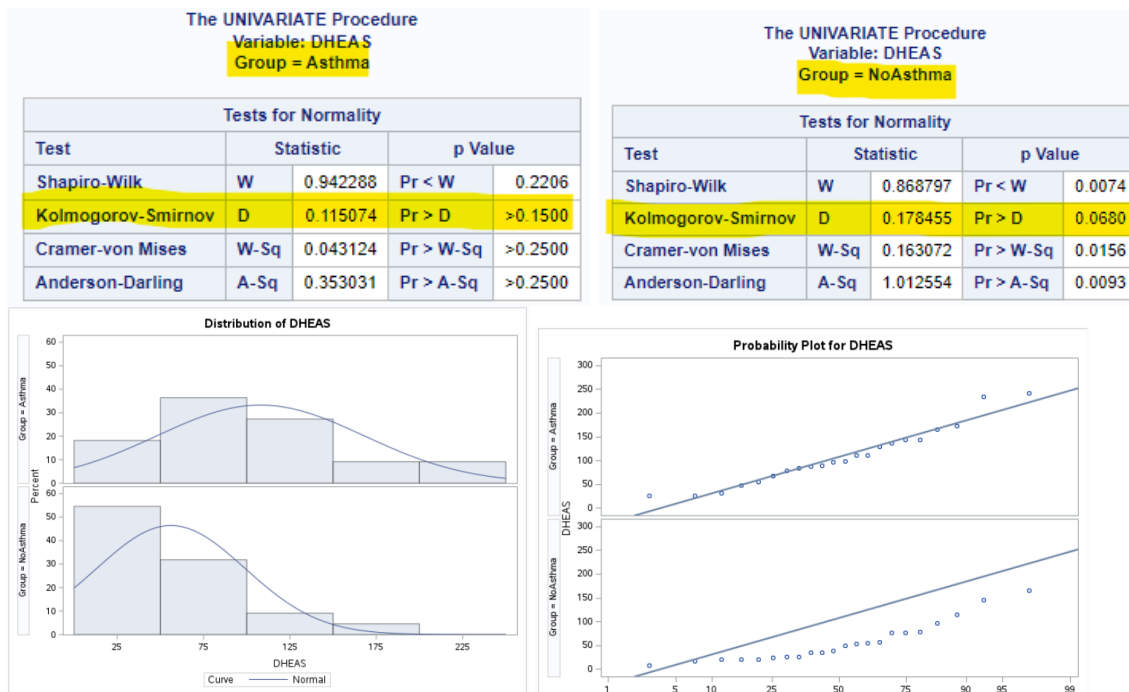
proc means data = Problem1 mean std median min max maxdec = 2;
    var Dheas;
    class group;
run;

proc freq data = Problem1;
    tables Group;
run;

/*** Checking for Normality ***/
```

```
proc univariate data=Problem1 normal;
    ods select ProbPlot;
    ods select Histogram;
    ods select TestsForNormality;
    class Group;
    var Dheas;
    probplot Dheas / normal(mu=est sigma=est);
    histogram Dheas / normal;
run;

title 'Two-sample t-test using the DHEAS data';
proc ttest data = Problem1 sides=2 plots(showh0);
    class Group;
    var Dheas;
run;
```



## Problem 2

In a study of violent victimization of women, Porcerelli et al. (A-26) collected information from 256 African American women and 388 Caucasian women ages 18 to 64 years at several family practice centers in the metropolitan Detroit area. Patients filled out a health history questionnaire that included a question about victimization. The following table shows the sample subjects cross-classified by race and the type of violent victimization reported among women. The victimization categories are defined as no victimization, partner victimization (and not by others), victimization by a person other than a partner (friend, family member, or stranger), and those who reported multiple victimization. Can we conclude on the basis of these data that for women, race and victimization status are not independent? Let  $\alpha = .05$ .

	No Victimization	Partner	Nonpartner	Multiple	Total
Caucasian	356	20	3	9	388
African-American	226	11	10	9	256
Total	582	31	13	18	644

Source: John H. Porcerelli, Rosemary Cogan, Patricia P. West, Edward A. Rose, Dawn Lambrecht, Karen E. Wilson, Richard K. Severson, and Dunia Karana, "Violent Victimization of Women and Men: Physical and Psychiatric Symptoms," *Journal of the American Board of Family Practice*, 16 (2003), 32–39.

**\*You'll need to manually enter the data from the table in SAS**

## Problem 2 - Solutions

### Methods:

1. A **Chi-Square test of independence** was used to test association between race and victimization status. A small p-value, say less than 0.05, is considered statistically significant.
2. The statistical hypotheses associated with the research question are:  **$H_0$  : race and victimization status are independent** versus  **$H_a$  : race and victimization status are not independent**.
3. The assumptions that are necessary for the validity of the statistical tests and the statistical techniques used to assess these assumptions:
  - a. We assume that the study sample is a **random sample** selected from the population of interest.
  - b. All **expected** frequency counts are larger than 5 (See Table 2 in Appendix).

### Results:

**Table 1** presents descriptive statistics for the sample, which comprised 644 women of African American and Caucasian backgrounds. Notably, the majority of both groups reported no experiences of victimization, with 91.75% of Caucasian women ( $n = 356$ ) and 88.28% of African American women ( $n=226$ ) falling into this category. Interestingly, African American women exhibited a higher incidence of non-partner victimization, accounting for 3.91% ( $n =10$ ) of the group compared to their Caucasian counterparts, where it was 0.77% ( $n = 3$ ).

To statistically examine the relationship between race and victimization type, we conducted a Chi-square test of independence, resulting in a chi-square value of 8.7308 with 3 degrees of freedom (DF) and a corresponding p-value of 0.0331. Given that the p-value is less than the conventional significance level of 0.05, we reject the null hypothesis. These findings provide statistical evidence supporting the assertion that there is an association between race and victimization status among women. In other words, race and the type of violent victimization are not independent variables in the context of this study.

**Table 1. Characteristics of the study sample**

Race	Type of violent victimization N (%)				P-value
	No Victimization	Partner	Nonpartner	Multiple	
Caucasian	356 (91.75)	20 (5.15)	3 (0.77)	9 (2.32)	0.0331
African-American	226 (88.28)	11 (4.30)	10 (3.91)	9 (3.52)	

## Appendix:

**The FREQ Procedure**

Expected	Table of race by status					
	race	status				
		None	Partner	Nonpartner	Multiple	Total
	Caucasian	350.65	18.677	7.8323	10.845	
	AA	231.35	12.323	5.1677	7.1553	
	Total	582	31	13	18	644

**Table 2: Expected Frequencies**

### 1. Submit your SAS code

```
***** Problem 2 Solutions: Chi-Square test of Independence *****;

** AA="American African", None="No Victimization" **;

data problem2;
  length race $14; *Specifies the number of bytes for storing variables;
  length status $14;
  input race $ status $ count;
  datalines;
Caucasian None 356
Caucasian Partner 20
Caucasian Nonpartner 3
Caucasian Multiple 9
AA None 226
AA Partner 11
AA Nonpartner 10
AA Multiple 9
;
run;

*Expected Frequencies;
proc freq data=problem2 order=data;
  weight count;
  tables race*status/nofreq nocol norow nopercent expected;
run;
```

```
*Chi-square test of independence*;
proc freq data=problem2 order=data;
    weight count;
    tables race*status/chisq expected;
run;
```

**Copyright Statement – PLEASE DO NOT SHARE THESE SOLUTIONS**

I, Dr. Hadiza Galadima, am the owner of all course materials, including but not limited to lecture notes, assignments, exams, solutions, and any other related content that I have created for MPHO 712. Consequently, students are strictly prohibited from reproducing, distributing, or publicly sharing my course materials without my obtaining my explicit written consent.