Yizhan Ao

Homework 3

April 4th, 2022

DATA HW3;

INFILE '/home/u58594663/my\_shared\_file\_links/schimiak/OldClassData.csv' delimiter=',' dsd;

INPUT Subject $

Gender $

Phone $

Campus $

Grade $

Car $

Optimist

Math

Siblings

Pets

Credit\_Hours

Social\_Media

Extra\_Curricular

Height

HS\_GPA

Exercise

Time\_To\_Get\_Ready

Distance;

1. Q1: Do a PROC FORMAT to change the following (answer to parts a and c put IN CODE.):

PROC FORMAT;

VALUE $New\_Gender ‘M’=‘Male’ ‘F’ =‘Female’;

RUN;

PROC FORMAT;

VALUE NEW\_MATH 1 ="I really like math."

2="I somewhat like math."

3="I could take math or leave it."

4="I really don't like math."

5="I'd rather have a root canal.";

RUN;

PROC PRINT;

FORMAT Gender $New\_Gender.;

FORMAT Math NEW\_MATH.;

RUN;

Application, table, Excel

Description automatically generated

1. Q2: Create a new variable for the letter grade (LETTER\_GRADE) for the HS\_GPA

if HS\_GPA>= 4.0 THEN LETTER\_GRADE = 'A';

else if HS\_GPA >= 3.0 AND HS\_GPA <=4.0 THEN LETTER\_GRADE = 'B';

else if HS\_GPA >= 2.0 AND HS\_GPA <=3.0 THEN LETTER\_GRADE = 'C';

else if HS\_GPA >= 2.0 AND HS\_GPA <=1.0 THEN LETTER\_GRADE = 'D';

else if HS\_GPA LT 1.0 THEN LETTER\_GRADE = 'F';

else if HS\_GPA = . THEN LETTER\_GRADE = '.';

Q3:

PROC GCHART DATA = HW3;

VBAR LETTER\_GRADE;

RUN;

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Q4: Do a hypothesis test: At a 0.05 significance level, test if there is a difference in the between the Male and Female High School GPA? State your conclusion based on your SAS results in the context of the question.

PROC TTEST ALPHA=0.05;

CLASS Gender;

VAR HS\_GPA;

RUN;

PROC NPAR1WAY WILCOXON;

CLASS Gender;

VAR HS\_GPA;

RUN;

Null Hypothesis (H0): GPA(males) = GPA(females).

Alternative Hypothesis (Ha): GPA(males) ≠ GPA(females).

We can't use a typical T-test since there's an outlier in the sample. For our scenario, we'll need to employ a Wilcoxon. Wilcoxon's p-value is 0.2288, which is larger than 0.05. This indicates that we do not reject the null hypothesis and that the alternative hypothesis is not supported by sufficient evidence.

**The TTEST Procedure**

**Variable: HS\_GPA**

| **Gender** | **Method** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **F** |  | 20 | 3.5000 | 0.4068 | 0.0910 | 2.7700 | 4.1200 |
| **M** |  | 22 | 3.3705 | 0.7228 | 0.1541 | 2.3000 | 4.8750 |
| **Diff (1-2)** | **Pooled** |  | 0.1295 | 0.5940 | 0.1835 |  |  |
| **Diff (1-2)** | **Satterthwaite** |  | 0.1295 |  | 0.1789 |  |  |

| **Gender** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **F** |  | 3.5000 | 3.3096 | 3.6904 | 0.4068 | 0.3093 | 0.5941 |
| **M** |  | 3.3705 | 3.0500 | 3.6909 | 0.7228 | 0.5561 | 1.0330 |
| **Diff (1-2)** | **Pooled** | 0.1295 | -0.2414 | 0.5005 | 0.5940 | 0.4877 | 0.7601 |
| **Diff (1-2)** | **Satterthwaite** | 0.1295 | -0.2343 | 0.4933 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 40 | 0.71 | 0.4844 |
| **Satterthwaite** | Unequal | 33.664 | 0.72 | 0.4741 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 21 | 19 | 3.16 | 0.0146 |

Chart, line chart

Description automatically generated

Chart, scatter chart

Description automatically generated

**The NPAR1WAY Procedure**

| **Wilcoxon Scores (Rank Sums) for Variable HS\_GPA Classified by Variable Gender** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Gender** | **N** | **Sum of Scores** | **Expected Under H0** | **Std Dev Under H0** | **Mean Score** |
| **Average scores were used for ties.** | | | | | |
| **F** | 20 | 474.0 | 430.0 | 39.544444 | 23.70 |
| **M** | 22 | 429.0 | 473.0 | 39.544444 | 19.50 |

| **Wilcoxon Two-Sample Test** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Statistic** | **Z** | **Pr > Z** | **Pr > |Z|** | **t Approximation** | |
| **Pr > Z** | **Pr > |Z|** |
| **Z includes a continuity correction of 0.5.** | | | | | |
| 474.0000 | 1.1000 | 0.1357 | 0.2713 | 0.1389 | 0.2777 |

| **Kruskal-Wallis Test** | | |
| --- | --- | --- |
| **Chi-Square** | **DF** | **Pr > ChiSq** |
| 1.2380 | 1 | 0.2658 |

Chart, box and whisker chart

Description automatically generated

Q5; Do a hypothesis test:

PROC TTEST ALPHA=0.05 H0=3.5;

VAR Exercise;

RUN;

PROC TTEST ALPHA=0.05 H0=3.5 SIDES=U;

VAR Exercise;

RUN;

Null Hypothesis (H0): Students exercise 3.5 times a week.

Alternative Hypothesis (Ha): Students exercise over 3.5 times a week.

Since there is no outlier in the sample, we can use a traditional T-test. The p-value in the T-test is 0.0184, which is smaller than 0.05. This means that we reject the null, and there is enough evidence to support the alternative hypothesis.

**The TTEST Procedure**

**Variable: Exercise**

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 46 | 4.0870 | 1.8895 | 0.2786 | 0 | 7.0000 |

| **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 4.0870 | 3.5259 | 4.6481 | 1.8895 | 1.5672 | 2.3798 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 45 | 2.11 | 0.0407 |

Chart, histogram

Description automatically generated

Chart, scatter chart

Description automatically generated

**The TTEST Procedure**

**Variable: Exercise**

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 46 | 4.0870 | 1.8895 | 0.2786 | 0 | 7.0000 |

| **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 4.0870 | 3.6191 | Infty | 1.8895 | 1.5672 | 2.3798 |

| **DF** | **t Value** | **Pr > t** |
| --- | --- | --- |
| 45 | 2.11 | 0.0204 |

Chart, histogram

Description automatically generated

Chart, scatter chart

Description automatically generated

Q6:

PROC MEANS CLM ALPHA=0.05 MAXDEC=2;

VAR TIME\_TO\_GET\_READY;

RUN;

People take between 38.77 and 55.14 minutes to get ready on average given a 95% confidence interval.

**The MEANS Procedure**

| **Analysis Variable : Time\_To\_Get\_Ready** | |
| --- | --- |
| **Lower 95% CL for Mean** | **Upper 95% CL for Mean** |
| 36.32 | 48.03 |