V506 Homework Exercise 4

# Authors: Jivitesh Poojary and Qiwen Zhu

# Part I: Do the following problems on the indicated pages from Lind, Marchal, and Wathen. Each problem is worth four points, for a total of 22 points in Part I.

Testing Differences Between Means and Proportions

**Page 354, Problem 1**

A sample of 40 observations is selected from one population with a population standard deviation of 5. The sample mean is 102. A sample of 50 observations is selected from a second population with a population standard deviation of 6. The sample mean is 99. Conduct the following test of hypothesis using the .04 significance level.

H0 : μ1 = μ2

H1 : μ1 ≠ μ2

**Answer:**

1. Is this a one-tailed or a two-tailed test?

This is a **two-tailed** test.

1. State the decision rule.

To find out the critical area for z value:

0.5000-0.04/2 = 0.4800

Then look up z table to find the z critical area value.

zα = ± 2.06

Thus, if z is positive, when z > 2.06, reject H0 or if z is negative, when z < -2.06, reject H0.

1. Compute the value of the test statistic.

Z = = = 2.5867

1. What is your decision regarding?

Since 2.5867 > 2.06, reject H0.

1. What is the p-value?

p-value = (0.5000 - 0.4952) \* 2 = 0.0096

**Page 355, Problem 4, 5**

4. As part of a study of corporate employees, the director of human resources for PNC Inc. wants to compare the distance traveled to work by employees at its office in downtown Cincinnati with the distance for those in downtown Pittsburgh. A sample of 35 Cincinnati employees showed they travel a mean of 370 miles per month. A sample of 40 Pittsburgh employees showed they travel a mean of 380 miles per month. The population standard deviation for the Cincinnati and Pittsburgh employees are 30 and 26 miles, respectively. At the .05 significance level, is there a difference in the mean number of miles traveled per month between Cincinnati and Pittsburgh employees?

**Answer:**

Step 1: state the null hypothesis and the alternate hypothesis

H0 : μc = μp

H1 : μc ≠ μp

Step 2: Select the level of significance

0.05

Step 3: Determine the test statistic

Since we assume the two population distributions are both normal and their standard deviations are known, we use z score as the test statistic.

Step 4: Formulate a decision rule

This is a two-tailed test, and z score is used as the test statistic; the significance level is 0.05.

To find out the critical area for z value:

0.5000-0.05/2= 0.4750

Then look up z table to find the z critical area value.

zα = ± 1.96

If the test statistic value (z) is positive, when z < -1.96 or if it’s negative, when z >+1.96, reject H0.

Step 5: Make the decision regarding H0.

Z = = = 1.5319

Since 1.5319 < +1.96, fail to reject H0.

Step 6: Interpret the result

There is no statistically significant difference in the mean number of miles traveled per month between Cincinnati and Pittsburgh employees.

The p-value = (0.5000-0.4370)\*2 = 0.126

5. Women’s height is a suspected factor for difficult deliveries, that is, shorter women are more likely to have Caesarean sections. A medical researcher found in a sample of 45 women who had a normal delivery that their mean height was 61.4 inches. A second sample of 39 women who had a Caesarean section had a mean height of 60.6 inches. Assume that the population of heights of normal deliveries has a population standard deviation of 1.2 inches. Also assume that the heights of the population of women who had Caesarean section births has a standard deviation of 1.1 inches. Are those who had a Caesarean section shorter? Use the .05 significance level. Find the p-value and explain what it means.

**Answer:**

Step 1: state the null hypothesis and the alternate hypothesis

H0 : μn ≤ μc  women who had c-section are not shorter

H1 : μn > μc  women who had c-section are shorter

Step 2: Select the level of significance

0.05

Step 3: Determine the test statistic

Since we assume the two population distributions are both normal and their standard deviations are known, we use z score as the test statistic.

Step 4: Formulate a decision rule

This is a one-tailed test, and z score is used as the test statistic; the significance level is 0.05.

To find out the critical area for z value:

0.5000-0.05= 0.4500

Then look up z table to find the z critical area value.

zα = ± 1.645

If the test statistic value (z) is negative, when z < -1.645 or if it’s positive, when z >+1.645, reject H0.

Step 5: Make the decision regarding H0.

Z = = = 3.1866

Since 3.1866 > +1.645, reject H0.

Step 6: Interpret the result

It is confident to say that women who had C-section are shorter.

The p-value < (0.5000-0.4993) = 0.0007, meaning the chance for the difference in the sample means that was caused by sampling errors is less than 0.001.

**Page 542, Problems 11 and 12**

11. A nationwide sample of influential Republicans and Democrats was asked as a part of a comprehensive survey whether they favored lowering environmental standards so that high-sulfur coal could be burned in coal-fired power plants. The results were:

Republicans Democrats

Number sampled 1,000 800

Number in favor 200 168

At the .02 level of significance, can we conclude that there is a larger proportion of Democrats in favor of lowering the standards? Determine the p-value.

**Answer:**

Step 1: state the null hypothesis and the alternate hypothesis

H0 : πd ≤ πr  proportion of democrats in favor of lowering the standards **is not larger**

H1 : πd > πr  proportion of democrats in favor of lowering the standards **is larger**

Step 2: Select the level of significance

0.02

Step 3: Determine the test statistic

The test statistic follows the standard normal distribution.

Step 4: Formulate a decision rule

This is a one-tailed test for two proportions, and z score is used as the test statistic; the significance level is 0.02.

The corresponding z critical area value = z(0.5-0.02) = z(0.48) = 2.06

If the test statistic value (z) is positive, when z < -2.06 or if it’s negative, when z >+2.06, reject H0.

Step 5: Make the decision regarding H0.

Pr:0.2

Pd:0.21

Pc = 0.20444

Z = = = -0.5227

Since -0.5227 > -2.06, **fail to reject H0**.

Step 6: Interpret the result

It is not safe to conclude that the proportion of democrats in favor of lowering the standards is larger.

p-value = 0.5-0.1985 = 0.3015

12. The research department at the home office of New Hampshire Insurance conducts ongoing research on the causes of automobile accidents, the characteristics of the drivers,

and so on. A random sample of 400 policies written on single persons revealed 120 had at least one accident in the previous three-year period. Similarly, a sample of 600 policies written on married persons revealed that 150 had been in at least one accident. At the .05 significance level, is there a significant difference in the proportions of single and married persons having an accident during a three-year period? Determine the p-value.

**Answer:**

Step 1: state the null hypothesis and the alternate hypothesis

H0 : πs = πm  there is no difference in the proportions of single and married persons

H1 : πd ≠ πr  there is difference in the proportions of single and married persons

Step 2: Select the level of significance

0.05

Step 3: Determine the test statistic

The test statistic follows the standard normal distribution.

Step 4: Formulate a decision rule

This is a two-tailed test for two proportions, and z score is used as the test statistic; the significance level is 0.05.

The corresponding z critical area value = z(0.5-0.05/2) = z(0.475) = 1.96

If the test statistic value (z) is positive, when z < -1.96 or if it’s negative, when z > +1.96, reject H0.

Step 5: Make the decision regarding H0.

Ps:0.3

Pm:0.25

Pc = 0.27

Z = = = 1.7447

Since 1.7447 < 1.96, fail to reject H0.

Step 6: Interpret the result

We cannot say that there is significant difference in the proportions of single and married persons having an accident during a three-year period.

p-value = (0.5-0.4591)\*2 = 0.0818

**Page 359, Problem 8,**

The null and alternate hypotheses are:

H0 : μ1 = μ2

H1 : μ1 ≠ μ2

A random sample of 15 observations from the first population revealed a sample mean of 350 and a sample standard deviation of 12. A random sample of 17 observations from

the second population revealed a sample mean of 342 and a sample standard deviation of 15. At the .10 significance level, is there a difference in the population means?

**Answer:**

(a). State the decision rule

df = 30, α = 0.10, two-tailed test, α/2 = 0.05, tα = 1.697

if t < -1.697 or t > 1.697, reject H0

(b). Compute the pooled estimate of the population variance

Using formula 11-3 on page 356,

S2p = = = 187.2

(c). Compute the test statistic

t = = = 1.6506

(d). State your decision about the null hypothesis

since 1.6506 < 1.697, fail to reject H0

(e). Estimate the p-value

p-value estimation: [0.10, 0.20]

**Page 360. Problems 10, and 11**

10. A recent study compared the time spent together by single- and dual-earner couples. According to the records kept by the wives during the study, the mean amount of time spent together watching television among the single-earner couples was 61 minutes per day, with a standard deviation of 15.5 minutes. For the dual-earner couples, the mean number of minutes spent watching television was 48.4 minutes, with a standard deviation of 18.1 minutes. At the .01 significance level, can we conclude that the single-earner couples on average spend more time watching television together? There were 15 single earner and 12 dual-earner couples studied.

**Answer:**

Step 1: state the null hypothesis and the alternate hypothesis

H0 : μs ≤ μd  single-earner couples time spend by single-earner couples watching tv together is less than or equal to that of dual-earner couples.

H1 : μs > μd  single-earner couples time spend by single-earner couples watching tv together is more than that of dual-earner couples.

Step 2: Select the level of significance

0.01

Step 3: Determine the test statistic

The population standard deviations are unknown and we assume the means follow normal distribution, so we use t-value

Step 4: Formulate a decision rule

df = 25, α = 0.01, one-tailed test, tα = 2.485

if t < -2.485 or t > 2.485, reject H0

Step 5: Make the decision regarding H0.

Pooled estimate of the population variance:

S2p = = = 278.6884

t = = = 1.9488

Since 1.9488 < 12.485, fail to reject H0.

Step 6: Interpret the result

At 0.01 significance level, we cannot say that single-earner couples spend more time watching tv together.

p-value is estimated between 0.025 and 0.05

11. Ms. Lisa Monnin is the budget director for Nexus Media Inc. She would like to compare the daily travel expenses for the sales staff and the audit staff. She collected the following sample information.

Sales ($) 131 135 146 165 136 142

Audit ($) 130 102 129 143 149 120 139

At the .10 significance level, can she conclude that the mean daily expenses are greater for the sales staff than the audit staff? What is the p-value?

**Answer:**

Step 1: state the null hypothesis and the alternate hypothesis

H0 : μs ≤ μa mean daily expenses are not greater for the sales staff

H1 : μs > μa mean daily expenses are greater for the sales staff

Step 2: Select the level of significance

0.10

Step 3: Determine the test statistic

The population standard deviations are unknown and we assume the means follow normal distribution, so we use t-value

Step 4: Formulate a decision rule

df = 11, α = 0.10, one-tailed test, tα = 1.363

if t < -1.363 or t > 1.363, reject H0

Step 5: Make the decision regarding H0.

Pooled estimate of the population variance:

= = 142.5

= = 130.2857

*==*149.9010

*==*249.2390

S2p = = = 204.0852

t = = = 1.5368

Since 1.5368 > 1.363, reject H0.

Step 6: Interpret the result

At 0.1 significance level, Ms. Lisa Monnin can conclude that sales staff’s mean daily expenses are greater than those of audit staff.

p-value is estimated between 0.05 and 0.1

**Page 364, Problem 15**

A recent survey compared the costs of adoption through public and private agencies. For a sample of 16 adoptions through public agency, the mean cost was $ 21,045, with a standard deviation of $835. For a sample of 18 adoptions through private agency, the mean cost was $22,840, with a standard deviation of $1,545. Can we conclude the mean cost is larger for adopting children through a private agency/ use the 0.5 significance level?

Answer:

H0: μprivate ≤ μpublic

H1: μprivate > μpublic

df =26.7

t = = = - 4.2764

df = = = 26.7362 (~26 )

(Conservative estimate is taken as 26)

α = 0.05, one-tailed test, tα = 1.706

if t < -1.706 or t > 1.706, reject H0

Since -4.276 < -1.706, reject H0

We cannot conclude that the mean cost is larger for adopting children through a private agency.

**Page 370, Problem 18**

The null and alternate hypotheses are:

H0 : μd = 0

H1 : μd ≠ 0

The following paired observations show the number of traffic citations given for speeding by Officer Dhondt and Officer Meredith of the South Carolina Highway Patrol for the last five months.

|  |
| --- |
| Number of Citations Used  May June July Aug Sept  Officer Dhondt 30 22 25 19 26  Officer Meredith 26 19 20 15 19 |

At the .05 significance level, is there a difference in the mean number of citations given by the two officers?

**Answer:**

Degree of freedom: 10-1 = 9

Sample size: 5

Sum of difference: (4 + 3 + 5 + 4 + 7) = 23

Sample average difference (): = = 4.6

Standard deviation of the difference (Sd) := = 1.51658

t-score:= = = 6.7823

df = 9, α= 0.05, two-tailed test, tα/2 = 2.262

if t < -2.262 or t > 2.262, reject H0

Since 6.782 > 2.262, reject H0

At the 0.05 significance level, there is a difference in the mean number of citations given by the two officers.

**Page 371, Problem 20**

The federal government recently granted funds for a special program designed to reduce crime in high-crime areas. A study of the results of the program in eight high-crime areas of Miami, Florida, yielded the following results.

|  |
| --- |
| Number of Crimes by Area  A B C D E F G H  Before 14 7 4 5 17 12 8 9  After 2 7 3 6 8 13 3 5 |

Has there been a decrease in the number of crimes since the inauguration of the program? Use the .01 significance level. Estimate the p-value.

**Answer:**

Step 1: state the null hypothesis and the alternate hypothesis

H0 : μd ≥ 0 there has been no decrease in the number of crimes since the inauguration of the program

H1 : μd < 0there has been a decrease in the number of crimes since the inauguration of the program

Step 2: Select the level of significance

0.01

Step 3: Determine the test statistic

This is a test for two dependent samples, so we use paired t-value

Step 4: Formulate a decision rule

Sum of difference: (12 + 0 + 1 + -1 + 9 + -1 + 5 + 4) = 29

Sample average difference (): = = 3.625

Standard deviation of the difference(Sd) := = 4.83846

df = 7, α = 0.010 one-tailed test, tα = 2.998

if t < -2.998 or t > 2.998, reject H0

Step 5: Make the decision regarding H0.

t = = = 2.11907

Since 2.119 < 2.998, reject H0.

Step 6: Interpret the result

At 0.01 significance level, there has been a decrease in the number of crimes since the inauguration of the program.

p-value is estimated between 0.025 and 0.05

# Part II: The following problems require you to use SAS. The data sets are in Canvas. Make sure you interpret all relevant results including P-values. There is a total of 28 points for Part II.

1. Computer Data Exercise Problem 47 (parts a-c) on pages 378 of the Lind, Marchal, and Wathen textbook. Additionally, do part d below. [8 pts, two for each part]

Refer to the Real Estate data, which report information on the homes sold in Goodyear, Arizona, last year.

TITLE "V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**DATA** V506.temp;

SET V506.RealEstate;

**RUN**;

/\* Answer 1(a)\*/

**PROC** **TTEST** DATA=V506.temp PLOTS=none ALPHA=**0.05** SIDES=**2**;

CLASS Pool;

VAR Price;

**RUN**;

/\* Answer 1(b)\*/

**PROC** **TTEST** DATA=V506.temp PLOTS=none ALPHA=**0.05** SIDES=**2**;

CLASS Garage;

VAR Price;

**RUN**;

/\* Answer 1(c)\*/

**DATA** V506.temp2;

SET V506.temp;

IF Twnship=**1** OR Twnship=**2**;

**RUN**;

**PROC** **TTEST** DATA=V506.temp2 PLOTS=none ALPHA=**0.05** SIDES=**2**;

CLASS Twnship;

VAR Price;

**RUN**;

/\* Answer 1(d)\*/

**PROC** **MEANS** DATA=v506.temp MEDIAN;

VAR Price;

**RUN**;

**DATA** v506.temp3;

SET v506.temp;

IF PRICE>=**213.5699959** THEN dPrice=**1**;

ELSE dPrice=**0**;

**RUN**;

**PROC** **FREQ** DATA=V506.temp3;

WHERE dPrice = **1**;

TABLES Pool/BINOMIAL(LEVEL=**2**) ALPHA=**.05**;

**RUN**;

**PROC** **FREQ** DATA=V506.temp3;

WHERE dPrice = **0**;

TABLES Pool/BINOMIAL(LEVEL=**2**) ALPHA=**.05**;

**RUN**;

1. At the .05 significance level, can we conclude that there is a difference in the mean selling price of homes with a pool and homes without a pool?

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: Price (Price)**

| **Pool** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 67 | 231.5 | 50.5750 | 6.1787 | 125.0 | 345.3 |
| **1** | 38 | 202.8 | 33.7066 | 5.4679 | 147.4 | 307.8 |
| **Diff (1-2)** |  | 28.6897 | 45.2452 | 9.1884 |  |  |

| **Pool** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** |  | 231.5 | 219.1 | 243.8 | 50.5750 | 43.2262 | 60.9579 |
| **1** |  | 202.8 | 191.7 | 213.9 | 33.7066 | 27.4798 | 43.6078 |
| **Diff (1-2)** | **Pooled** | 28.6897 | 10.4668 | 46.9126 | 45.2452 | 39.8205 | 52.3945 |
| **Diff (1-2)** | **Satterthwaite** | 28.6897 | 12.3209 | 45.0585 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 103 | 3.12 | 0.0023 |
| **Satterthwaite** | Unequal | 100.22 | 3.48 | 0.0008 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 66 | 37 | 2.25 | 0.0088 |

Yes, we conclude that there is a difference in the mean selling price of homes with a pool and homes without a pool.

1. At the .05 significance level, can we conclude that there is a difference in the mean selling price of homes with an attached garage and homes without an attached garage?

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: Price (Price)**

| **Garage** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 34 | 185.4 | 28.0053 | 4.8029 | 125.0 | 242.1 |
| **1** | 71 | 238.2 | 44.8815 | 5.3264 | 166.2 | 345.3 |
| **Diff (1-2)** |  | -52.7310 | 40.2523 | 8.3949 |  |  |

| **Garage** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** |  | 185.4 | 175.7 | 195.2 | 28.0053 | 22.5884 | 36.8627 |
| **1** |  | 238.2 | 227.6 | 248.8 | 44.8815 | 38.5213 | 53.7768 |
| **Diff (1-2)** | **Pooled** | -52.7310 | -69.3804 | -36.0817 | 40.2523 | 35.4263 | 46.6128 |
| **Diff (1-2)** | **Satterthwaite** | -52.7310 | -66.9679 | -38.4942 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 103 | -6.28 | <.0001 |
| **Satterthwaite** | Unequal | 95.786 | -7.35 | <.0001 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 70 | 33 | 2.57 | 0.0037 |

Yes, we conclude that there is a difference in the mean selling price of homes with an attached garage and homes without an attached garage

1. At the .05 significance level, can we conclude that there is a difference in the mean selling price of homes in Township 1 and Township 2?

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: Price (Price)**

| **Twnship** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | 15 | 196.9 | 35.7891 | 9.2407 | 125.9 | 245.4 |
| **2** | 20 | 227.4 | 44.1970 | 9.8827 | 154.3 | 307.8 |
| **Diff (1-2)** |  | -30.5312 | 40.8420 | 13.9502 |  |  |

| **Twnship** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1** |  | 196.9 | 177.1 | 216.7 | 35.7891 | 26.2022 | 56.4430 |
| **2** |  | 227.4 | 206.8 | 248.1 | 44.1970 | 33.6114 | 64.5529 |
| **Diff (1-2)** | **Pooled** | -30.5312 | -58.9130 | -2.1493 | 40.8420 | 32.9422 | 53.7594 |
| **Diff (1-2)** | **Satterthwaite** | -30.5312 | -58.0657 | -2.9966 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 33 | -2.19 | 0.0358 |
| **Satterthwaite** | Unequal | 32.761 | -2.26 | 0.0308 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 19 | 14 | 1.53 | 0.4252 |

Yes, we conclude that there is a difference in the mean selling price of homes in Township 1 and Township 2.

1. Find the median selling price of the homes. Divide the homes into two groups, those that sold for more than (or equal to) the median price and those that sold for less. Is there a difference in the proportion of homes with a pool for those that sold at or above the median price versus those that sold for less than the median price? Use the 0.05 significance level.

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The MEANS Procedure**

| **Analysis Variable : Price Price** |
| --- |
| **Median** |
| 213.5699959 |

The median selling price of the homes is **213.5699959.**

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The FREQ Procedure**

| **Pool** | | | | |
| --- | --- | --- | --- | --- |
| **Pool** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **0** | 42 | 80.77 | 42 | 80.77 |
| **1** | 10 | 19.23 | 52 | 100.00 |

| **Binomial Proportion** | |
| --- | --- |
| **Pool = 1** | |
| **Proportion** | 0.1923 |
| **ASE** | 0.0547 |
| **95% Lower Conf Limit** | 0.0852 |
| **95% Upper Conf Limit** | 0.2994 |
|  |  |
| **Exact Conf Limits** |  |
| **95% Lower Conf Limit** | 0.0963 |
| **95% Upper Conf Limit** | 0.3253 |

| **Test of H0: Proportion = 0.5** | |
| --- | --- |
| **ASE under H0** | 0.0693 |
| **Z** | -4.4376 |
| **One-sided Pr < Z** | <.0001 |
| **Two-sided Pr > |Z|** | <.0001 |

|  |
| --- |
| **Sample Size = 52** |

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The FREQ Procedure**

| **Pool** | | | | |
| --- | --- | --- | --- | --- |
| **Pool** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **0** | 25 | 47.17 | 25 | 47.17 |
| **1** | 28 | 52.83 | 53 | 100.00 |

| **Binomial Proportion** | |
| --- | --- |
| **Pool = 1** | |
| **Proportion** | 0.5283 |
| **ASE** | 0.0686 |
| **95% Lower Conf Limit** | 0.3939 |
| **95% Upper Conf Limit** | 0.6627 |
|  |  |
| **Exact Conf Limits** |  |
| **95% Lower Conf Limit** | 0.3864 |
| **95% Upper Conf Limit** | 0.6670 |

| **Test of H0: Proportion = 0.5** | |
| --- | --- |
| **ASE under H0** | 0.0687 |
| **Z** | 0.4121 |
| **One-sided Pr > Z** | 0.3401 |
| **Two-sided Pr > |Z|** | 0.6803 |

|  |
| --- |
| **Sample Size = 53** |

Yes, there is a difference in the proportion of homes with a pool for those that sold at or above the median price versus those that sold for less than the median price .

2. Refer to the country data in the International-2005.xlsx file, which reports information on demographic and economic characteristics for 46 countries from the CIA International Data Book. (Documentation for this data is in the excel file under the second tab.)

TITLE "V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**DATA** V506.ctemp;

SET V506.CiaData;

**RUN**;

/\* Answer 2(a)\*/

**PROC** **TTEST** DATA=V506.ctemp PLOTS=none ALPHA=**0.05** SIDES=**2**;

CLASS G20;

VAR LitRate;

**RUN**;

/\* Answer 2(b)\*/

**DATA** v506.ctemp2;

SET v506.ctemp;

IF GDPpCap>=**20** THEN WCountry=**1**;

ELSE WCountry=**0**;

**RUN**;

**PROC** **TTEST** DATA=V506.ctemp2 PLOTS=none ALPHA=**0.01** SIDES=**2**;

CLASS WCountry;

VAR LifeExpc;

**RUN**;

1. Conduct a test of hypothesis to determine whether the literacy rate in G20 countries is different from those that are not G20 countries. Use a .05 significance level. [3 pts]

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: LitRate (LitRate)**

| **G20** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 27 | 89.6370 | 14.0683 | 2.7074 | 57.1000 | 100.0 |
| **1** | 19 | 89.4368 | 13.3129 | 3.0542 | 52.0000 | 100.0 |
| **Diff (1-2)** |  | 0.2002 | 13.7643 | 4.1217 |  |  |

| **G20** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** |  | 89.6370 | 84.0718 | 95.2023 | 14.0683 | 11.0790 | 19.2796 |
| **1** |  | 89.4368 | 83.0202 | 95.8535 | 13.3129 | 10.0594 | 19.6875 |
| **Diff (1-2)** | **Pooled** | 0.2002 | -8.1065 | 8.5069 | 13.7643 | 11.3948 | 17.3870 |
| **Diff (1-2)** | **Satterthwaite** | 0.2002 | -8.0474 | 8.4478 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 44 | 0.05 | 0.9615 |
| **Satterthwaite** | Unequal | 40.213 | 0.05 | 0.9611 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 26 | 18 | 1.12 | 0.8225 |

Yes, the literacy rate in G20 countries is different from those that are not G20 countries.

1. The United Nations wants to see if there is a relationship between Gross Domestic Product per capita and life expectancy. They feel that wealthy countries will also have high average life expectancy. The UN has defined wealthy countries as those with GDP per capita of 20 thousand dollars or more. Test this hypothesis at a .01 significance level. [3 pts]

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: LifeExpc (LifeExpc)**

| **WCountry** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 25 | 70.2648 | 7.6023 | 1.5205 | 48.0900 | 78.9300 |
| **1** | 21 | 78.0210 | 1.8934 | 0.4132 | 72.6200 | 80.8000 |
| **Diff (1-2)** |  | -7.7562 | 5.7579 | 1.7044 |  |  |

| **WCountry** | **Method** | **Mean** | **99% CL Mean** | | **Std Dev** | **99% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** |  | 70.2648 | 66.0122 | 74.5174 | 7.6023 | 5.5178 | 11.8450 |
| **1** |  | 78.0210 | 76.8453 | 79.1966 | 1.8934 | 1.3389 | 3.1057 |
| **Diff (1-2)** | **Pooled** | -7.7562 | -12.3448 | -3.1675 | 5.7579 | 4.5045 | 7.8648 |
| **Diff (1-2)** | **Satterthwaite** | -7.7562 | -12.1157 | -3.3966 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 44 | -4.55 | <.0001 |
| **Satterthwaite** | Unequal | 27.496 | -4.92 | <.0001 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 24 | 20 | 16.12 | <.0001 |

As we can see from the above computations the countries with higher Gross Domestic Product per capita also have a higher literacy rate and hence we can conclude that there is a relationship between Gross Domestic Product per capita and life expectancy.

3. Refer to the Schools.xls data set, which contains a sample of 94 of the 600+ school districts in Ohio. (Documentation for this data set is now under the documentation tab of the spreadsheet in Canvas.) Divide the school districts into two groups. Include all schools with less than 2,000 students (small districts) in one group and those with 2,000 or more (large districts) in the other.

TITLE "V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**DATA** V506.Stemp;

SET V506.Schools;

IF Students>=**2000** THEN SGroup=**1**;

ELSE SGroup=**0**;

**RUN**;

/\* Answer 3(a)\*/

**PROC** **TTEST** DATA=V506.Stemp PLOTS=none ALPHA=**0.05** SIDES=L;

CLASS SGroup;

VAR Salary;

**RUN**;

/\* Answer 3(b)\*/

**PROC** **TTEST** DATA=V506.Stemp PLOTS=none ALPHA=**0.05** SIDES=**2**;

CLASS SGroup;

VAR Instruct;

**RUN**;

/\* Answer 3(c)\*/

**PROC** **TTEST** DATA=V506.Stemp PLOTS=none ALPHA=**0.05** SIDES=**2**;

CLASS SGroup;

VAR Attend;

**RUN**;

1. Compute the mean teacher salary for the two groups. At the .05 significance level, can we conclude that the mean teacher salary is higher in the larger school districts? [2 pts]

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: Salary (Salary)**

| **SGroup** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 64 | 31772.8 | 2851.4 | 356.4 | 26125.0 | 43256.0 |
| **1** | 30 | 36186.4 | 2999.8 | 547.7 | 30907.0 | 42734.0 |
| **Diff (1-2)** |  | -4413.5 | 2899.0 | 641.4 |  |  |

| **SGroup** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** |  | 31772.8 | 31060.6 | 32485.1 | 2851.4 | 2428.8 | 3453.4 |
| **1** |  | 36186.4 | 35066.2 | 37306.5 | 2999.8 | 2389.0 | 4032.6 |
| **Diff (1-2)** | **Pooled** | -4413.5 | -Infty | -3347.7 | 2899.0 | 2533.8 | 3388.1 |
| **Diff (1-2)** | **Satterthwaite** | -4413.5 | -Infty | -3320.1 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr < t** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 92 | -6.88 | <.0001 |
| **Satterthwaite** | Unequal | 54.284 | -6.75 | <.0001 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 29 | 63 | 1.11 | 0.7196 |

Yes, we can conclude that the mean teacher salary is higher in the larger school districts

1. Compute the mean amount spent per pupil for the large and the small districts. At the .05 significance level, can we conclude that there is a difference in the mean amount spent? [2 pts]

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: Instruct (Instruct)**

| **SGroup** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 64 | 2718.2 | 1307.9 | 163.5 | 1916.0 | 11226.0 |
| **1** | 30 | 2738.3 | 361.1 | 65.9219 | 2250.0 | 3933.0 |
| **Diff (1-2)** |  | -20.1125 | 1101.1 | 243.6 |  |  |

| **SGroup** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** |  | 2718.2 | 2391.5 | 3044.9 | 1307.9 | 1114.0 | 1584.0 |
| **1** |  | 2738.3 | 2603.5 | 2873.1 | 361.1 | 287.6 | 485.4 |
| **Diff (1-2)** | **Pooled** | -20.1125 | -504.0 | 463.8 | 1101.1 | 962.4 | 1286.9 |
| **Diff (1-2)** | **Satterthwaite** | -20.1125 | -370.9 | 330.7 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 92 | -0.08 | 0.9344 |
| **Satterthwaite** | Unequal | 80.527 | -0.11 | 0.9094 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 63 | 29 | 13.12 | <.0001 |

Yes, we can conclude that there is a difference in the mean amount spent per pupil for the large and the small districts.

1. Compute the mean daily percent of students attending for the large and small districts. At the .05 significance level, can we conclude that there is a difference in the mean daily attendance? [2 pts]

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: Attend (Attend)**

| **SGroup** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 64 | 95.6688 | 0.8483 | 0.1060 | 93.9000 | 99.8000 |
| **1** | 30 | 94.4233 | 1.1181 | 0.2041 | 90.7000 | 96.1000 |
| **Diff (1-2)** |  | 1.2454 | 0.9417 | 0.2084 |  |  |

| **SGroup** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** |  | 95.6688 | 95.4568 | 95.8807 | 0.8483 | 0.7226 | 1.0274 |
| **1** |  | 94.4233 | 94.0058 | 94.8408 | 1.1181 | 0.8905 | 1.5031 |
| **Diff (1-2)** | **Pooled** | 1.2454 | 0.8316 | 1.6593 | 0.9417 | 0.8231 | 1.1006 |
| **Diff (1-2)** | **Satterthwaite** | 1.2454 | 0.7822 | 1.7087 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 92 | 5.98 | <.0001 |
| **Satterthwaite** | Unequal | 45.246 | 5.41 | <.0001 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 29 | 63 | 1.74 | 0.0688 |

Yes, we can conclude that there is a difference in the mean daily attendance of students attending for the large and small districts.

4. Using data from the Baltimore Longitudinal Study of Aging (described on the handout passed out for the last exercise):

TITLE "V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**DATA** V506.Btemp;

SET V506.Blsa2;

**RUN**;

/\* Answer 4(a)\*/

**PROC** **TTEST** DATA=V506.Btemp PLOTS=none ALPHA=**0.01** SIDES=**2**;

CLASS SEX;

VAR DBP;

**RUN**;

/\* Answer 4(b)\*/

**PROC** **TTEST** DATA=V506.Btemp PLOTS=none ALPHA=**0.05** SIDES=**2**;

WHERE SEX ='F';

CLASS Smoker;

VAR weight;

**RUN**;

1. At the .01 significance level, can we conclude that the mean diastolic blood pressure (dbp) of males is different from the mean dbp of females? [3 pts]

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: DBP**

| **sex** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **F** | 359 | 76.1922 | 10.4491 | 0.5515 | 40.0000 | 118.0 |
| **M** | 360 | 79.9944 | 10.9252 | 0.5758 | 56.0000 | 130.0 |
| **Diff (1-2)** |  | -3.8022 | 10.6901 | 0.7973 |  |  |

| **sex** | **Method** | **Mean** | **99% CL Mean** | | **Std Dev** | **99% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **F** |  | 76.1922 | 74.7641 | 77.6203 | 10.4491 | 9.5268 | 11.5534 |
| **M** |  | 79.9944 | 78.5033 | 81.4856 | 10.9252 | 9.9620 | 12.0780 |
| **Diff (1-2)** | **Pooled** | -3.8022 | -5.8616 | -1.7429 | 10.6901 | 10.0066 | 11.4663 |
| **Diff (1-2)** | **Satterthwaite** | -3.8022 | -5.8614 | -1.7430 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 717 | -4.77 | <.0001 |
| **Satterthwaite** | Unequal | 715.75 | -4.77 | <.0001 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 359 | 358 | 1.09 | 0.3993 |

Yes, we can conclude that the mean diastolic blood pressure (dbp) of males is different from the mean dbp of females

1. At the .05 significance level, can we conclude that the mean weight of females who smoke is different from the mean weight of females who do not smoke? [3 pts]

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: weight**

| **smoker** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **N** | 278 | 62.6802 | 10.6509 | 0.6388 | 40.8000 | 111.7 |
| **Y** | 81 | 60.9432 | 10.7943 | 1.1994 | 44.2000 | 107.2 |
| **Diff (1-2)** |  | 1.7370 | 10.6832 | 1.3489 |  |  |

| **smoker** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **N** |  | 62.6802 | 61.4227 | 63.9377 | 10.6509 | 9.8331 | 11.6184 |
| **Y** |  | 60.9432 | 58.5564 | 63.3300 | 10.7943 | 9.3498 | 12.7709 |
| **Diff (1-2)** | **Pooled** | 1.7370 | -0.9158 | 4.3898 | 10.6832 | 9.9538 | 11.5289 |
| **Diff (1-2)** | **Satterthwaite** | 1.7370 | -0.9516 | 4.4256 |  |  |  |

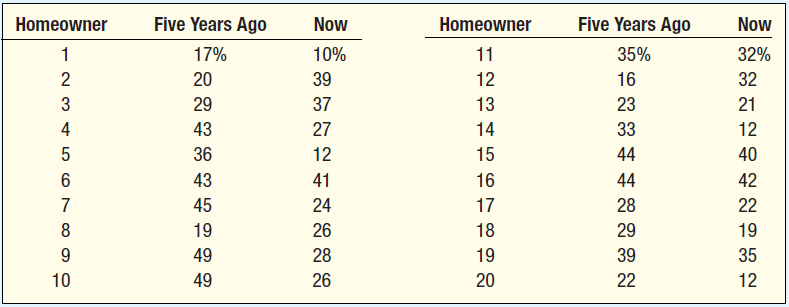
| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 357 | 1.29 | 0.1987 |
| **Satterthwaite** | Unequal | 128.83 | 1.28 | 0.2035 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 80 | 277 | 1.03 | 0.8555 |

Yes, we conclude that the mean weight of females who smoke is different from the mean weight of females who do not smoke

5. Problem 44 on page 376 of the Lind, Marchal, and Wathen textbook. Use the NewEnglandHomes.csv data set. [2 pts]

The amount of income spent on housing is an important component of the cost of living. The total costs of housing for homeowners might include mortgage payments, property taxes, and utility costs (water, heat, electricity). An economist selected a sample of 20 homeowners in New England and then calculated these total housing costs as a percent of monthly income, five years ago and now. The information is reported below. Is it reasonable to conclude the percent is less now than five years ago?



TITLE "V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

/\* Answer 5\*/

**PROC** **TTEST** DATA=V506.NEHomes PLOTS=none;

VAR FiveYearsAgo Now;

**RUN**;

|  |
| --- |
| **V506 HOMEWORK04 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The TTEST Procedure**

**Variable: FiveYearsAgo (FiveYearsAgo)**

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 20 | 33.1500 | 11.0847 | 2.4786 | 16.0000 | 49.0000 |

| **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 33.1500 | 27.9622 | 38.3378 | 11.0847 | 8.4298 | 16.1900 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 19 | 13.37 | <.0001 |

|  |
| --- |
| **Variable: Now (Now)** |

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 20 | 26.8500 | 10.3836 | 2.3218 | 10.0000 | 42.0000 |

| **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 26.8500 | 26.8500 | 21.9903 | 31.7097 | 10.3836 | 7.8966 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 19 | 11.56 | <.0001 |

Yes, we can conclude the total housing costs as a percent of monthly income is less now than five years ago.