V506 Homework Exercise 3

***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 1.5 points except where noted.

**Point and Interval Estimation**

**Page 288, Problem 2 [1 pt]**

A sample of 81 observations is taken from a normal population with a standard deviation of 5. The sample mean is 40. Determine the 95 percent confidence interval for the population mean.

**Answer**:

Confidence Intervals for a Mean (µ) when σ Known =

= 40

z value for 95 percent confidence interval = 1.96

= 5

= 81

Here, n>30

Thus, CI = = =

Thus, 38.9111 < µ < 41.0889

**Page 289, Problem 5 [1 pt]**

A research firm conducted a survey to determine the mean amount steady smokers spend on cigarettes during a week. They found the distribution of amounts spent per week followed the normal distribution with a population standard deviation of $5. A sample of 49 steady smokers revealed that. *X* = $20

a. What is the point estimate of the population mean? Explain what it indicates.

b. Using the 95 percent level of confidence, determine the confidence interval for µ. Explain what it indicates.

**Answer**:

1. The point estimate of population mean (µ) is the sample mean (, it tells us how close our estimate is to the population parameter.
2. Confidence Intervals for a Mean (µ) when σ Known =

= 20

z value for 95 percent confidence interval = 1.96

= 5

= 49

Here, n>30

Thus , CI = = =

Thus, 18.6 < µ < 21.4

This shows that the population mean lies between the interval 18.6 and 21.4. Thus we can conclude that about 95 percent of the intervals constructed in the same manner will have the population mean in their range.

**Page 289, Problem 7 [1 pt]**

Bob Nale is the owner of Nale’s Quick Fill. Bob would like to estimate the mean number of gallons of gasoline sold to his customers. Assume the number of gallons sold follows the normal distribution with a population standard deviation of 2.30 gallons. From his records, he selects a random sample of 60 sales and finds the mean number of gallons sold is 8.60.

a. What is the point estimate of the population mean?

b. Develop a 99 percent confidence interval for the population mean.

c. Interpret the meaning of part (b).

**Answer**:

1. The point estimate of population mean
2. Confidence Intervals for a Mean (µ) when σ Known =

= 8.60

z value for 99 percent confidence interval = 2.58

= 2.30

= 60

Here, n>30

Thus , CI = = =

Thus, 7.8339 < µ < 9.3661

This shows that the population mean lies between the interval 7.8339 and 9.3661. Thus we can conclude that about 99 percent of the intervals constructed in the same manner will have the population mean in their range.

**Page 296, Problem 9**

Use Appendix B.2 to locate the value of t under the following conditions.

a. The sample size is 12 and the level of confidence is 95 percent.

b. The sample size is 20 and the level of confidence is 90 percent.

c. The sample size is 8 and the level of confidence is 99 percent.

**Answer**:

a. The sample size is 12 and the level of confidence is 95 percent. = 2.201

b. The sample size is 20 and the level of confidence is 90 percent. = 1.729

c. The sample size is 8 and the level of confidence is 99 percent.= 3.499

**Page 297, Problems 13 and 14**

Merrill Lynch Securities and Health Care Retirement Inc. are two large employers in downtown Toledo, Ohio. They are considering jointly offering child care for their employees. As a part of the feasibility study, they wish to estimate the mean weekly child-care cost of their employees.

A sample of 10 employees who use child care reveals the following amounts spent last week.

**$107 $92 $97 $95 $105 $101 $91 $99 $95 $104**

Develop a 90 percent confidence interval for the population mean. Interpret the result.

**Answer**:

*Mean () = =* ***98.6***

*Variance* = *=* ***30.7111***

*Standard Deviation (s) = =* ***5.5418***

t value for 99 percent confidence interval = 1.833

Thus, CI = = =

This shows that the population mean lies between the interval 95.3877 and 101.8123. Thus we can conclude that about 99 percent of the intervals constructed in the same manner will have the population mean in their range.

The Greater Pittsburgh Area Chamber of Commerce wants to estimate the mean time workers who are employed in the downtown area spend getting to work. A sample of 15 workers reveals the following number of minutes spent traveling.

**29 38 38 33 38 21 45 34 40 37 37 42 30 29 35**

Develop a 98 percent confidence interval for the population mean. Interpret the result.

**Answer**:

*Mean() = =* ***35.0667***

*Variance* ==***36.2095***

*Standard Deviation (s) ==* ***6.0174***

t value for 98 percent confidence interval = 2.624

Thus, CI = = =

This shows that the population mean lies between the interval 30.9898 and 39.1435

**Problem 16 and 18**

Ms. Maria Wilson is considering running for mayor of the town of Bear Gulch, Montana. Before completing the petitions, she decides to conduct a survey of voters in Bear Gulch.

A sample of 400 voters reveals that 300 would support her in the November election.

a. Estimate the value of the population proportion.

b. Develop a 99 percent confidence interval for the population proportion.

c. Interpret your findings.

**Answer**:

Confidence Intervals for Population proportion =

Population proportion (p) = = 0.75

z value for 99 percent confidence interval = 2.58

CI = =

CI = 0.75 0.0559

Thus, CI will lie between 0.6941 and 0.8059

**Problem 18**

Schadek Silkscreen Printing Inc. purchases plastic cups on which to print logos for sporting events, proms, birthdays, and other special occasions. Zack Schadek, the owner, received a large shipment this morning. To ensure the quality of the shipment, he selected

a random sample of 300 cups. He found 15 to be defective.

a. What is the estimated proportion defective in the population?

b. Develop a 95 percent confidence interval for the proportion defective.

c. Zack has an agreement with his supplier that he is to return lots that are 10 percent or more defective. Should he return this lot? Explain your decision.

**Answer**:

1. The estimated proportion defective in the population is,

* Population proportion (p) = = = 0.05

1. Confidence Intervals for Population proportion =

z value for 95 percent confidence interval = 1.96

CI = =

CI = 0.05 0.02466

Thus, CI will lie between 0.02534 and 0.07466.

1. As we can see the value maximum value (7.5%) is less than 10% we can say that Zack would not have to return this shipment.

**Problem 24**

A processor of carrots cuts the green top off each carrot, washes the carrots, and inserts six to a package. Twenty packages are inserted in a box for shipment. To test the weight of the boxes, a few were checked. The mean weight was 20.4 pounds; the standard deviation, 0.5 pounds. How many boxes must the processor sample to be 95 percent confident that the sample mean does not differ from the population mean by more than 0.2 pounds?

**Answer**:

Given Data:

= 20.4

z value for 95 percent confidence interval = 1.96

= 0.5

= ?

Error in finding endpoints (E) =

Sample size for estimating the population mean (n) =

Thus , (n) = = = 25

Hence , the processor must sample 25 boxes.

**Problem 48**

During a national debate on changes to health care, a cable news service performs an opinion poll of 500 small-business owners. It shows that 65 percent of small-business owners do not approve of the changes. Develop a 95 percent confidence interval for the proportion opposing health care changes. Comment on the result.

**Answer**:

Confidence Intervals for Population proportion =

Population proportion (p) = 0.65

z value for 95 percent confidence interval = 1.96

Number of small business owners = 500

CI = =

CI = 0.65 0.0418

Thus, CI will lie between 0.6082 and 0.6918.

Interval values, (0.6082 x 500) = 304 and (0.6918 x 500) = 346.

As we can see the values, the proportion of small-business owners do not approve of the changes range from 0.6082 to 0.6918.

**Problem 53**

It is estimated that 60 percent of U.S. households subscribe to cable TV. You would like to verify this statement for your class in mass communications. If you want your estimate to be within 5 percentage points, with a 95 percent level of confidence, how large of a sample is required?

**Answer**:

Given Data:

= 0.60

z value for 95 percent confidence interval = 1.96

= 0.05

= ?

Error in finding endpoints (E) =

Sample size for estimating the population mean (n) =

Thus, (n) = = = 368.8

Hence, if we can to verify the above statement we would be requiring a sample size of 369 people.

## **Hypothesis Testing, Population Means, and Proportions**

**Problem 4**

A sample of 64 observations is selected from a normal population. The sample mean is 215, and the population standard deviation is 15. Conduct the following test of hypothesis using the .025 significance level.

H0 : μ ≥ 220

H1 : μ < 220

**Answer:**

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

* This is a one-tailed test, because the alternative hypothesis, H1, states a direction: the mean is less than 220.
* zα = -1.96, for the significance level of 0.025

1. What is the decision rule?

* Reject H0, when z < - zα

1. What is the value of the test statistic?

* z-value = = = -2.67

1. What is your decision regarding H0?

* -2.67 < - 1.96, as a result our null hypothesis condition is satisfied.
* Hence we reject H0

1. What is the p-value?

* p-value is 0.0038; because 0.0038 is less than 0.01, this means that we have extremely strong evidence that H0 is not true. This validates that H0 should be rejected.

**Problem 5**

The manufacturer of the X-15 steel-belted radial truck tire claims that the mean mileage the tire can be driven before the tread wears out is 60,000 miles. Assume the mileage wear follows the normal distribution and the standard deviation of the distribution is 5,000 miles. Crosset Truck Company bought 48 tires and found that the mean mileage for its trucks is 59,500 miles. Is Crosset’s experience different from that claimed by the manufacturer at the .05 significance level?

**Answer**:

1. State the null hypothesis and the alternate hypothesis.

* The Keyword here is ‘different from’, this become a two tail hypothesis

H0 : μ = 60,000

H1 : μ 60,000

1. State the decision rule.

* Reject H0 if |Z| > Zα/2
* In our case the significance level (α) is 0.05, the corresponding z value for α/2 is 1.96
* So, H0 becomes z < - 1.96 or z > 1.96.

1. Compute the value of the test statistic.

* z-value = = = -0.6928

1. What is your decision regarding H?

* -0.6928 > -1.96, as a result our null hypothesis condition is not satisfied.
* Hence we do not reject H0

1. What is the p-value? Interpret it.

* The significance level for our calculated z value (0.6928) = 0.2549
* The p value then becomes = 2(0.5000 – 0.2549) = 0.4902
* If the null hypothesis holds true then the probability of the claim that - mean mileage the tire can be driven before the tread wears out is 60,000 miles, is 0.4902

**Problem 8**

At the time she was hired as a server at the Grumney Family Restaurant, Beth Brigden was told, “You can average $80 a day in tips.” Assume the population of daily tips is normally distributed with a standard deviation of $9.95. Over the first 35 days she was employed at the restaurant, the mean daily amount of her tips was $84.85. At the .01 significance level, can Ms. Brigden conclude that her daily tips average more than $80?

**Answer**:

1. State the null hypothesis and the alternate hypothesis.

* The Keyword here is ‘more than’, this become a one tail hypothesis

H0 : μ ≤ 80

H1 : μ > 80

1. State the decision rule.

* Reject H0 if Z > Zα
* In our case the significance level (α) is 0.01, the corresponding z value is 2.326
* So, H0 becomes z > 2.33.

1. Compute the value of the test statistic.

* z-value = = = 2.8837

1. What is your decision regarding H?

* 2.8837 > 2.33, as a result our null hypothesis condition is satisfied.
* Hence we reject H0

1. What is the p-value? Interpret it.

* The significance level for our calculated z value (2.88) = 0.4980
* The p value is (0.5000 – 0.4980) = 0.002
* If the null hypothesis holds true then the probability of the claim that - the daily tips of Ms. Bridgen average less than $80, is 0.002

**Problem 19**

A Washington, D.C., “think tank” announces the typical teenager sent 50 text messages per day in 2009. To update that estimate, you phone a sample of teenagers and ask them how many text messages they sent the previous day. Their responses were:

**51 175 47 49 44 54 145 203 21 59 42 100**

At the .05 level, can you conclude that the mean number is greater than 50? Estimate the p-value and describe what it tells you.

**Answer**:

Calculation:

Mean (=  *=* ***82.5***

*Variance* ==***3244.4167***

*Standard Deviation (s) ==* ***59.4925***

1. State the null hypothesis and the alternate hypothesis.

* The Keyword here is ‘is greater’, this become a one tail hypothesis

H0 : μ ≤ 50

H1 : μ > 50

1. State the decision rule.

* Reject H0 if t > tα , n-1
* In our case the significance level (α) is 0.05, df is 12-1=11, the corresponding t value is 1.796
* So, H0 becomes t > 1.796.

1. Compute the value of the test statistic.

* t-value = = = 1.8924

1. What is your decision regarding H?

* 1.8924 > 1.796, as a result our null hypothesis condition is satisfied.
* Hence we reject H0

1. What is the p-value? Interpret it.

* The t value (1.89) with df (11) is between 1.796 and 2.201, so the p-value is between 0.05 and 0.025, meaning there is less than 5% chance that a teenager sends less than or equal to 50 texts daily; If the null hypothesis holds true then the probability of the claim that - the mean number is lesser than 50, is less than 0.05.

**Problem 20**

Hugger Polls contends that an agent conducts a mean of 53 in-depth home surveys every week. A streamlined survey form has been introduced, and Hugger wants to evaluate its effectiveness. The number of in-depth surveys conducted during a week by a random sample of agents are:

**53 57 50 55 58 54 60 52 59 62 60 60 51 59 56**

At the .05 level of significance, can we conclude that the mean number of interviews conducted by the agents is more than 53 per week? Estimate the p-value.

**Answer**:

Calculation:

Mean (=  *=* ***56.4***

*Variance* ==***13.9714***

*Standard Deviation (s) = =* ***3.7378***

1. State the null hypothesis and the alternate hypothesis.

* The Keyword here is ‘more than’, this become a one tail hypothesis

H0 : μ ≤ 53

H1 : μ > 53

1. State the decision rule.

* Reject H0 if t > tα , n-1
* In our case the significance level (α) is 0.05, df is 15-1=14, the corresponding t value is 1.761
* So, H0 becomes t > 1.761.

1. Compute the value of the test statistic.

* t-value = = = 3.5229

1. What is your decision regarding H?

* 3.5229 > 1.761, as a result our null hypothesis condition is satisfied.
* Hence we reject H0

1. What is the p-value? Interpret it.

* The p-value is less than 0.005, meaning only 0.5% chance of the number of interviews during a week is less than or equal to 53.

**Problem 23**

According to the local union president, the mean gross income of plumbers in the Salt Lake City area follows the normal probability distribution with a mean of $45,000 and a standard deviation of $3,000. A recent investigative reporter for KYAK TV found, for a sample of 120 plumbers, the mean gross income was $45,500. At the .10 significance level, is it reasonable to conclude that the mean income is not equal to $45,000? Determine the *p*-value.

**Answer**:

1. State the null hypothesis and the alternate hypothesis.

* The Keyword here is ‘not equal’, this become a two tail hypothesis

H0 : μ = 45,000

H1 : μ ≠ 45,000

1. State the decision rule.

* Reject H0 when z is positive and z > zα/2 or when z is negative and z < - zα/2
* In our case the significance level (α) is 0.10, the corresponding z value for α/2 is 1.645

1. Compute the value of the test statistic.

* z-value = = = 1.83

1. What is your decision regarding H0?

* 1.83 > 1.645, as a result our null hypothesis condition is satisfied.
* Hence reject H0
* At the 0.1 level of significance, it is reasonable to conclude that the mean income is not equal to 45,000

1. Determine the p-value

* P-value =2\*(0.5000-0.4664) = 0.0672

**Problem 27**

According to a recent survey, Americans get a mean of 7 hours of sleep per night. A random sample of 50 students at West Virginia University revealed the mean number of hours slept last night was 6 hours and 48 minutes (6.8 hours). The standard deviation of the sample was 0.9 hours. Is it reasonable to conclude that students at West Virginia sleep less than the typical American? Compute the *p*-value.

**Answer**:

1. State the null hypothesis and the alternate hypothesis.

* The Keyword here is ‘less than’, this become a one tail hypothesis

H0 : μ ≥ 7

H1 : μ < 7

1. State the decision rule.

* Reject H0 , when t-value is negative and when t < - tα ,df(n-1)
* tα ,df(n-1) = 1.677

1. Compute the value of the test statistic.

* t-value = = = -1.5713
* since t-value is negative and t > -1.677, fail to reject H0.
* At the 5% level of significance, we cannot safely conclude that students at West Virginia sleep less than the typical American

1. What is the p value?

* If t value is -1.5713 we get the range of the p value to be between 0.05 and 0.10
* Hence we cannot reject H0

**Problem 34**

The postanesthesia care area (recovery room) at St. Luke’s Hospital in Maumee, Ohio, was recently enlarged. The hope was that with the enlargement the mean number of patients per day would be more than 25. A random sample of 15 days revealed the following numbers of patients.

25 27 25 26 25 28 28 27 24 26 25 29 25 27 24

At the .01 significance level, can we conclude that the mean number of patients per day is more than 25? Estimate the p-value and interpret it.

**Answer**:

Calculation:

Mean (=  *=* ***26.0667***

*Variance* ==***2.3524***

*Standard Deviation (s) ==* ***1.5337***

1. State the null hypothesis and the alternate hypothesis.

* The Keyword here is ‘more than’, this become a two tail hypothesis

H0 : μ ≤ 25

H1 : μ > 25

1. State the decision rule.

* When t < -tα , n-1, reject H0;
* In our case the significance level (α) is 0.01, df =15-1=14, the corresponding t value is 2.624

1. Compute the value of the test statistic.

* t-value = = = 2.6937

1. Estimate the p-value:

* P value for the t statistic lies between 0.01 and 0.005

1. What is your decision regarding H?

* 2.6937 > 2.624, as a result our null hypothesis condition is satisfied.
* Hence reject H0
* At the given level of significance, we conclude that the mean number of patients per day is more than 25.

**Problem 44**

The American Water Works Association reports that the per capita water use in a single family home is 69 gallons per day. Legacy Ranch is a relatively new housing development consisting of 100 homes. The builders installed more efficient water fixtures, such as lowflush toilets, and subsequently conducted a survey of the residences. Thirty-six homes responded, and the sample mean water use per day was 64 gallons with a standard deviation of 8.8 gallons per day. At the .10 level of significance, is that enough evidence to conclude that residents of Legacy Ranch use less water on average?

**Answer**:

1. State the null hypothesis and the alternate hypothesis.

* The Keyword here is ‘less‘, this become a one tail hypothesis

H0 : μ ≥ 69

H1 : μ < 69

1. State the decision rule.

* When t < -tα , n-1, reject H0;
* In our case the significance level (α) is 0.10, df=36-1=35, the corresponding t value is -1.306

1. Compute the value of the test statistic.

* t-value = = = -3.41

1. What is your decision regarding H?

* -3.41 < -1. 306, as a result our null hypothesis condition is satisfied.
* Hence we can reject H0
* At the 10% level of significance, this is strong evidence to conclude that residents of Legacy Ranch use less water on average.
* The p-value is estimated by looking at the t-value (3.41) with df(35). P-value is between 0.005 and 0.0005.

**Problem 48**

An insurance company, based on past experience, estimates the mean damage for a natural disaster in its area is $5,000. After introducing several plans to prevent loss, it randomly samples 200 policyholders and finds the mean amount per claim was $4,800 with a standard deviation of $1,300. Does it appear the prevention plans were effective in reducing the mean amount of a claim? Use the .05 significance level.

**Answer**:

1. State the null hypothesis and the alternate hypothesis.

* The Keyword here is ‘reducing ‘, this become a two tail hypothesis

H0 : μ ≥ 5,000

H1 : μ < 5,000

1. State the decision rule.

* If t-value is negative, when t < -tα , n-1, reject H0; if t-value is positive, when t > tα , n-1 , reject H0.
* In our case the significance level (α) is 0.05, df = 200-1=199 the corresponding t value is -1.653

1. Compute the value of the test statistic.

* t-value = = = -2.1757

1. What is your decision regarding H?

* -1.653 > -2.1757, as a result our null hypothesis condition is satisfied.
* Hence we can reject H0
* At the 5% level of significance, it appears that the prevention plans were effective in reducing the mean amount of a claim.

**Part II:** The following problems require SAS. You will need to use the “IF” or “WHERE” statements to subset the data when appropriate, PROC MEANS to generate confidence intervals and default hypothesis test statistics, and PROC TTEST for general hypothesis testing. Don’t forget to copy the respective data files from Canvas and then use either a permanent SAS data library to import the spreadsheet data or the INFILE statement to import the text data. You must submit a copy of the relevant pages of your SAS output and interpretations when you turn in the exercise.

1. Data Set Exercise 69 (Parts a-c), on page 311 of the textbook. This problem uses the Goodyear, Arizona Real Estate data (REAL-ESTATE-2003.csv) that is documented on page 716 of the textbook and is available via Canvas. [3 pts]

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

a. Develop a 95 percent confidence interval for the mean selling price of the homes.

b. Develop a 95 percent confidence interval for the mean distance the home is from the center of the city.

c. Develop a 95 percent confidence interval for the proportion of homes with an attached garage.

**Answer**:

1. and (b)

TITLE "V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**PROC** **TTEST** DATA=v506.RealEstate ALPHA=**.05** SIDES=**2**;

VAR Price Distance;

**RUN**;

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

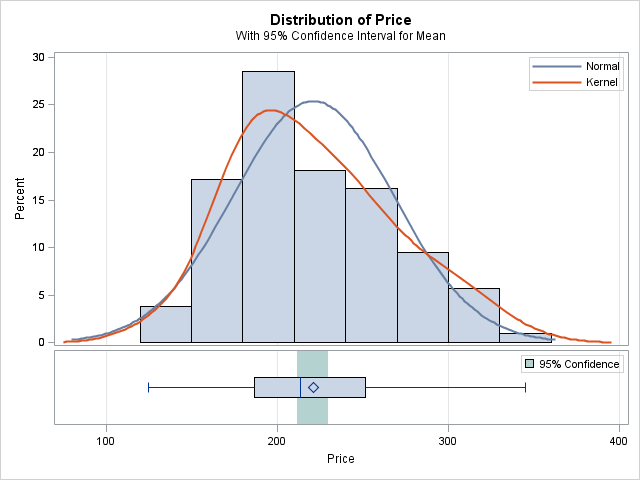
The TTEST Procedure

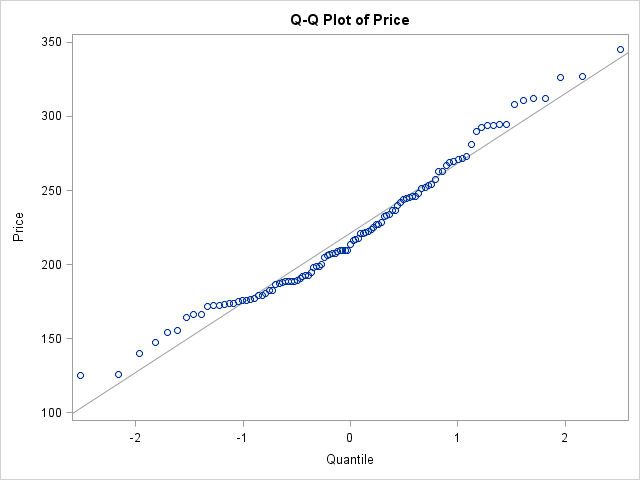
Variable: Price (Price)

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 105 | 221.1 | 47.1100 | 4.5975 | 125.0 | 345.3 |

| **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 221.1 | 212.0 | 230.2 | 47.1100 | 41.4856 | 54.5126 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 104 | 48.09 | <.0001 |



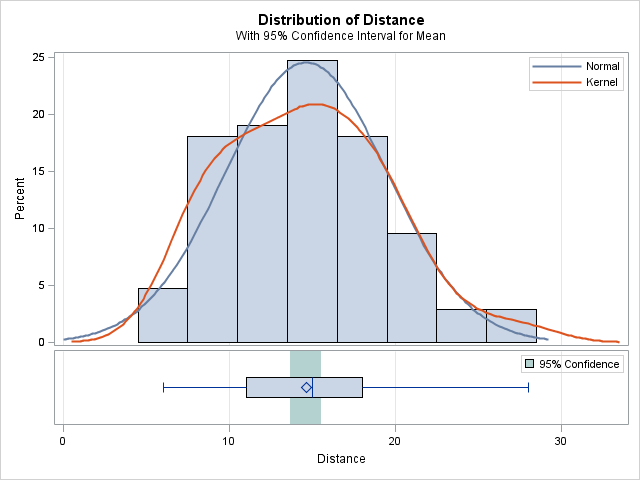


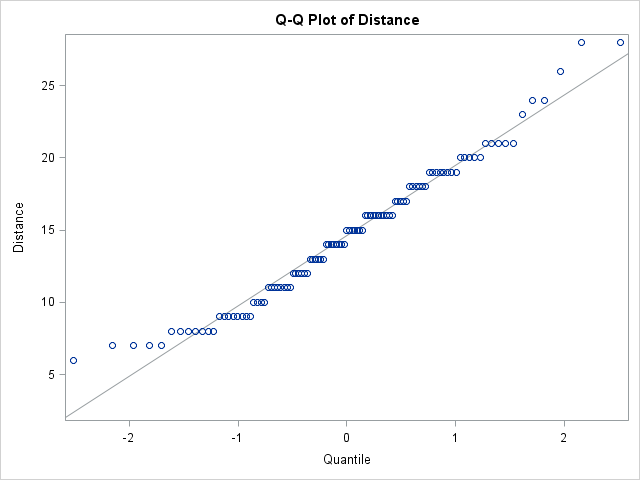
|  |
| --- |
| Variable: Distance (Distance) |

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 105 | 14.6286 | 4.8739 | 0.4756 | 6.0000 | 28.0000 |

| **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 14.6286 | 13.6854 | 15.5718 | 4.8739 | 4.2920 | 5.6398 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 104 | 30.76 | <.0001 |





**Answer** (c):

**PROC** **FREQ** DATA=v506.RealEstate;

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

**RUN**;

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

The FREQ Procedure

| **Garage** | | | | |
| --- | --- | --- | --- | --- |
| **Garage** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 71 | 67.62 | 71 | 67.62 |
| **0** | 34 | 32.38 | 105 | 100.00 |

| **Binomial Proportion** | |
| --- | --- |
| **Garage = 1** | |
| **Proportion** | 0.6762 |
| **ASE** | 0.0457 |
| **95% Lower Conf Limit** | 0.5867 |
| **95% Upper Conf Limit** | 0.7657 |
|  |  |
| **Exact Conf Limits** |  |
| **95% Lower Conf Limit** | 0.5779 |
| **95% Upper Conf Limit** | 0.7643 |

| **Test of H0: Proportion = 0.5** | |
| --- | --- |
| **ASE under H0** | 0.0488 |
| **Z** | 3.6108 |
| **One-sided Pr < Z** | 0.0002 |
| **Two-sided Pr > |Z|** | 0.0003 |

|  |
| --- |
| Sample Size = 105 |

1. Computer Data Exercise 50, on page 347 of the textbook. This problem also uses the REAL-ESTATE-2003.csv dataset. [6 pts]
2. A recent article in the Arizona Republic indicated that the mean selling price of the homes in the area is more than $220,000. Can we conclude that the mean selling price in the Goodyear, AZ, area is more than $220,000? Use the .01 significance level. What is the p-value?
3. The same article reported the mean size was more than 2,100 square feet. Can we conclude that the mean size of homes sold in the Goodyear, AZ, area is more than 2,100 square feet? Use the .01 significance level. What is the p-value?

Also complete the following questions:

1. Determine the proportion of homes that have an attached garage. At the .05 significance level, can we conclude that more than 60 percent of the homes sold in Goodyear, AZ, area had an attached garage? What is the p-value?
2. Determine the proportion of homes that have a pool. At the .05 significance level, can we conclude that more than 60 percent of homes sold in the Goodyear, AZ, area had a pool? What is the p-value?

**Answer (a)**:

This is a one sided hypothesis test; we are testing on the right side of the distribution.

TITLE "V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**PROC** **TTEST** H0=**220** ALPHA=**.01** SIDES=U DATA=v506.RealEstate;

VAR Price;

**RUN**;

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

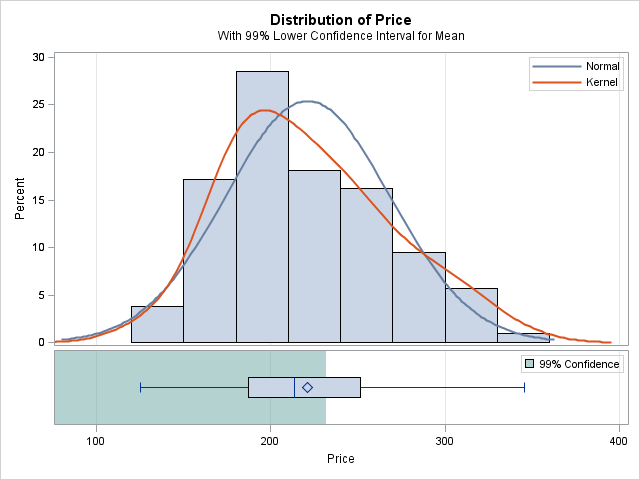
The TTEST Procedure

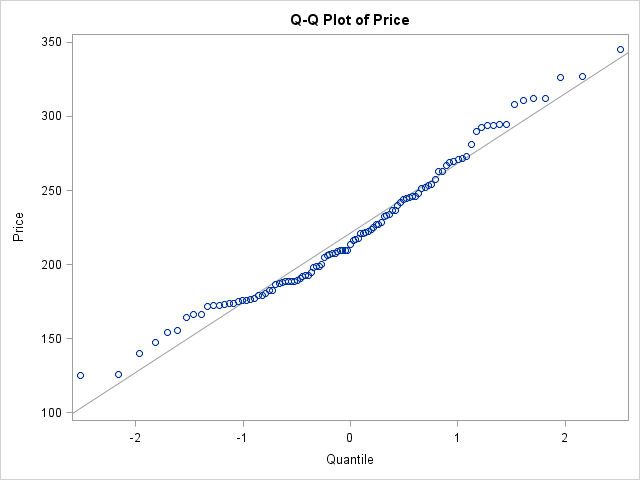
Variable: Price (Price)

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 105 | 221.1 | 47.1100 | 4.5975 | 125.0 | 345.3 |

| **Mean** | **99% CL Mean** | | **Std Dev** | **99% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 221.1 | -Infty | 232.0 | 47.1100 | 39.9124 | 57.1752 |

| **DF** | **t Value** | **Pr < t** |
| --- | --- | --- |
| 104 | 0.24 | 0.5944 |





From the distribution and table we cannot conclude that the mean selling price in the Goodyear, AZ, area is more than $220,000.

**Answer (b)**:

This is a one sided hypothesis test; we are testing on the right side of the distribution.

**PROC** **TTEST** H0=**2100** ALPHA=**.01** SIDES=L DATA=v506.RealEstate;

VAR Size;

**RUN**;

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

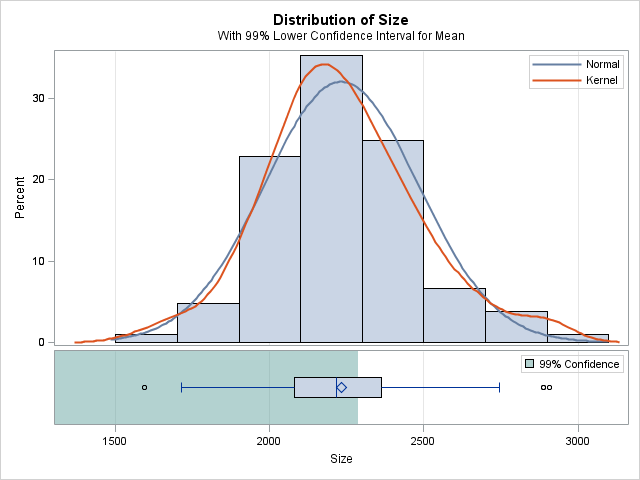
The TTEST Procedure

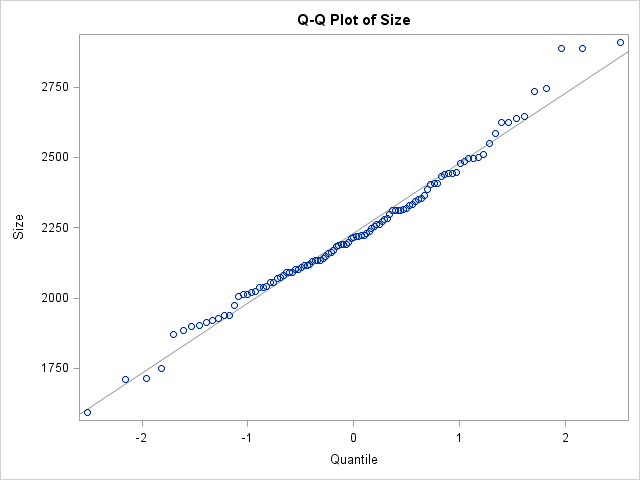
Variable: Size (Size)

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 105 | 2231.4 | 249.3 | 24.3307 | 1593.0 | 2908.4 |

| **Mean** | **99% CL Mean** | | **Std Dev** | **99% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 2231.4 | -Infty | 2288.9 | 249.3 | 211.2 | 302.6 |

| **DF** | **t Value** | **Pr < t** |
| --- | --- | --- |
| 104 | 5.40 | 1.0000 |





We cannot reject our null hypothesis as it has a high probability. Hence we cannot confidently conclude that the mean size of homes sold in the Goodyear, AZ, area is more than 2,100 square feet.

**Answer (c)**:

This is a hypothesis test; we are testing using the proportion of the distribution.

**PROC** **FREQ** DATA=v506.RealEstate;

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

**RUN**;

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

The FREQ Procedure

| **Garage** | | | | |
| --- | --- | --- | --- | --- |
| **Garage** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 71 | 67.62 | 71 | 67.62 |
| **0** | 34 | 32.38 | 105 | 100.00 |

| **Binomial Proportion** | |
| --- | --- |
| **Garage = 1** | |
| **Proportion** | 0.6762 |
| **ASE** | 0.0457 |
| **95% Lower Conf Limit** | 0.5867 |
| **95% Upper Conf Limit** | 0.7657 |
|  |  |
| **Exact Conf Limits** |  |
| **95% Lower Conf Limit** | 0.5779 |
| **95% Upper Conf Limit** | 0.7643 |

| **Test of H0: Proportion = 0.5** | |
| --- | --- |
| **ASE under H0** | 0.0488 |
| **Z** | 3.6108 |
| **One-sided Pr < Z** | 0.0002 |
| **Two-sided Pr > |Z|** | 0.0003 |

Yes we can conclude that more than 60 percent of the homes sold in Goodyear, AZ, area had an attached garage .The p value is highlighted above.

**Answer (d)**:

**PROC** **FREQ** DATA=v506.RealEstate;

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

**RUN**;

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

The FREQ Procedure

| **Pool** | | | | |
| --- | --- | --- | --- | --- |
| **Pool** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 38 | 36.19 | 38 | 36.19 |
| **0** | 67 | 63.81 | 105 | 100.00 |

| **Binomial Proportion** | |
| --- | --- |
| **Pool = 1** | |
| **Proportion** | 0.3619 |
| **ASE** | 0.0469 |
| **95% Lower Conf Limit** | 0.2700 |
| **95% Upper Conf Limit** | 0.4538 |
|  |  |
| **Exact Conf Limits** |  |
| **95% Lower Conf Limit** | 0.2704 |
| **95% Upper Conf Limit** | 0.4615 |

| **Test of H0: Proportion = 0.5** | |
| --- | --- |
| **ASE under H0** | 0.0488 |
| **Z** | 2.8301 |
| **One-sided Pr > Z** | 0.0023 |
| **Two-sided Pr > |Z|** | 0.0047 |
|  |

Our interpretation is that for the given confidence level **proportion** the proportion of homes that have a pool is 0.3619. From the given confidence level, we can say that the required proportion lie between the range 0.2700 to 0.4538. As we did not know the proportion to be tested, the system assumes as 50-50 ratio for the smoking and non-smoking males. This Hypothesis is however rejected.

1. Data Set Exercise 52, page 347 of the textbook. This problem uses the Buena School District Bus data (Buena School District.csv) that is documented on page 759 of the textbook and is available via Canvas. [6 pts]
   1. Select the variable for the number of miles traveled last month. Conduct a test of hypothesis to determine whether the mean number of miles traveled is equal to 840. Use the .01 significance level. Find the p-value and explain what it means.
   2. Using the maintenance cost variable, conduct a test of hypothesis to determine whether the mean maintenance cost is less than $500 at the .05 significance level. Determine the p-value and interpret the result.

Also complete the following questions:

* 1. Suppose we consider a bus “old” if it is more than eight years old. At the .01 significance level, can we conclude that less than 40 percent of the buses are old? Report the p-value.

**Answer (a)**:

TITLE "V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**PROC** **TTEST** H0=**840** data=v506.Buses SIDES=**2** ALPHA=**.01**;

var Miles;

**RUN**;

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

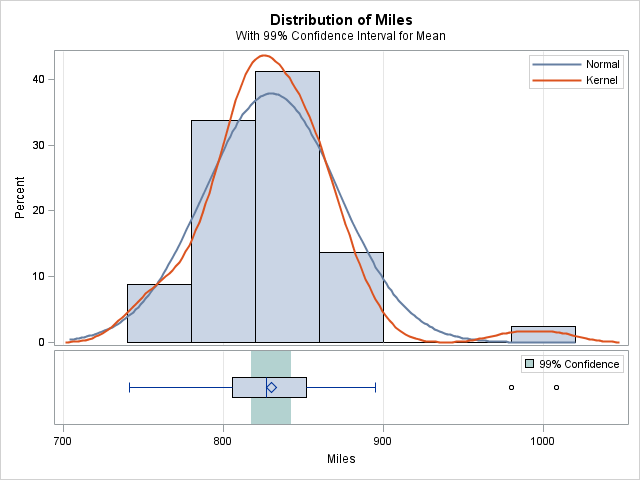
The TTEST Procedure

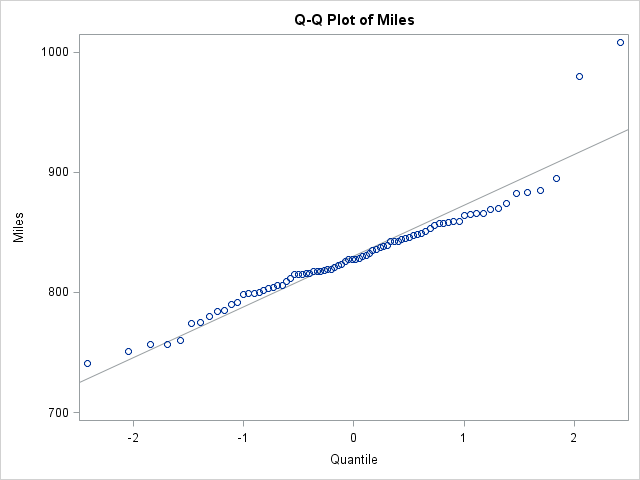
Variable: Miles

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 80 | 830.1 | 42.1882 | 4.7168 | 741.0 | 1008.0 |

| **Mean** | **99% CL Mean** | | **Std Dev** | **99% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 830.1 | 817.7 | 842.6 | 42.1882 | 34.9491 | 52.8315 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 79 | -2.10 | 0.0393 |





We can see form the above table that the probability of the mean number of miles is low in the tails and thus we can say our null hypothesis cannot be rejected for the given scenario.

**Answer (b)**:

**PROC** **TTEST** H0=**500** data=v506.Buses SIDES=U ALPHA=**.05**;

var Maintenance;

**RUN**;

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

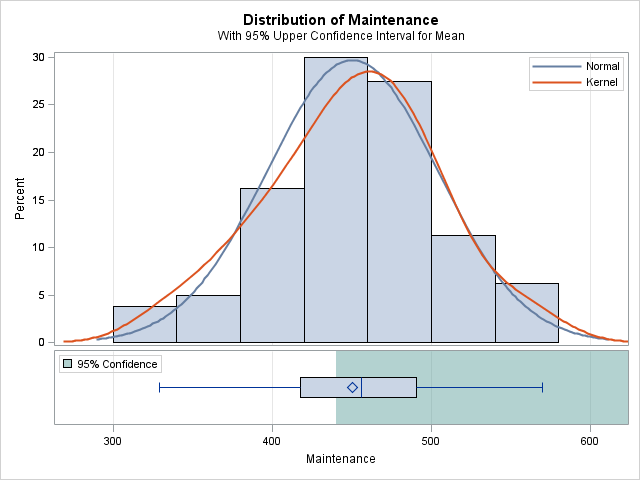
The TTEST Procedure

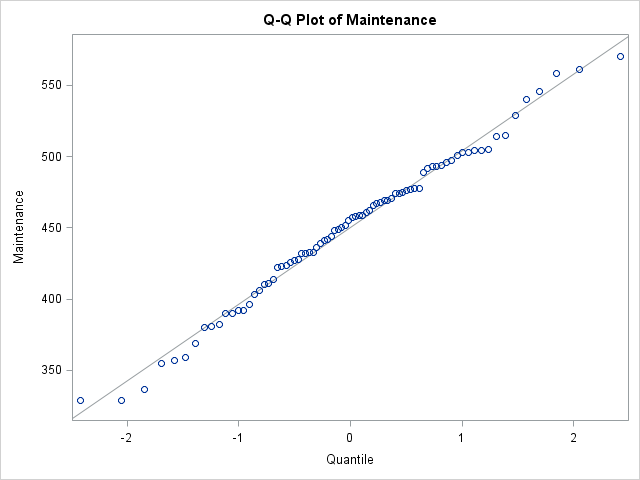
Variable: Maintenance

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 80 | 450.3 | 53.6860 | 6.0023 | 329.0 | 570.0 |

| **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 450.3 | 440.3 | Infty | 53.6860 | 46.4627 | 63.5895 |

| **DF** | **t Value** | **Pr > t** |
| --- | --- | --- |
| 79 | -8.28 | 1.0000 |





From the above table we can see that we cannot reject the hypothesis as the probability of the maintenance cost greater than $500 is very high.

**Answer (c):**

**DATA** temp;

SET v506.Buses;

IF age>=**8** THEN dage=**1**;

ELSE dage=**0**;

**RUN**;

**PROC** **FREQ** DATA=temp;

TABLES dage / BINOMIAL(p=0.4 LEVEL=**2**) ALPHA=**.01**;

**RUN**;

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

The FREQ Procedure

| **dage** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **0** | 42 | 52.50 | 42 | 52.50 |
| **1** | 38 | 47.50 | 80 | 100.00 |

| **Binomial Proportion** | |
| --- | --- |
| **dage = 1** | |
| **Proportion (P)** | 0.4750 |
| **ASE** | 0.0558 |
| **99% Lower Conf Limit** | 0.3312 |
| **99% Upper Conf Limit** | 0.6188 |
| **Exact Conf Limits** |  |
| **99% Lower Conf Limit** | 0.3302 |
| **99% Upper Conf Limit** | 0.6229 |

| **Test of H0: Proportion = 0.4** | |
| --- | --- |
| **ASE under H0** | 0.0548 |
| **Z** | 1.3693 |
| **One-sided Pr > Z** | 0.0855 |
| **Two-sided Pr > |Z|** | 0.1709 |
| **Exact Test** |  |
| **One-sided Pr >= P** | 0.1053 |
| **Two-sided = 2 \* One-sided** | 0.2107 |

|  |
| --- |
| **Sample Size = 80** |

Here the p- values for one tail and two tail are - 0.0855 and 0.1709 respectively.

4. Carry out the following tasks with the data from the Baltimore Longitudinal Study of Aging. This data set is named: BLSA.DAT, which is a text file available from Canvas. [5 pts]

a. Construct a 95% confidence interval for the **mean** diastolic blood pressure of all males. Interpret the results.

b. Construct a 99% confidence interval for the **proportion** of males that smoke. Interpret the results.

c. Construct a 90% confidence interval for the **mean** systolic blood pressure of all females who are over 30 years old. Interpret the results.

d. Test the claim, at the .05 level of significance, that males who are 50 and older have a mean systolic blood pressure that is greater than the mean systolic blood pressure for all men. Use the p-value method.

e. Test the claim, at the .01 level, that females who have lower weight than the mean weight of all females also have a lower mean systolic blood pressure than the average female. Use the p-value method.

The variables in the BLSA data set are, in order, subject number, sex, age, smoker, systolic blood pressure (SBP), diastolic blood pressure (DBP), height, and weight for a random sample of 720 adults. Ages are given in years, blood pressure readings are in millimeters of mercury, heights are given in centimeters, and weights are given in kilograms. Male gender is indicated with a value of M and female gender with a value of F. A smoker is given a value of Y and a non-smoker a value of N. To use an IF statement to subset the data to only include males with a character variable such as sex, use the following IF statement:

IF sex=”M”;

Note the quotes around the M, which is required in SAS whenever a character value is used in an IF statement.

TITLE "V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**PROC** **TTEST** data=v506.Blsa2 ALPHA=**.05**;

WHERE SEX = "M";

var DBP;

**RUN**;

**PROC** **FREQ** DATA=v506.Blsa2;

WHERE SEX = "M";

TABLES smoker / BINOMIAL(LEVEL=**2**) ALPHA=**.01**;

**RUN**;

**PROC** **TTEST** data=v506.Blsa2 ALPHA=**.1**;

WHERE AGE > **30**;

var SBP;

**RUN**;

**PROC** **UNIVARIATE** DATA=v506.Blsa2;

WHERE SEX = "M";

var SBP;

**RUN**;

**PROC** **TTEST** H0=**130.841667** ALPHA=**.05** SIDES=**2** DATA=v506.Blsa2;

WHERE AGE > **50** AND SEX = "M";

VAR SBP;

**RUN**;

**PROC** **UNIVARIATE** DATA=v506.Blsa2;

WHERE SEX = "F";

var weight;

**RUN**;

**PROC** **TTEST** H0=**130.841667** ALPHA=**.01** SIDES=U DATA=v506.Blsa2;

WHERE weight < **62.2883008** AND SEX = "F";

VAR SBP;

**RUN**;

**Answer (a):**

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

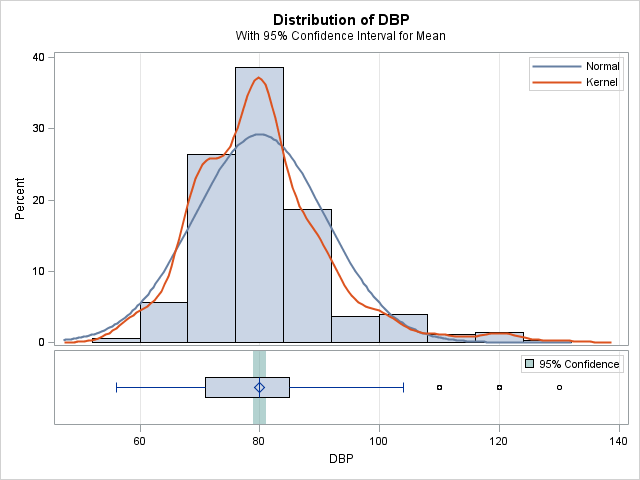
The TTEST Procedure

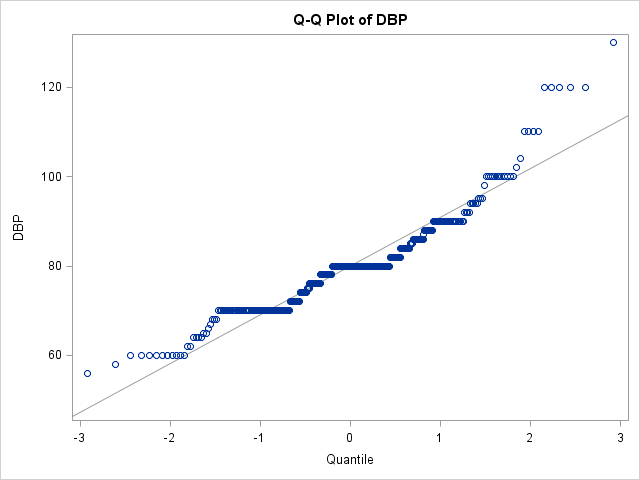
Variable: DBP

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 360 | 79.9944 | 10.9252 | 0.5758 | **56.0000** | 130.0 |

| **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 79.9944 | 78.8621 | 81.1268 | 10.9252 | 10.1812 | 11.7874 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 359 | 138.93 | <.0001 |





Our interpretation is that for the given confidence level the **mean** diastolic blood pressure of all males falls in the range of 78.8621 to 81.1268

|  |
| --- |
| **Answer (b):**  V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

The FREQ Procedure

| **smoker** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **Y** | 129 | 35.83 | 129 | 35.83 |
| **N** | 231 | 64.17 | 360 | 100.00 |

| **Binomial Proportion** | |
| --- | --- |
| **smoker = Y** | |
| **Proportion** | 0.3583 |
| **ASE** | 0.0253 |
| **99% Lower Conf Limit** | 0.2932 |
| **99% Upper Conf Limit** | 0.4234 |
|  |  |
| **Exact Conf Limits** |  |
| **99% Lower Conf Limit** | 0.2941 |
| **99% Upper Conf Limit** | 0.4264 |

| **Test of H0: Proportion = 0.5** | |
| --- | --- |
| **ASE under H0** | 0.0264 |
| **Z** | -5.3759 |
| **One-sided Pr < Z** | <.0001 |
| **Two-sided Pr > |Z|** | <.0001 |

|  |
| --- |
| **Sample Size = 360** |

Our interpretation is that for the given confidence level **proportion** of males that smoke is 0.3583. From the given confidence level, we can say that the required proportion lie between the range 0.2932 to 0.4234. As we did not know the proportion to be tested, the system assumes as 50-50 ratio for the smoking and non-smoking males. This Hypothesis is however rejected.

**Answer (c):**

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

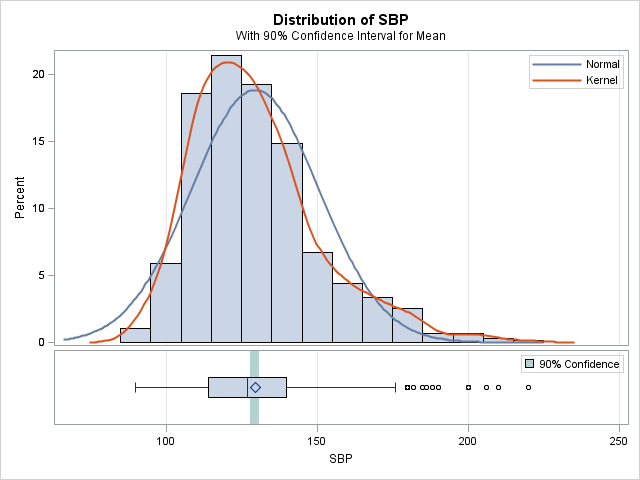
The TTEST Procedure

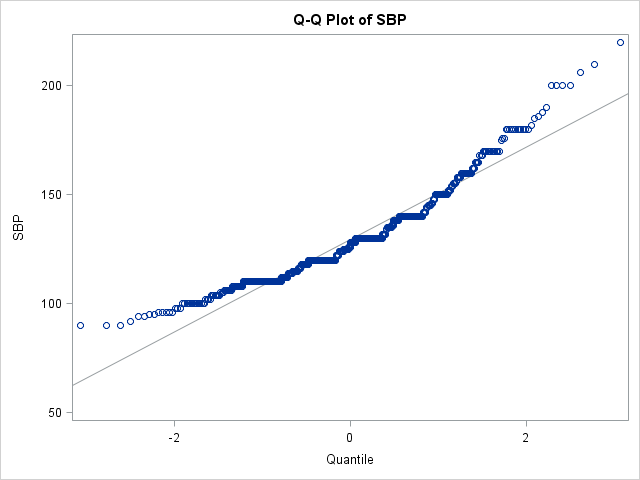
Variable: SBP

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 592 | 129.5 | 21.1446 | 0.8690 | 90.0000 | 220.0 |

| **Mean** | **90% CL Mean** | | **Std Dev** | **90% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 129.5 | 128.1 | 130.9 | 21.1446 | 20.1829 | 22.2110 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 591 | 149.02 | <.0001 |





Our interpretation is that for the given confidence level the **mean** diastolic blood pressure of all males falls in the range of 128.1 to 130.9

|  |
| --- |
| **Answer (d):**  V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

The UNIVARIATE Procedure

Variable: SBP

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 360 | **Sum Weights** | 360 |
| **Mean** | 130.841667 | **Sum Observations** | 47103 |
| **Std Deviation** | 20.2791092 | **Variance** | 411.24227 |
| **Skewness** | 1.19381913 | **Kurtosis** | 1.80123291 |
| **Uncorrected SS** | 6310671 | **Corrected SS** | 147635.975 |
| **Coeff Variation** | 15.4989689 | **Std Error Mean** | 1.0688029 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 130.8417 | **Std Deviation** | 20.27911 |
| **Median** | 129.0000 | **Variance** | 411.24227 |
| **Mode** | 130.0000 | **Range** | 130.00000 |
|  |  | **Interquartile Range** | 22.00000 |

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

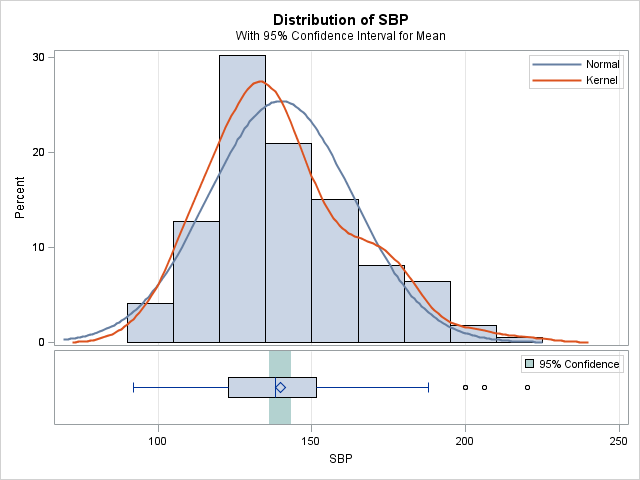
The TTEST Procedure

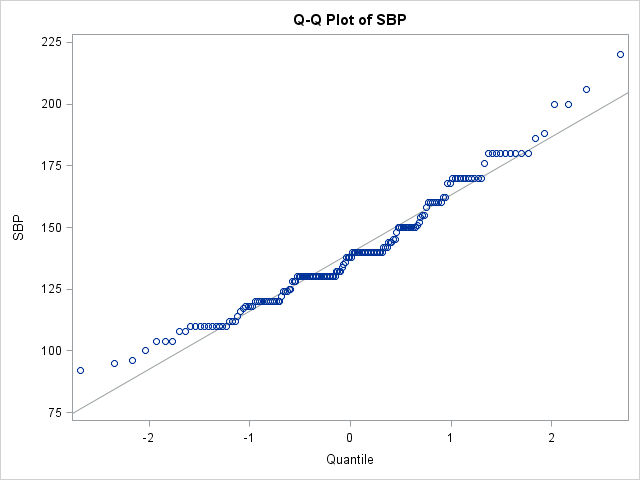
Variable: SBP

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 172 | 139.8 | 23.5334 | 1.7944 | 92.0000 | 220.0 |

| **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| 139.8 | 136.3 | 143.3 | 23.5334 | 21.2816 | 26.3223 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 171 | 4.99 | <.0001 |





As we can see from our calculations the p-value is large. Hence we can say that our hypothesis is true.

|  |
| --- |
| **Answer (e):**  V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

The UNIVARIATE Procedure

Variable: weight

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 359 | **Sum Weights** | 359 |
| **Mean** | 62.2883008 | **Sum Observations** | 22361.5 |
| **Std Deviation** | 10.6930499 | **Variance** | 114.341315 |
| **Skewness** | 1.19157885 | **Kurtosis** | 2.53589568 |
| **Uncorrected SS** | 1433794.03 | **Corrected SS** | 40934.1909 |
| **Coeff Variation** | 17.1670277 | **Std Error Mean** | 0.56435759 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 62.28830 | **Std Deviation** | 10.69305 |
| **Median** | 60.60000 | **Variance** | 114.34132 |
| **Mode** | 53.80000 | **Range** | 70.90000 |
|  |  | **Interquartile Range** | 12.70000 |

|  |
| --- |
| V506 HOMEWORK03 PART 2 - JIVITESH POOJARY AND QIWEN ZHU |

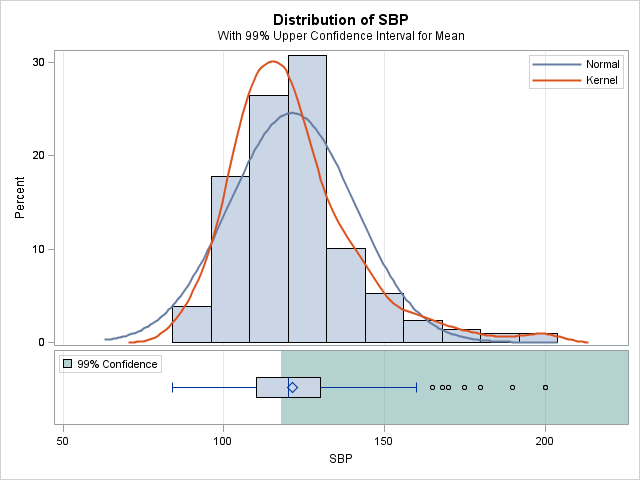
The TTEST Procedure

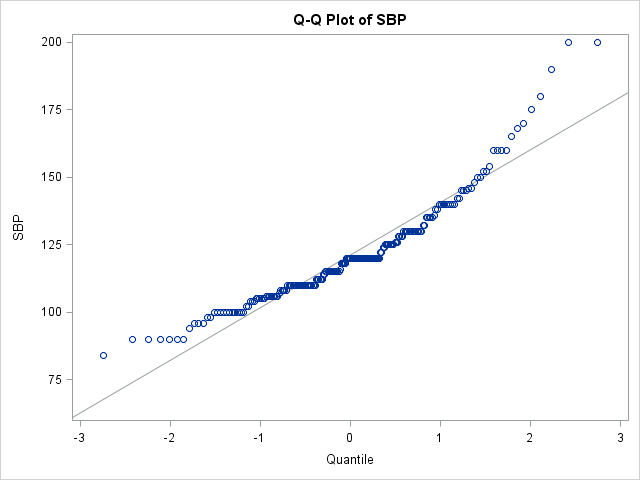
**Variable: SBP**

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| **208** | **121.2** | **19.4565** | **1.3491** | **84.0000** | **200.0** |

| **Mean** | **99% CL Mean** | | **Std Dev** | **99% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- |
| **121.2** | **118.0** | **Infty** | **19.4565** | **17.2560** | **22.2456** |

| **DF** | **t Value** | **Pr > t** |
| --- | --- | --- |
| **207** | **-7.14** | **1.0000** |





As we can see from our calculations the p-value is large. Hence we can say that our hypothesis cannot be rejected.