SPEA V506 Homework Exercise 2

***Authors: Jivitesh Poojary and Qiwen Zhu***

Probability Distributions and the Binomial

Page 181, Problem 6 (1 pt)

The director of admissions at Kinzua University in Nova Scotia estimated the distribution of student admissions for the fall semester on the basis of past experience. What is the expected number of admissions for the fall semester? Compute the variance and the standard deviation of the number of admissions.

**Answer**:

Given Information:

|  |  |
| --- | --- |
| Admissions | Probability |
| 1000 | 0.6 |
| 1200 | 0.3 |
| 1500 | 0.1 |

**Mean (µ) =**

= (1000 x 0.6) + (1200 x 0.3) + (1500 x 0.1) = 1100

The **expected number of admission** for the fall semester is **1100.**

**Variance (σ2) =**

= ((1000-1100)2 x 0.6) + ((1200-1100)2 x 0.3) + ((1500-1100)2 x 0.1) = **25000**

**Standard Deviation (σ)** = = **158.1139**

Page 188, Problem 14 (2 pts)

The United States Postal Service reports 95 percent of first class mail within the same city is delivered within two days of the time of mailing. Six letters are randomly sent to different locations.

a. What is the probability that all six arrive within two days?

b. What is the probability that exactly five arrive within two days?

c. Find the mean number of letters that will arrive within two days.

d. Compute the variance and standard deviation of the number that will arrive within two days.

**Answer**:

1. ***P*(*x*) = *nCx* πx (1 – π) n - x**

*n = 6,* π = 0.95 *and x = 6*

*P*(*x*) = *6C6* (0.95)6 (1 – 0.95) 6 - 6 = 0.7351

1. ***P*(*x*) = *nCx* πx (1 – π) n - x**

*n = 6,* π = 0.95 *and x = 5*

*P*(*x*) = *6C5* (0.95)5 (1 – 0.95) 6 - 5 = 0.2321

1. **Mean (µ) = nπ** = 6\*0.95 = 5.7
2. **Variance (σ2) = nπ(1- π)** = 6\*0.95\*0.05 = 0.285

**Standard Deviation (σ)** = = 0.5339

Page 189, Problem 17 (1.5 pts)

A recent survey by the American Accounting Association revealed 23 percent of students graduating with a major in accounting select public accounting. Suppose we select a sample of 15 recent graduates.

a. What is the probability two select public accounting?

b. What is the probability five select public accounting?

c. How many graduates would you expect to select public accounting?

**Answer**:

1. ***P*(*x*) = *nCx* πx (1 – π) n - x**

*n = 15,* π = 0.23 *and x = 2*

*P*(*x*) = *15C2* (0.23)2 (1 – 0.23) 15 - 2 = 0.1861

1. ***P*(*x*) = *nCx* πx (1 – π) n - x**

*n = 15,* π = 0.23 *and x = 5*

*P*(*x*) = *15C5* (0.23)5 (1 – 0.23) 15 - 5 = 0.1415

1. **Mean (µ) = nπ**

= 15\*0.23 = 3.45

3 to 4 graduates are expected to select public accounting.

Page 191, Problem 22 (1.5 pts)

A manufacturer of window frames knows from long experience that 5 percent of the production will have some type of minor defect that will require an adjustment. What is the probability that in a sample of 20 window frames:

a. None will need adjustment?

b. At least one will need adjustment?

c. More than two will need adjustment?

**Answer**:

1. the probability that none of the 20 frames will need adjustment: n = 20

***P*(*x*) = *nCx* πx (1 – π) n - x**

*P*(*x*) = *20C0* (0.05)0 (1 – 0.05) 20 - 0 = 0.3585

1. P(>=1) = 1-P(<1) = 1 – P(0) = 1-0.3585 = 0.6415
2. P(>2) = 1 - P(<=2) = 1 - P(2) - P(1) - P(0) = 0.0754

***P*(*x*) = *nCx* πx (1 – π) n - x**

*P*(*x*) = *20C0* (0.05)0 (1 – 0.05) 20 - 0 = 0.3585

*P*(*x*) = *20C1* (0.05)1 (1 – 0.05) 20 - 1 = 0.3774

*P*(*x*) = *20C2* (0.05)2 (1 – 0.05) 20 - 2 = 0.1887

Thus,

1 - P(2) - P(1) - P(0)

= 1 - 0.3585 - 0.3774 - 0.1887

= 0.0754

The Normal Probability Distribution

Page 217, Problem 10 (1.5 pts)

The mean of a normal probability distribution is 60; the standard deviation is 5.

a. About what percent of the observations lie between 55 and 65?

b. About what percent of the observations lie between 50 and 70?

c. About what percent of the observations lie between 45 and 75?

**Answer**:

1. Z =

Z = = (-1) and Z = = 1

The observation at a distance of 1 standard deviation on both sides.

P(observation) = P(>55 and <60) + P(<65 and <60)

= 0.3413 + 0.3413

= 68.26%

1. Z =

Z = = (-2) and Z = = 2

The observation at a distance of 2 standard deviations on both sides.

P(observation) = P(>50 and <60) + P(<70 and <60)

= 0.4772 + 0.4772

= 95.45%

1. Z =

Z = = (-3) and Z = = 3

The observation at a distance of 3 standard deviations on both sides.

P(observation) = P(>45 and <60) + P(<75 and <60)

= 0.4987 + 0.4987

= 99.73%

practically all of the observations lie between 45 and 75; or 99.73%

Page 221, Problem 14 (1.5 pts)

A normal population has a mean of 12.2 and a standard deviation of 2.5.

a. Compute the z value associated with 14.3.

b. What proportion of the population is between 12.2 and 14.3?

c. What proportion of the population is less than 10.0?

**Answer**:

1. z-value = = = 0.84
2. 0.2995
3. z-value = = = 0.88

P(between 10 and 12.2) = 0.3106

P(<10) = 0.5000 - 0.3106 = 0.1894

Page 221, Problem 15 (1.5 pts)

A recent study of the hourly wages of maintenance crew members for major airlines showed that the mean hourly salary was $20.50, with a standard deviation of $3.50. Assume the distribution of hourly wages follows the normal probability distribution. If we select a crew member at random, what is the probability the crew member earns:

a. Between $20.50 and $24.00 per hour?

b. More than $24.00 per hour?

c. Less than $19.00 per hour?

**Answer:**

1. z-value = = = 1

P(between 20.50 and 24) = 0.3413

1. P(>24) = 0.5000 - P(between 20.50 and 24)

= 0.5000-0.3413

= 0.1587

1. z-value = = = 0.4236

P(between 20.50 and 19) = 0.1664

P(<19) = 0.5000 - P(between 20.50 and 19)

= 0.5000-0.1644

= 0.3336

Page 223, Problem 17 (1.5 pts)

A normal distribution has a mean of 50 and a standard deviation of 4.

a. Compute the probability of a value between 44.0 and 55.0.

b. Compute the probability of a value greater than 55.0.

c. Compute the probability of a value between 52.0 and 55.0.

**Answer:**

1. z-value = = = 1.5

P(between 44 and 50) = 0.4332

z-value = = = 1.2

P(between 55 and 50) = 0.3944

P(between 44 and 55) = P(between 44 and 50) + P(between 55 and 50)

= 0.4332 + 0.3944

= 0.8276

1. z-value = = = 1.2

P(between 50 and 55) = 0.3944

P(>55) = 0.5000 - P(between 50 and 55)

= 0.5000-0.3944

= 0.1056

1. z-value = = = 0.5

P(between 52 and 50) = 0.3944

z-value = = = 1.2

P(between 55 and 50) = 0.1915

P(between 52 and 55) = P(between 55 and 50) - P(between 52 and 50)

= 0.3944-0.1915

= 0.2029

Page 223, Problem 20 (1.5 pts)

The number of viewers of American Idol has a mean of 29 million with a standard deviation of 5 million. Assume this distribution follows a normal distribution. What is the probability that next week’s show will:

a. Have between 30 and 34 million viewers?

b. Have at least 23 million viewers?

c. Exceed 40 million viewers?

**Answer**:

1. z-value = = = 0.2

P(between 29 and 30) = 0.0793

z-value = = = 1.0

P(between 29 and 34) = 0.3413

P(between 30 and 34) = P(between 29 and 34) - P(between 29 and 30)

= 0.3413-0.0793

= 0.2620

1. z-value = = = -1.2

P(between 23 and 29) = 0.3849

P(At least 23) = P(between 23 and 29) + 0.5000

= 0.3849+0.5000

= 0.8849

1. z-value = = = 2.2

P(between 29 and 40) = 0.4861

P(>40) = 0.5000 - P(between 29 and 40)

= 0.5000-0.4861

= 0.0139

Page 224, Problem 21 (1.5 pts)

WNAE, an all-news AM station, finds that the distribution of the lengths of time listeners are tuned to the station follows the normal distribution. The mean of the distribution is 15.0 minutes and the standard deviation is 3.5 minutes. What is the probability that a particular listener will tune in:

a. More than 20 minutes?

b. For 20 minutes or less?

c. Between 10 and 12 minutes?

**Answer**:

1. z-value = = = 1.4286

P(between 15 and 20) = 0.4236

P(>20) = 0.5000 - P(between 15 and 20)

= 0.5000-0.4236

= 0.0764

1. z-value = = = 1.4286

P(between 15 and 20) = 0.4236

P(>20) = 0.5000 + P(between 15 and 20)

= 0.5000+0.4236

= 0.9236

1. z-value = = = -0.8571

P(between 12 and 15) = 0.4236

z-value = = = -1.4286

P(between 10 and 15) = 0.3051

P(between 10 and 12) = P(between 12 and 15) - P(between 10 and 12)

= 0.4236-0.3051

= 0.1185

Page 226, Problem 23 (1 pt)

A normal distribution has a mean of 50 and a standard deviation of 4. Determine the value below which 95 percent of the observations will occur.

**Answer**:

1. Out of the 95% of the required observation 50% of the total values will fall before the mean on the standard deviation curve and the remaining 45% will be on the right side of the mean.

The z-value of the 45% region = ~1.65

z-value =

1.65 =

4 x 1.65 = X – 50

6.6 = X – 50

X = 50 + 6.6 = 56.6

Page 226, Problem 28 (2 pts)

For the most recent year available, the mean annual cost to attend a private university in the United States was $26,889. Assume the distribution of annual costs follows the normal probability distribution and the standard deviation is $4,500. Ninety-five percent of all students at private universities pay less than what amount?

**Answer**:

1. Out of the 95% of the required observation 50% of the total values will fall before the mean on the standard deviation curve and the remaining 45% will be on the right side of the mean.

The z-value of the 45% region = ~1.65

z-value =

1.65 =

4500 x 1.65 = X – 26889

7425 = X – 26889

X = 26889+7425 = 34314

The Normal Approximation to the Binomial

Page 239, Problem 56 (2 pts)

A recent report in BusinessWeek indicated that 20 percent of all employees steal from their company each year. If a company employs 50 people, what is the probability that:

a. Fewer than 5 employees steal?

b. More than 5 employees steal?

c. Exactly 5 employees steal?

d. More than 5 but fewer than 15 employees steal?

**Answer**:

1. Mean ( = n = 50 x 0.2 = 10

Variance) = n = 50 x 0.2 (1 – 0.2) = 50 x 0.2 x 0.8 = 8

Standard Deviation ( = = 2.8284

P(<5) = 0.5000 - P(4.5)

Z-value = = = -1.944

P(4.5) = 0.4738

Thus, P(<5) = 0.5000 - 0.4738 = 0.0262

1. Mean ( = n = 50 x 0.2 = 10

Variance) = n = 50 x 0.2 (1 – 0.2) = 50 x 0.2 x 0.8 = 8

Standard Deviation ( = = 2.8284

P(>5) = 0.5000 + P(5.5)

Z-value = = = -1.5910

P(5.5) = 0.4441

P(>5) = 0.5000 + 0.4441= **0.9441**

1. *P*(*x*) = *nCx* πx (1 – π) n - x

*P*(*5*) = *50C5* (0.20)5 (1 – 0.20) 50 - 5

P(5) = **0.0295312043**

OR

Mean ( = n = 50 x 0.2 = 10

Variance) = n = 50 x 0.2 (1 – 0.2) = 50 x 0.2 x 0.8 = 8

Standard Deviation ( = = 2.8284

Z-value = = = -1.944

P(4.5) = 0.4738

Z-value = = = -1.5910

P(5.5) = 0.4441

P(5) = P(4.5) - P(5.5) = 0.4738 - 0.4441 = 0.0297

1. Mean ( = n = 50 x 0.2 = 10

Variance) = n = 50 x 0.2 (1 – 0.2) = 50 x 0.2 x 0.8 = 8

Standard Deviation ( = = 2.8284

Z-value = = = -1.5910

P(5.5) = 0.4441

Z-value = = = 1.5910

P(5.5) = 0.4441

P(5.5)+P(14.5) = 0.4441 + 0.4441 = **0.8882**

Page 239, Problem 57 (2 pts)

The Orange County Register, as part of its Sunday health supplement, reported that 64 percent of American men over the age of 18 consider nutrition a top priority in their lives. Suppose we select a sample of 60 men. What is the likelihood that:

a. 32 or more consider nutrition important?

b. 44 or more consider nutrition important?

c. More than 32 but fewer than 43 consider nutrition important?

d. Exactly 44 consider diet important?

**Answer**:

1. Mean ( = n = 60 x 0.64 = 38.4

Variance) = n = 60 x 0.64 (1 – 0.64) = 60 x 0.64 x 0.36 = 13.824

Standard Deviation ( = = 3.7181

P(>=32) = 0.5000 + P(31.5)

Z-value = = = -1.8558

P(31.5) = 0.4686

P(>=32) = 0.5000 + 0.4686 = 0.9686

1. Mean ( = n = 60 x 0.64 = 38.4

Variance) = n = 60 x 0.64 (1 – 0.64) = 60 x 0.64 x 0.36 = 13.824

Standard Deviation ( = = 3.7181

P(>=44) = 0.5000 - P(43.5)

Z-value = = = 1.3717

P(43.5) = 0.4147

P(>=44) = 0.5000 - 0.4147 = 0.0853

1. Mean ( = n = 60 x 0.64 = 38.4

Variance) = n = 60 x 0.64 (1 – 0.64) = 60 x 0.64 x 0.36 = 13.824

Standard Deviation ( = = 3.7181

Z-value = = = 1.5868

P(32.5) = 0.4441

Z-value = = = 1.3717

P(42.5) = 0.3643

P(more than 32 but less than 43) = P(32.5)+P(42.5) = 0.4441+0.3643 = 0.8084

1. *P*(*x*) = *nCx* πx (1 – π) n - x

*P*(*44*) = *60C44* (0.64)44 (1 – 0.64) 60 - 44

P(44) = 0.035294

OR

Mean ( = n = 60 x 0.64 = 38.4

Variance) = n = 60 x 0.64 (1 – 0.64) = 60 x 0.64 x 0.36 = 13.824

Standard Deviation ( = = 3.7181

Z-value = = = 1.3717

P(43.5) = 0.4147

Z-value = = = 1.6406

P(44.5) = 0.4495

P(44) = P(44.5) - P(43.5) = 0.4495 - 0.4147 = 0.0348

The Central Limit Theorem

Page 272, Problem 16 (2 pts)

A normal population has a mean of 75 and a standard deviation of 5. You select a sample of 40. Compute the probability the sample mean is:

a. Less than 74.

b. Between 74 and 76.

c. Between 76 and 77.

d. Greater than 77.

**Answer**:

1. Z = = -1.2649

P(74) = 0.3962

P(Ms<74) = 0.5 – P(74)

= 0.5 - 0.3962

= 0.1038

1. Z = = -1.2649

P(74) = 0.3962

Z = = 1.2649

P(76) = 0.3962

P(74<Ms <76) = P(76) + P(74)

= 0.3962 + 0.3962

= 0.7924

1. Z = = -2.530

P(77) = 0.4943

Z = = 1.2649

P(76) = 0.3962

P(76<Ms<77) = P(77) - P(76)

= 0.4943 – 0.3962

= 0.0981

1. Z = = -2.530

P(77) = 0.4943

P(Ms>77) = 0.5000 - P(77)

= 0.5000 - 0.4943

= 0.0057

Page 272, Problem 17 (1 pt)

The rent for a one-bedroom apartment in Southern California follows the normal distribution with a mean of $2,200 per month and a standard deviation of $250 per month. The distribution of the monthly costs does not follow the normal distribution. In fact, it is positively skewed. What is the probability of selecting a sample of 50 one-bedroom apartments and finding the mean to be at least $1,950 per month?

**Answer**:

1. Z = = -7.0710

P(>=1950) = ~1

(This value is almost equal to 1 as the z score obtained is an outlier for the distribution)

Page 272, Problem 18 (2 pts)

According to an IRS study, it takes a mean of 330 minutes for taxpayers to prepare, copy, and electronically file a 1040 tax form. This distribution of times follows the normal distribution and the standard deviation is 80 minutes. A consumer watchdog agency selects a random sample of 40 taxpayers.

a. What is the standard error of the mean in this example?

b. What is the likelihood the sample mean is greater than 320 minutes?

c. What is the likelihood the sample mean is between 320 and 350 minutes?

d. What is the likelihood the sample mean is greater than 350 minutes?

**Answer**:

1. Standard error of the mean = = =12.6491
2. Z = = -0.7906

P(320) = 0.2852

P(>320) = 0.5000 + P(320)

= 0.5000 + 0.2852

= 0.7852

1. Z = = -0.7906

P(320) = 0.2852

Z = = 1.5811

P(350) = 0.4429

P(between 320 and 350) = P(320) + P(350)

= 0.2852+0.4429

= 0.7281

1. Z = = 1.5811

P(350) = 0.4429

P(>350) = 0.5000 - P(320)

= 0.5000 - 0.4429

= 0.0571

Page 276, Problem 35 (2 pts)

The mean age at which men in the United States marry for the first time follows the normal distribution with a mean of 24.8 years. The standard deviation of the distribution is 2.5 years. For a random sample of 60 men, what is the likelihood that the age at which they were married for the first time is less than 25.1 years?

**Answer**:

Z = = 0.9295

P(25.1) = 0.3238

P(<25.1) = 0.5000 + P(25.1)

= 0.5000 + 0.3238

= 0.8238

Page 276, Problem 36 (2 pts)

A recent study by the Greater Los Angeles Taxi Drivers Association showed that the mean fare charged for service from Hermosa Beach to Los Angeles International Airport is $21 and the standard deviation is $3.50. We select a sample of 15 fares.

a. What is the likelihood that the sample mean is between $20 and $23?

b. What must you assume to make the above calculation?

**Answer**:

Z = = -1.1066

P(20) = 0.3665

Z = = 2.2132

P(23) = 0.4864

P(between 20 and 23) = P(20) + P(23)

= 0.3665 + 0.4864

= 0.8529

1. To make the above calculation, we must assume that the population follows a normal distribution. And also that the population standard deviation can be applied for calculating the z-value of a sample distribution

**Part II:** The following problems require you to use SAS to perform the indicated tasks. The data sets are available under the “Data Sets” area in the Canvas Files section. You should copy the respective CSV files to your local workstation, Box account, or USB Flash Drive and then use SAS to import the data. You must submit an electronic copy of the relevant pages of your SAS output when you turn in the exercise.

1. For the variables Number of Students (Students), the Percent of Families Receiving Welfare (Welfare), the Amount Spent per Pupil (Instruct), and Average Instructor Salary (Salary) in the “SCHOOLS.CSV” data set, find the mean, median, mode, range, variance, standard deviation, and standard error. (1 pt)

title "V506 HOMEWORK02 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**data** hw02.mydata;

set hw02.schools;

**run**;

**proc** **univariate** data=hw02.mydata;

var Students Welfare Instruct Salary;

**run**;

**PROC** **PRINT**;

**RUN**;

**The UNIVARIATE Procedure**

**Variable: Students**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 94 | **Sum Weights** | 94 |
| **Mean** | 2134.05319 | **Sum Observations** | 200601 |
| **Std Deviation** | 3895.04411 | **Variance** | 15171368.6 |
| **Skewness** | 7.80213605 | **Kurtosis** | 68.899592 |
| **Uncorrected SS** | 1839030485 | **Corrected SS** | 1410937281 |
| **Coeff Variation** | 182.518605 | **Std Error Mean** | 401.743125 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 2134.053 | **Std Deviation** | 3895 |
| **Median** | 1226.500 | **Variance** | 15171369 |
| **Mode** | . | **Range** | 36770 |
|  |  | **Interquartile Range** | 1482 |

**The UNIVARIATE Procedure**

**Variable: Welfare**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 94 | **Sum Weights** | 94 |
| **Mean** | 7.23297872 | **Sum Observations** | 679.9 |
| **Std Deviation** | 6.52225946 | **Variance** | 42.5398685 |
| **Skewness** | 2.91648029 | **Kurtosis** | 11.5166209 |
| **Uncorrected SS** | 8873.91 | **Corrected SS** | 3956.20777 |
| **Coeff Variation** | 90.1739063 | **Std Error Mean** | 0.6727197 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 7.232979 | **Std Deviation** | 6.52226 |
| **Median** | 5.350000 | **Variance** | 42.53987 |
| **Mode** | 3.900000 | **Range** | 42.60000 |
|  |  | **Interquartile Range** | 4.80000 |

**The UNIVARIATE Procedure**

**Variable: Instruct**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 94 | **Sum Weights** | 94 |
| **Mean** | 2724.60638 | **Sum Observations** | 256113 |
| **Std Deviation** | 1095.21792 | **Variance** | 1199502.28 |
| **Skewness** | 6.26024991 | **Kurtosis** | 43.9989606 |
| **Uncorrected SS** | 809360827 | **Corrected SS** | 111553712 |
| **Coeff Variation** | 40.1972895 | **Std Error Mean** | 112.963103 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 2724.606 | **Std Deviation** | 1095 |
| **Median** | 2508.500 | **Variance** | 1199502 |
| **Mode** | 2250.000 | **Range** | 9310 |
|  |  | **Interquartile Range** | 400.00000 |

**The UNIVARIATE Procedure**

**Variable: Salary**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 94 | **Sum Weights** | 94 |
| **Mean** | 33181.4043 | **Sum Observations** | 3119052 |
| **Std Deviation** | 3548.50437 | **Variance** | 12591883.3 |
| **Skewness** | 0.6702765 | **Kurtosis** | 0.27556739 |
| **Uncorrected SS** | 1.04666E11 | **Corrected SS** | 1171045145 |
| **Coeff Variation** | 10.6942562 | **Std Error Mean** | 366.00028 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 33181.40 | **Std Deviation** | 3549 |
| **Median** | 32708.00 | **Variance** | 12591883 |
| **Mode** | . | **Range** | 17131 |
|  |  | **Interquartile Range** | 4680 |

2. Create a new variable called “Amt\_Per\_Pupil\_Std\_Score” which is equal to the Amount Spent per Pupil (Instruct) converted to standard scores (i.e., z-scores) and print out the original and new variable values. Then interpret the highest and lowest values for the Amt\_Per\_Pupil\_Std\_Score. (2 pts)

title "V506 HOMEWORK02 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**data** hw02.mydata;

set hw02.schools;

**run**;

**data** schools;

set hw02.mydata;

Amt\_Per\_Pupil\_Std\_Score = (Instruct - **2724.606**)/**1095** ;

**run**;

**PROC** **UNIVARIATE**;

VAR Instruct Amt\_Per\_Pupil\_Std\_Score;

**RUN**;

**PROC** **PRINT**;

**RUN**;

|  |
| --- |
| **V506 HOMEWORK02 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

| **Obs** | **Schools** | **Students** | **Income** | **Property** | **Welfare** | **Salary** | **Instruct** | **Attend** | **Passing** | **Amt\_Per\_Pupil\_Std\_Score** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | Bluffton | 1132 | 24487 | 62678 | 1.8 | 31221 | 2130 | 95.7 | 85 | -0.54302 |
| **2** | Shawnee | 2472 | 29777 | 130910 | 2.6 | 34860 | 2570 | 94.7 | 73 | -0.14119 |
| **3** | Spencerville | 1026 | 23161 | 51645 | 5.4 | 30155 | 2262 | 95.5 | 68 | -0.42247 |
| **4** | Delphos | 1104 | 21792 | 88453 | 6.2 | 32273 | 2506 | 96.5 | 65 | -0.19964 |
| **5** | Elida | 3204 | 24446 | 65550 | 8.9 | 32876 | 2250 | 94.1 | 62 | -0.43343 |
| **6** | Lima | 5963 | 18394 | 44138 | 33.8 | 33142 | 2657 | 92.3 | 40 | -0.06174 |
| **7** | Northeastern | 1194 | 26428 | 88789 | 1.7 | 30919 | 2431 | 96.1 | 72 | -0.26813 |
| **8** | Ayersville | 921 | 28228 | 82707 | 3.9 | 32850 | 2693 | 95.6 | 68 | -0.02886 |
| **9** | Defiance | 3046 | 23812 | 56333 | 11.2 | 34750 | 2438 | 94.2 | 63 | -0.26174 |
| **10** | Hicksville | 990 | 22448 | 56411 | 7 | 34224 | 2351 | 95.7 | 59 | -0.34119 |
| **11** | Central | 1216 | 24189 | 53923 | 4.6 | 34430 | 2496 | 94.8 | 56 | -0.20877 |
| **12** | Berlin-Milan | 1593 | 25223 | 76878 | 4 | 32166 | 2564 | 96.1 | 77 | -0.14667 |
| **13** | Perkins | 2038 | 25586 | 117545 | 3.3 | 39352 | 2861 | 95.8 | 74 | 0.12456 |
| **14** | Huron | 1494 | 27135 | 105588 | 4.6 | 33433 | 2968 | 95.4 | 74 | 0.22228 |
| **15** | Margaretta | 1560 | 23849 | 74601 | 2.8 | 37084 | 2464 | 95.5 | 66 | -0.23800 |
| **16** | Sandusky | 4426 | 19529 | 72425 | 25.2 | 36042 | 2766 | 93 | 37 | 0.03780 |
| **17** | Kelleys Island | 20 | 19854 | 802081 | 11.4 | 27144 | 11226 | 95 | 100 | 7.76383 |
| **18** | Pettisville | 503 | 25079 | 53948 | 3.9 | 31159 | 2834 | 96.1 | 78 | 0.09990 |
| **19** | Wauseon | 1864 | 23408 | 60896 | 4.9 | 32499 | 2252 | 95.3 | 75 | -0.43160 |
| **20** | Evergreen | 1238 | 23826 | 69432 | 5.2 | 32353 | 2250 | 95 | 72 | -0.43343 |
| **21** | Archbold-Area | 1401 | 26706 | 107547 | 4.1 | 35982 | 2837 | 96.2 | 69 | 0.10264 |
| **22** | Pike-Delta-Yor | 1559 | 23396 | 48638 | 6.1 | 31310 | 2309 | 94.8 | 66 | -0.37955 |
| **23** | Gorham-Fayette | 487 | 22405 | 57221 | 6.9 | 33166 | 2492 | 94.6 | 51 | -0.21243 |
| **24** | Swanton | 1725 | 24596 | 69320 | 6.3 | 33690 | 2615 | 94.9 | 50 | -0.10010 |
| **25** | Arlington | 685 | 26175 | 55478 | 2.7 | 31821 | 2205 | 96.5 | 84 | -0.47453 |
| **26** | Vanlue | 346 | 24709 | 49606 | 4.1 | 28411 | 2420 | 96.2 | 83 | -0.27818 |
| **27** | Liberty-Benton | 954 | 28718 | 77503 | 1.5 | 30330 | 2063 | 96.7 | 78 | -0.60421 |
| **28** | Van Buren | 840 | 28964 | 151992 | 1.8 | 33447 | 2584 | 96.4 | 75 | -0.12841 |
| **29** | Cory-Rawson | 794 | 23904 | 69242 | 2.4 | 31241 | 2416 | 96.4 | 73 | -0.28183 |
| **30** | Arcadia | 597 | 24305 | 78102 | 2.5 | 30738 | 2752 | 95.9 | 64 | 0.02502 |
| **31** | McComb | 805 | 23754 | 69347 | 5.7 | 28986 | 2321 | 95.9 | 61 | -0.36859 |
| **32** | Findlay | 5758 | 24269 | 92648 | 7.6 | 35879 | 2860 | 94.9 | 60 | 0.12365 |
| **33** | Ada | 855 | 23029 | 52655 | 5.3 | 28479 | 2380 | 95.9 | 69 | -0.31471 |
| **34** | Kenton | 2228 | 20418 | 61155 | 13.8 | 30907 | 2512 | 93.7 | 54 | -0.19416 |
| **35** | Liberty Center | 1009 | 24723 | 57685 | 4.1 | 30904 | 2431 | 95.8 | 82 | -0.26813 |
| **36** | Patrick Henry | 1176 | 23061 | 63134 | 5.2 | 31895 | 2552 | 95.9 | 75 | -0.15763 |
| **37** | Napoleon Area | 2331 | 25304 | 84245 | 6.6 | 32773 | 2422 | 94.6 | 73 | -0.27635 |
| **38** | Holgate | 605 | 23962 | 49709 | 6 | 31324 | 2454 | 94.9 | 71 | -0.24713 |
| **39** | Monroeville | 686 | 22942 | 63103 | 4.3 | 30838 | 2474 | 95.8 | 64 | -0.22886 |
| **40** | Bellevue | 2276 | 24025 | 66912 | 6.9 | 32164 | 2374 | 95.1 | 55 | -0.32019 |
| **41** | Willard | 2300 | 23304 | 58832 | 14.3 | 35042 | 2347 | 94.6 | 53 | -0.34485 |
| **42** | Norwalk | 2650 | 21551 | 72266 | 12.2 | 37145 | 2384 | 94 | 50 | -0.31106 |
| **43** | Ottawa Hills | 933 | 45723 | 122356 | 0.2 | 43256 | 4150 | 95.7 | 95 | 1.30173 |
| **44** | Anthony Wayne | 3178 | 29215 | 88004 | 3.1 | 35617 | 2844 | 95.3 | 75 | 0.10904 |
| **45** | Sylvania | 7822 | 32114 | 101503 | 3.8 | 39684 | 2943 | 95.5 | 72 | 0.19945 |
| **46** | Maumee | 3009 | 27604 | 117921 | 3.9 | 41634 | 3933 | 95 | 69 | 1.10356 |
| **47** | Oregon | 3594 | 24525 | 123599 | 8.7 | 35848 | 2941 | 94.7 | 52 | 0.19762 |
| **48** | Washington | 7154 | 23507 | 102485 | 11.8 | 39155 | 2997 | 93.7 | 51 | 0.24876 |
| **49** | Springfield | 3575 | 26048 | 98346 | 12.2 | 34437 | 2774 | 93.8 | 49 | 0.04511 |
| **50** | Toledo | 36790 | 21079 | 62668 | 42.8 | 36190 | 2611 | 90.7 | 28 | -0.10375 |
| **51** | Benton Carroll | 2063 | 23899 | 237206 | 4.7 | 42734 | 3444 | 95.6 | 81 | 0.65698 |
| **52** | Danbury | 635 | 21325 | 182360 | 3.2 | 34971 | 3158 | 95.3 | 65 | 0.39579 |
| **53** | Genoa | 1584 | 25321 | 53120 | 4.8 | 34661 | 2845 | 95.9 | 65 | 0.10995 |
| **54** | Port Clinton | 2238 | 20941 | 129961 | 10.5 | 39542 | 2926 | 94.5 | 50 | 0.18392 |
| **55** | Put-in-Bay | 70 | 19266 | 426419 | 1.4 | 30242 | 7824 | 94.5 | 33 | 4.65698 |
| **56** | Paulding | 1993 | 22677 | 46163 | 8.6 | 32928 | 2560 | 94.5 | 59 | -0.15033 |
| **57** | Ottoville | 610 | 24128 | 46582 | 0.2 | 26125 | 2588 | 99.8 | 86 | -0.12475 |
| **58** | Columbus Grove | 866 | 23562 | 55568 | 8.4 | 30476 | 2174 | 96.2 | 84 | -0.50284 |
| **59** | Kalida | 775 | 24456 | 44267 | 0.7 | 28962 | 2274 | 96.8 | 81 | -0.41151 |
| **60** | Continental | 792 | 23625 | 37277 | 10 | 28945 | 2225 | 95.5 | 79 | -0.45626 |
| **61** | Ottawa-glandor | 1749 | 25363 | 64288 | 6.5 | 31185 | 2154 | 96.5 | 77 | -0.52110 |
| **62** | Pandora-Gilboa | 632 | 23806 | 55446 | 1.2 | 27693 | 2078 | 96.5 | 67 | -0.59051 |
| **63** | Leipsic | 748 | 20941 | 62648 | 16.1 | 30282 | 2811 | 93.9 | 47 | 0.07890 |
| **64** | Gibsonburg | 983 | 23312 | 46098 | 17.8 | 31244 | 2242 | 94.4 | 71 | -0.44074 |
| **65** | Lakota | 1332 | 22678 | 55933 | 10.5 | 30765 | 2306 | 95.2 | 66 | -0.38229 |
| **66** | Fremont | 5156 | 22327 | 74874 | 16.2 | 37759 | 2616 | 94.7 | 57 | -0.09918 |
| **67** | Woodmore | 1141 | 26460 | 90484 | 10.9 | 32296 | 2227 | 96.5 | 56 | -0.45443 |
| **68** | Clyde-Green Sp | 2368 | 23854 | 55724 | 7 | 33998 | 2383 | 94.6 | 47 | -0.31197 |
| **69** | Bettsville | 347 | 22103 | 37269 | 6.1 | 27466 | 2394 | 95.8 | 83 | -0.30192 |
| **70** | Seneca East | 1183 | 22656 | 50895 | 4.8 | 29940 | 2435 | 95.9 | 80 | -0.26448 |
| **71** | Old Fort | 540 | 23208 | 50712 | 19.3 | 28195 | 2743 | 96 | 78 | 0.01680 |
| **72** | Hopewell-Loudo | 870 | 22103 | 72201 | 4.7 | 30644 | 2564 | 94.6 | 74 | -0.14667 |
| **73** | New Riegel | 459 | 23314 | 41376 | 2.7 | 29099 | 2501 | 97 | 73 | -0.20421 |
| **74** | Tiffin | 3632 | 21246 | 65291 | 9 | 35513 | 2506 | 94.7 | 67 | -0.19964 |
| **75** | Fostoria | 2742 | 20809 | 62268 | 20.4 | 34241 | 2455 | 92.7 | 34 | -0.24622 |
| **76** | Van Wert | 2504 | 22728 | 67932 | 7.1 | 33885 | 2511 | 95.3 | 61 | -0.19507 |
| **77** | Edon-Northwest | 744 | 23035 | 38462 | 2.7 | 30833 | 1916 | 95.1 | 73 | -0.73845 |
| **78** | Milcreek-West | 788 | 21302 | 40239 | 3.2 | 31582 | 2382 | 95.9 | 71 | -0.31288 |
| **79** | Bryan | 2266 | 22607 | 81152 | 4.2 | 32643 | 2706 | 95.2 | 69 | -0.01699 |
| **80** | North Central | 757 | 21871 | 59396 | 2.9 | 31978 | 2349 | 95.6 | 68 | -0.34302 |
| **81** | Montpelier | 1172 | 20787 | 44383 | 8.2 | 33243 | 2654 | 94.8 | 67 | -0.06448 |
| **82** | Edgerton | 767 | 22429 | 54040 | 4.1 | 28975 | 2470 | 95.9 | 61 | -0.23252 |
| **83** | Stryker | 579 | 24084 | 65532 | 3.4 | 33855 | 2617 | 95.7 | 47 | -0.09827 |
| **84** | Perrysburg | 3839 | 32773 | 97888 | 2.6 | 40320 | 3011 | 96.1 | 98 | 0.26155 |
| **85** | Elmwood | 1237 | 22179 | 47644 | 7.5 | 30434 | 2643 | 95.2 | 68 | -0.07453 |
| **86** | Bowling Green | 3534 | 21307 | 84682 | 7.5 | 37983 | 2849 | 94.7 | 68 | 0.11360 |
| **87** | Otsego | 1643 | 24614 | 57601 | 5 | 36065 | 2539 | 95.4 | 64 | -0.16950 |
| **88** | Northwood | 1091 | 25905 | 77077 | 7.9 | 35536 | 2979 | 94.9 | 61 | 0.23232 |
| **89** | Eastwood | 1739 | 25043 | 67929 | 4.7 | 35742 | 2499 | 95.9 | 60 | -0.20603 |
| **90** | Lake | 1665 | 23559 | 95859 | 6.2 | 38046 | 2820 | 95.2 | 55 | 0.08712 |
| **91** | Rossford | 2087 | 25360 | 123725 | 7.2 | 39476 | 3258 | 94.9 | 54 | 0.48712 |
| **92** | North Baltimor | 839 | 22075 | 58383 | 10.6 | 29579 | 2331 | 94.5 | 50 | -0.35946 |
| **93** | Upper Sandusky | 1801 | 21063 | 68348 | 3.9 | 32778 | 2267 | 95.6 | 62 | -0.41791 |
| **94** | Carey | 915 | 21658 | 51497 | 6 | 30968 | 2513 | 95.5 | 59 | -0.19325 |

* Interpretation of highest and lowest values for the variable **Amt\_Per\_Pupil\_Std\_Score**. Mean of the variable is 3.49752E-7. This value is very close to 0. The below table gives us the extreme values for the variable **Amt\_Per\_Pupil\_Std\_Score**.

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| -0.738453 | 77 | 0.656981 | 51 |
| -0.604206 | 27 | 1.103556 | 46 |
| -0.590508 | 62 | 1.301730 | 43 |
| -0.543019 | 1 | 4.656981 | 55 |
| -0.521101 | 61 | 7.763830 | 17 |

* The highest and lowest values are highlighted in the above table. The lowest value is closer to the mean than the highest value that is present in the dataset. In fact, the standard deviation of the variable is 1.00019901. And the skewness is 6.26024991. We can interpret that the distribution of the variable is positively-skewed.

3. Compare the standard deviation and standard error for Average Instructor Salary, and explain why they are different (1 pt).

**Variable: Instruct**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 94 | **Sum Weights** | 94 |
| **Mean** | 2724.60638 | **Sum Observations** | 256113 |
| **Std Deviation** | 1095.21792 | **Variance** | 1199502.28 |
| **Skewness** | 6.26024991 | **Kurtosis** | 43.9989606 |
| **Uncorrected SS** | 809360827 | **Corrected SS** | 111553712 |
| **Coeff Variation** | 40.1972895 | **Std Error Mean** | 112.963103 |

The Standard deviation of the sample is the degree to which individuals within the sample differ from the sample mean. In the above scenario we are making use of the population dataset itself the standard deviation of the sample is the standard deviation of the population ().

Standard distribution of sample (s) = Standard distribution of sample ()

The Standard error is applied on a sample distribution of a population. It gives us an idea of the error that is obtained in calculating the mean of the population () based on the mean of the sample distribution (s). In the above scenario as we can see the standard error is less (112.963103). This observation is minimum as we are testing our observations on the population dataset itself.

Standard Error of sample =

The difference in the values standard deviation and standard error is understandable as we are not doing a sample distribution in the above case. The distribution is for population data set. The standard deviation in the above case is the population standard deviation (). The standard error is actually meaningless in this case as the sample mean and population mean should be one and the same in the above case.

Standard Error of sample = = = 112.963103 (As obtained)

4. Using the REAL-ESTATE-2003.CSV data set, create two new variables called “Price\_Per\_Bed” and “Price\_Per\_Bath”, which represent Selling Price per Number of Bedrooms and Selling Price per Number of Bathrooms, respectively. Then find the mean, median, mode, range, variance, standard deviation, and standard error of these new variables. (2 pts)

title "V506 HOMEWORK02 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**data** hw02.mydata;

set hw02.RealEstate;

**run**;

**data** RealE;

set hw02.mydata;

Price\_Per\_Bed = Price / Bedrooms;

Price\_Per\_Bath = Price / Baths;

**run**;

**PROC** **UNIVARIATE**;

VAR Price\_Per\_Bed Price\_Per\_Bath Distance;

**RUN**;

**The UNIVARIATE Procedure**

**Variable: Price\_Per\_Bed**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 105 | **Sum Weights** | 105 |
| **Mean** | 65.7641848 | **Sum Observations** | 6905.2394 |
| **Std Deviation** | 25.2525004 | **Variance** | 637.688774 |
| **Skewness** | 1.00996979 | **Kurtosis** | 0.66536814 |
| **Uncorrected SS** | 520437.073 | **Corrected SS** | 66319.6325 |
| **Coeff Variation** | 38.3985606 | **Std Error Mean** | 2.46439169 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 65.76418 | **Std Deviation** | 25.25250 |
| **Median** | 59.66667 | **Variance** | 637.68877 |
| **Mode** | 31.38333 | **Range** | 122.43333 |
|  |  | **Interquartile Range** | 31.72500 |

**The UNIVARIATE Procedure**

**Variable: Price\_Per\_Bath**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 105 | **Sum Weights** | 105 |
| **Mean** | 108.336476 | **Sum Observations** | 11375.33 |
| **Std Deviation** | 24.355649 | **Variance** | 593.197636 |
| **Skewness** | 0.43735698 | **Kurtosis** | -0.5278742 |
| **Uncorrected SS** | 1294055.72 | **Corrected SS** | 61692.5541 |
| **Coeff Variation** | 22.4814853 | **Std Error Mean** | 2.37686796 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 108.3365 | **Std Deviation** | 24.35565 |
| **Median** | 104.8500 | **Variance** | 593.19764 |
| **Mode** | 89.5000 | **Range** | 109.88333 |
|  |  | **Interquartile Range** | 36.98333 |

**The UNIVARIATE Procedure**

**Variable: Distance**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 105 | **Sum Weights** | 105 |
| **Mean** | 14.6285714 | **Sum Observations** | 1536 |
| **Std Deviation** | 4.8739045 | **Variance** | 23.7549451 |
| **Skewness** | 0.40190999 | **Kurtosis** | -0.1736331 |
| **Uncorrected SS** | 24940 | **Corrected SS** | 2470.51429 |
| **Coeff Variation** | 33.3177065 | **Std Error Mean** | 0.47564438 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 14.62857 | **Std Deviation** | 4.87390 |
| **Median** | 15.00000 | **Variance** | 23.75495 |
| **Mode** | 16.00000 | **Range** | 22.00000 |
|  |  | **Interquartile Range** | 7.00000 |

5. Compare the standard deviation and standard error for the variable representing Distance from the City Center, and explain why they are different. (1.5 pts)

**The UNIVARIATE Procedure**

**Variable: Distance**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 105 | **Sum Weights** | 105 |
| **Mean** | 14.6285714 | **Sum Observations** | 1536 |
| **Std Deviation** | 4.8739045 | **Variance** | 23.7549451 |
| **Skewness** | 0.40190999 | **Kurtosis** | -0.1736331 |
| **Uncorrected SS** | 24940 | **Corrected SS** | 2470.51429 |
| **Coeff Variation** | 33.3177065 | **Std Error Mean** | 0.47564438 |

The Standard deviation of the sample is the degree to which individuals within the sample differ from the sample mean. In the above scenario we are making use of the population dataset itself the standard deviation of the sample is the standard deviation of the population ().

Standard distribution of sample (s) = Standard distribution of sample ()

The Standard error is applied on a sample distribution of a population. It gives us an idea of the error that is obtained in calculating the mean of the population () based on the mean of the sample distribution (s). In the above scenario as we can see the standard error is fractional (0.47564438). This observation is minimum as we are testing our observations on the population dataset itself.

Standard Error of sample =

The difference in the values standard deviation and standard error is understandable as we are not doing a sample distribution in the above case. The distribution is for population data set. The standard deviation in the above case is the population standard deviation (). The standard error is actually meaningless in this case as the sample mean and population mean should be one and the same in the above case.

Standard Error of sample = = = 0.47564438 (As obtained)