**V506 Homework 1 Answers**

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

**Part I (37 points):**

Problems 4, 5, 7, 8, 12, 14, and 15 are worth 3 points each; and all other problems are each worth 2 points.

1. (Page 55, #5) Compute the mean of the following sample values: 16.25, 12.91, 14.58

**Answer:**

*Mean = =* ***14.58***

1. The worldwide airline fatalities since 1980, according to the National Safety Council, are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year** | **Number of Passenger Deaths** |  | **Year** | **Number of Passenger Deaths** |
| 1980 | 814 |  | 1987 | 890 |
| 1981 | 362 |  | 1988 | 699 |
| 1982 | 764 |  | 1989 | 817 |
| 1983 | 809 |  | 1990 | 440 |
| 1984 | 223 |  | 1991 | 510 |
| 1985 | 1,066 |  | 1992 | 990 |
| 1986 | 331 |  |  |  |

What is the annual arithmetic mean number of passenger deaths worldwide since 1980?

**Answer:**

*Mean = =* ***670.385***

1. (Page 65, #24) The Bookstall, Inc., is a specialty bookstore concentrating on used books sold via the internet. Paperbacks are $1.00 each, and hardcover books are $3.50. Of the 50 books sold Tuesday morning, 40 were paperback and the rest were hardcover. What was the weighted mean price of a book?

**Answer:**

Weighted mean price of a book = = ***1.5***

1. (Page 84, #60) SCCoast, an internet provider in the Southeast, developed the following frequency distribution on the age of internet users. Estimate the mean and standard deviation

|  |  |
| --- | --- |
| **Age (years)** | **Frequency** |
| 10 up to 20 | 3 |
| 20 up to 30 | 7 |
| 30 up to 40 | 18 |
| 40 up to 50 | 20 |
| 50 up to 60 | 12 |

**Answer:**

In the above problem first we shall use the mean of individual to find the total age of an age group after that we find the total age of the entire group. In the end we shall divide this value with our total frequency.

We get the following table once we find the mean of each value range.

|  |  |  |
| --- | --- | --- |
| **Age (years)** | **Mean of Age Group** | **Frequency** |
| 10 up to 20 | 15 | 3 |
| 20 up to 30 | 25 | 7 |
| 30 up to 40 | 35 | 18 |
| 40 up to 50 | 45 | 20 |
| 50 up to 60 | 55 | 12 |

*Mean (X) =*

*=*

*=* ***40.167***

Once we have got the median we can work on the standard deviation now,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Age (years)** | **Mean (M)** | **Frequency (f)** | **(M – X)** | **f(M – X)2** |
| 10 up to 20 | 15 | 3 | -25.167 | 1900.134 |
| 20 up to 30 | 25 | 7 | -15.167 | 1610.265 |
| 30 up to 40 | 35 | 18 | -5.167 | 480.562 |
| 40 up to 50 | 45 | 20 | 4.833 | 467.158 |
| 50 up to 60 | 55 | 12 | 14.833 | 2640.215 |
|  |  | 60 |  | 7098.334 |

Standard deviation = = **= *10.969***

1. (Page 87, #65) A sample of households that subscribe to the United Bell Phone Company for land line phone service revealed the following numbers of calls received per household last week. Determine the mean and the median number of calls received.

52 43 30 38 30 42 12 46 39 37 34 46 32 18 41 5

**Answer: =**

*Mean = =* ***34.0625***

Arranging the above numbers in an ascending order we get,

5 12 18 30 30 32 34 **37 38** 39 42 41 43 46 52

As there are 16 numbers in the above list we will be considering the 8th and 9th numbers in the above order to calculate the median.

Median *= =* ***37.5***

1. (Page 75, #42) Consider these six values a population: 13, 3, 8, 10, 8, and 6.
   1. Determine the mean of the population.
   2. Determine the variance.

**Answer:**

1. *Mean = =* ***8***
2. Variance =  *=* ***9.67***
3. (Page 91, #85) Bidwell Electronics, Inc. recently surveyed a sample of employees to determine how far they lived from corporate headquarters. The results are shown below. Compute the mean and the standard deviation. Interpret your results.

|  |  |  |
| --- | --- | --- |
| **Distance (Miles)** | **Frequency** | **M** |
| 0 up to 5 | 4 | 2.5 |
| 5 up to 10 | 15 | 7.5 |
| 10 up to 15 | 27 | 12.5 |
| 15 up to 20 | 18 | 17.5 |
| 20 up to 25 | 6 | 22.5 |

**Answer:**

*Mean (X) =*

*= =*

*=* ***13***

Once we have got the median we can work on the standard deviation now,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Age (years)** | **Frequency (f)** | **Mean (M)** | **(M – X)** | **(M – X)2** | **f(M – X)2** |
| 0 up to 5 | 4 | 2.5 | -10.5 | 110.25 | 441 |
| 5 up to 10 | 15 | 7.5 | -5.5 | 30.25 | 453.75 |
| 10 up to 15 | 27 | 12.5 | -0.5 | 0.25 | 6.75 |
| 15 up to 20 | 18 | 17.5 | 4.5 | 20.25 | 364.5 |
| 20 up to 25 | 6 | 22.5 | 9.5 | 90.25 | 541.5 |
|  | 70 |  |  |  | 1807.5 |

Standard deviation = = **= *5.1182***

**Interpretation:**

* On an average the employees of Bidwell Electronics, Inc. live 13 miles away from the headquarters. However we have a standard deviation of 5.1183 miles. Thus we can give the below interpretation.
* 68% live around -> in the range of [13 + 5.1183] miles to [13 - 5.1183] miles away from the headquarters.
* 95% live around -> in the range of [13 + (2 x 5.1183)] miles to [13 - (2 x 5.1183)] miles away from the headquarters.
* 99.7% live around -> in the range of [13 + (3 x 5.1183] miles to [13 - (2 x 5.1183)] miles away from the headquarters.

1. (Page 145, #11) The events A and B are mutually exclusive. Suppose P(A)=0.30 and P(B)=0.20. What is the probability of either A or B occurring? What is the probability that neither A nor B will happen?

**Answer:**

The events A and B are mutually exclusive. Here special rule of addition is applied.

P(A)=0.30 and P(B)=0.20

P(A or B) = 0.30 + 0.20 (Since the events are mutually exclusive)

P(A or B) = 0.50

P(neither A nor B) = 1 – P(A or B) = 1 – 0.50 = 0.50

1. (Page 146 #18) Let P(X) = 0.55 and P(Y) = 0.35. Assume that the probability that they both occur is 0.20. What is the probability of either X or Y occurring?

**Answer:**

P(X) = 0.55 and P(Y) = 0.35

P(X and Y) = P(X U Y) = 0.20

P(X U Y) = P(X) + P(Y) - P(X ∩ Y)

Hence, P(X ∩ Y) = P(X) + P(Y) - P(X U Y)

= 0.55 + 0.35 – 0.20

P(X ∩ Y) = 0.70

1. (Page 146, #20) A student is taking two courses, history and math. The probability that the student will pass the history course is 0.60, and the probability of passing the math course is 0.70. The probability of passing both is 0.50. What is the probability of passing at least one?

**Answer:**

Probability of student passing in history = P(H) = 0.60

Probability of student passing in math = P(M) = 0.70

P(H ∩ M) = 0.50

P(H U M) = P(H) + P(M) - P(H ∩ M)

= 0.60 + 0.70 – 0.50

P(H U M) = 0.80

1. (Page 155, #27) Refer to the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Second Event** | **First Event** | | | **Total** |
| **A1** | **A2** | **A3** |
| B1 | 2 | 1 | 3 | 6 |
| B2 | 1 | 2 | 1 | 4 |
| **Total** | 3 | 3 | 4 | 10 |

* 1. Determine P(A1).
  2. Determine P(B1|A2).
  3. Determine P(B2 and A3).

**Answer:**

1. P(A1) = 3/10
2. P(B1|A2) = 1/3
3. P(B2 and A3) = P(B2)P(A3|B2) = 4/10 x 1/4 = 1/10
4. (Page 165, #41) A pollster randomly selected 4 of 10 available people. How many different groups of 4 are possible?

**Answer:**

Order doesn’t matter here, so combination is applied.

n = 10 and r = 4

=  *= = 210*

1. (Page 165, #46) A company is creating three new divisions and seven managers are eligible to be appointed head of a division. How many different ways could the three new heads be appointed? Hint: Assume the division assignment makes a difference.

**Answer:**

Order doesn’t matter here, so permutation is applied.

n = 7 and r = 3

=  *= = 210*

1. (Page 168, #56) Assume the likelihood that any flight on Delta Airlines arrives within 15 minutes of the scheduled time is 0.90. We randomly selected a Delta flight on four different days.
   1. What is the likelihood all four of the selected flights arrived within 15 minutes of the scheduled time?
   2. What is the likelihood that none of the selected flights arrived within 15 minutes of the scheduled time?
   3. What is the likelihood that at least one of the selected flights does not arrive within 15 minutes of the scheduled time?

**Answer:**

1. Four independent events, thus

P(four flights on schedule)

= P(one flight on schedule)4 = 0.9 x 0.9 x 0.9 x 0.9 = ***0.6561***

1. Four independent events, thus

P(four flights not on schedule)

= P(one flight not on schedule)4 = 0.1 x 0.1 x 0.1 x 0.1 = ***0.0001***

1. When you say atleast on flight doesn’t arrive on schedule we mean that there is atleast one of the 4 flights that is on schedule

P(atleast one on time) = 1 – P(No flight is on time) = 1 – 0.0001 ***= 0.9999***

1. (Page 169, #66) A recent survey reported in Bloomberg BusinessWeek dealt with the salaries of CEOs at large corporations and whether company shareholders made money or lost money.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CEO paid more than $1 million** | **CEO paid less than $1 million** | **Total** |
| Shareholders made money | 2 | 11 | 13 |
| Shareholders lost money | 4 | 3 | 7 |
| **Total** | 6 | 14 | 20 |

If a company is randomly selected from the list of 20 studied, what is the probability:

* 1. The CEO made more than $1 million?
  2. The CEO made more than $1 million or the shareholders lost money?
  3. The CEO made more than $1 million given the shareholders lost money?
  4. Of selecting 2 CEOs and finding they both made more than $1 million?

**Answer:**

1. P(CEO made more than $1 million) = =
2. P(CEO made more than $1 million or the shareholders lost money)

= + - =

1. P(CEO made more than $1 million given the shareholders lost money)

=

1. P(Of selecting 2 CEOs and finding they both made more than $1 million)

= x =

**Part II (13 points):** The following problems require you to use SAS to perform the indicated tasks. The data set is available in the “Data Sets” folder in the Canvas /Files/Data Sets section. You should copy the respective files to your workstation or a USB Flash Drive and then use SAS to import the data. You must submit a copy of the relevant pages of your SAS output when you turn in the exercise (histograms are relevant). Note that you should not include frequency distributions.

1. Import the “General Social Survey 2008” data set file in to SAS. Note that the file on Canvas is a Comma Separated Values file or .csv. Take a look at the dataset. You will notice that it contains dozens of variables and thousands of observations. You will only work with a few variables for this assignment, but we will look at this data again in the future. Below are relevant variable definitions, but you can find all variable descriptions in the General Social Survey Codebook.
2. Derive basic descriptive statistics including mean, median, mode, range, variance, and standard deviation for the following variables: **AGE**, **HRS1**, **TVHOURS**,and **EDUC**. (4 points)
3. Generate histograms for the following variables: **AGE**, **HAPMAR**, and **LIFE.** (3 points)
4. What does the histogram for **AGE** tell us? Is there anything distinctive about its shape? (1 point).
5. What do you notice about the histograms for **HAPMAR** and **LIFE**? (2 points)
6. Do people in the survey generally find their lives exciting, routine, or dull? (1 point)
7. How do people describe their marriages? (1 point)
8. Approximately what percent of individuals in the sample are older than 75 years? (1 point)

**Variables Definitions:**

**MARITAL:** Current marital status

1=married, 2=widowed, 3=divorced, 4=separated, 5=never married, 9=no answer

**EDUC:** Years of education

0=no formal education, 1-20=years of school (up to 8 years in university study), 98=don’t know, 99=no answer

**HAPMAR:** Taking things all together, how would you describe your marriage?

1=very happy, 2=pretty happy, 3=not too happy, 8=don’t know, 9=no answer, 0=not married

**LIFE:** Do you find your life exciting, pretty routine, or dull?

1=exciting, 2=routine, 3=dull, 8, 9, and 0= no answer or N/A

**AGE:** In years

**HRS1:** Number of hours worked the previous week (1-89 hours):

98=don’t know, 99=no answer, -1=not employed

**TVHOURS:** Total number of hours watched on a daily basis (0-24):

98=don’t know, 99=no answer, -1=not applicable

**Answer:**

1. We imported the “General Social Survey 2008” CSV data set file in to SAS.

Library used ‘hw01’ and data source name ‘pop’.

Code used:

title "V506 HOMEWORK01 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**data** hw01.gss2008cp;

set hw01.pop;

**run**;

**proc** **univariate** data=hw01.gss2008cp;

var AGE HRS1 TVHOURS EDUC hapmar life;

histogram age;

histogram hapmar;

histogram life;

**run**;

1. **Variable: AGE**

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Mean** | **47.96194** | **Std Deviation** | **17.67794** |
| **Median** | **47.00000** | **Variance** | **312.50943** |
| **Mode** | **50.00000** | **Range** | **81.00000** |
|  |  | **Interquartile Range** | **26.00000** |

**Variable: HRS1**

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Mean** | **25.21354** | **Std Deviation** | **24.52325** |
| **Median** | **30.00000** | **Variance** | **601.38959** |
| **Mode** | **-1.00000** | **Range** | **100.00000** |
|  |  | **Interquartile Range** | **43.00000** |

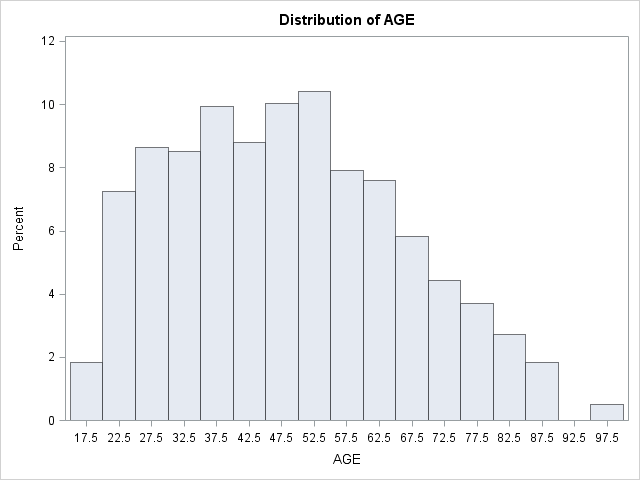
**Variable: TVHOURS**

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Mean** | **1.85072** | **Std Deviation** | **5.57801** |
| **Median** | **1.00000** | **Variance** | **31.11420** |
| **Mode** | **-1.00000** | **Range** | **99.00000** |
|  |  | **Interquartile Range** | **4.00000** |

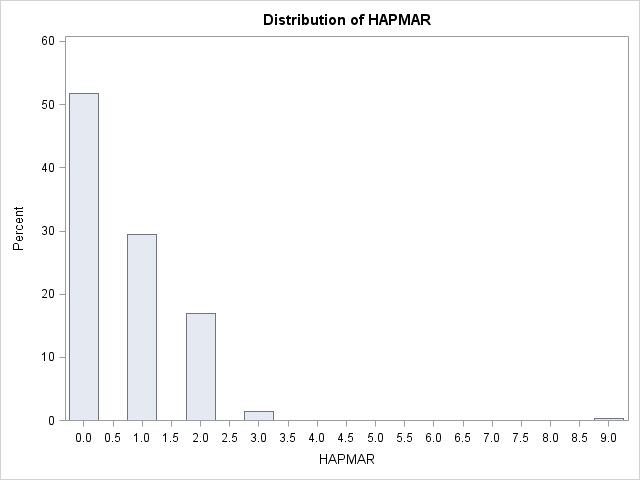
**Variable: EDUC**

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Mean** | **13.64212** | **Std Deviation** | **5.22164** |
| **Median** | **13.00000** | **Variance** | **27.26553** |
| **Mode** | **12.00000** | **Range** | **99.00000** |
|  |  | **Interquartile Range** | **4.00000** |

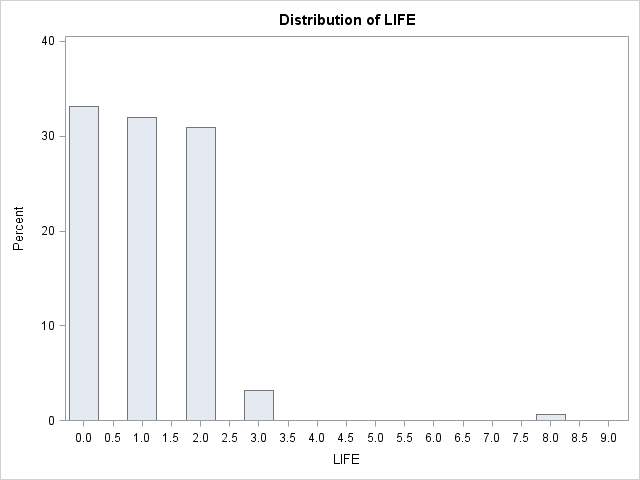
1. AGE



HAPMAR



LIFE



1. The histogram for **AGE** has a nonsymmetrical or skewed distribution, the relationship among the three measure changes. We can infer that respondents were from a younger age group. Following are some of the properties of the graph:

Mode: **50.00000** Median: **47.00000** Mean: **47.96194**

It is skewed towards the right. However the mean is not the largest value of the three measures. As per the textbook definition it does not satisfy the definition of a positively skewed distribution.

1. **HAPMAR: -**

The histogram of HAPMAR is a decreasing graph. It is definitely nonsymmetrical and heavily skewed towards the right. Following are some of the properties of the graph:

Mode: **0.00000** Median: **0.00000** Mean: **0.713297**

On the basis of the relationship among the three measures we can definitely conclude that this is a positively skewed histogram.

More than half of the people surveyed were not married. Out of the remaining ~47% people most of them rated that they were ‘very happy’ with their marriage. This was followed by people who were ‘pretty happy’ with their marriages.

A small percent of the surveyed people were ‘not happy’ with their marriages. Also there were a tiny proportion of surveyed individuals who choose not to answer the particular question in the survey.

We can conclude that most people did choose to give some kind of answer in this category and we get a rough estimate of the happiness of their marriages.

**LIFE: -**

The histogram of LIFE is again decreasing graph. It is definitely nonsymmetrical and heavily skewed towards the right. Following are some of the properties of the graph:

Mode: **0.00000** Median: **1.00000** Mean: **1.093426**

On the basis of the relationship among the three measures we can definitely conclude that this is a positively skewed histogram.

Most of the surveyed individuals choose not to answer or skipped the question altogether. This was about ~33% of surveyed people. Although 8,9 and 0 were of the same category a tiny proportion of people chose to give their opinions by answer in the ‘8’ category of the group.

Next an almost equal number of people rated their lives ‘exciting’, ~32%. Similarly an equal number of people, but less than the above two categories chose their life was ‘routine’, ~31%. Finally, a small number of surveyed people ~3% were of the opinion that their lives were ‘dull’.

1. People describe their marriages:

On the basis of HAPMAR histogram:

More than half of the people surveyed were not married. Out of the remaining ~47% people most of them rated that they were ‘very happy’ with their marriage. This was followed by people who were ‘pretty happy’ with their marriages.

A small percent of the surveyed people were ‘not happy’ with their marriages. Also there were a tiny proportion of surveyed individuals who choose not to answer the particular question in the survey.

We can conclude that most people did choose to give some kind of answer in this category and we get a rough estimate of the happiness of their marriages.

1. Percent of individuals in the sample are older than 75 years:

* We can give a rough estimate of people who are above older than 75 years
* Age Distribution:

|  |  |
| --- | --- |
| **Age Group** | **Rough Percentage** |
| (75 – 80) | 3.7 |
| (80 – 85) | 2.9 |
| (85 – 90) | 1.8 |
| (90 – 95) | 0.0 |
| (95 – 100) | 0.5 |
| **Total** | **8.9** |

* This we can conclude that roughly 8.9% of people of the total population surveyed are above the age of 75 years.