Homework #2

 In order to construct a stem and leaf diagram first you have to dissect the numbers into stems and leaves. The first digit will be the stem and the last two digits will be the leaves. The leaves also can be interrupted into frequencies, since every stem is dividing into class. For example; class 6 will be a category that includes 600-699.

|  |  |
| --- | --- |
| **Stems** | **Leaves** |
| 3 | 67 |
| 4 | 25 |
| 5 | 00,75,95 |
| 6 | 30,50,60,82 |
| 7 | 10,20,49,60,70 |
| 8 | 20,43,49 |
| 9 | 45,50 |
| 10 | 16,60,90 |
| 11 | 20 |
| 12 | 95 |
| 13 | * NA-
 |
| 14 | 80 |

The frequency chart can be used to construct a frequency histogram

**Example 3.1**:

1. Find the value of ∑x for the recorded observation are 2, 1, 3,2,3

Solution: The ∑ explains the sum of values in the data set, therefore,

∑x = 2+1+3+2+3 = 11

1. Find the value of ∑x2

∑x2 = (2)2+ (1)2+(3)2+(2)2+(3)2= 4+1+9+4+9= 27

**Example 3.2**:

1. Find ∑(x-3)

∑(x-3) = (2-3)+ (1-3)+(3-3)+(2-3)+(3-3)= -1-2+0-1+0= -4

1. Find ∑(x-3)2

∑(x-3)2= (2-3)2+ (1-3)2+(3-3)2+(2-3)2+(3-3)2= (-1)2+(-2)2+(0)2+(-1)2+(0)2= 1+4+0+1+0= 6

1. Find ∑x2-3

∑x2-3= (2)2+ (1)2+(3)2+(2)2+(3)2-3 = 4+1+9+4+9-3= 24

**Example 3.8**: Find the variance for the sample measurements 3, 7, 2, 1, 8

 The variance S2 of a set of n samples measurement is equal to the sum of square of deviations of the measurement … mean divided by the value of (n-1)-

S2 = ∑(x-)2/(n-1)

Solution:

x= ∑x/n = (3+7+2+1+8)/5 = 21/5 = 4.2

**Example 3.14**: Calculating the z- score provided value of sale price (x), mean( ) and standard deviation (s) are given

Given: x= $200,000;  = $106,405 and s= $85,414

Solution: The formula to calculate z-score is-

 z = (x- )/ s

 z = (200,000-106,405)/ 85,414 = **1.10**

**Since the z-score value is positive it means that the $200,000 sale price lies a distance of 1.10 standard deviation to the right of the mean $106,405**

Homework #3

**Example 4.8** Applying Probability: the probabilities of the outcomes that occur in that particular event. For example, probability is determined by formula: P (7) = 1/36 + 1/36 + 1/36 + 1/36 + 1/36 + 1/36 = 6/36

**Example 4.13**

History has shown that a manufacturing operation produces, on the average, only one defective unit in 10 units. Defective units are removed from the production line, repaired, and returned to the warehouse. Suppose that during a given period of time you observe five defective units emerging in sequence from the production line.

1. Assume Randomness L What is the probability of observing five consecutive defective units?
2. If the event “a)” occurred, what could be concluded about the process

SOLUTION:

1. The events observed are considered unconditional probabilities and the defectives occur at random. Therefore, the units are independent of each other and cannot be treated any other way.

P (all 5 are defective) = (1/10) (1/10) (1/10) (1/10) (1/10) = 1/100,000 = 0.000001

STEPS:

* Define the events:
	+ D1: first unit is defective/D 2: second unit is defective /D 3: third unit defective/D 4: fourth unit defective / D 5: fifth unit defective
* Determine probability of defective units: 1 defect per 10 units = 1/10
	+ P(D1) = P(D 2) = P(D 3) = P(D 4) = P(D 5) = 1/10
1. By observing historical data it is easy to understand observing in production five consecutive defective units is strongly unlikely.

STEPS:

* Examine historical data to find out if anything similar has happened in the past.

**Example 4.15**

SOLUTION:

Suppose an experiment consists of tossing a pair of coins and observing the upper faces. Define the following events

A: Observe at least one head

B: Observe at least one tail

 You must then use the additive law of probability

P (A or B or both) =P (A U B) = P (A) +P (B) –P (A and B)

1. P (A) = P (head) = 3/4

2. P (B) =P (tail) = 3/4

P (A) + P (B) - P (A and B)

The chance of tossing either a head or tail on each coin is ½

(3/4)+(3/4)-(1/2)= 1

**Computer Implementation:**

The JavaScript supported the above calculations

**Example 4.16**

SOLUTIONs:

1. P (A and B) = .14
2. PA = .24 and PB = .39
	1. PA + PAbar = 1
	2. 1 - .24 = .76
3. The probability that A or B occur can be using the previous solutions:
	1. .24 + .39 - .14 = 49

**Computer Implementation:**

The use of the JavaScript mirrored the outcomes of the calculations

**Example 5.2**

SOLUTION:

Construct a relative frequency distribution for the 100,000 values of X

 1. To determine the values of relative frequency of sample, the value can be obtained by dividing the frequency of each x value by the total number of observations

|  |  |  |
| --- | --- | --- |
| X | Frequency  | Relative frequency |
| 0 | 32891 | .32891 |
| 1 | 40929 | .40929 |
| 2 | 20473 | .20473 |
| 3 | 5104 | .05104 |
| 4 | 599 | .00599 |
| 5 | 4 | .00004 |

2. Show that the properties of a probability distribution are satisfied

 To determine sum up all values of X and determine whether distribution is between 0 and 1.

**Computer Implementation:** The above charts and graphs were created in Excel to visually reflect the findings for part one and two.

**Example 5.3**

SOLUTION:

Find the mean for example 5.2

Using the formula- find the expected value of X- You find the expected value of X by referring to table 5.1. X is the observations of X in the sample size of N=5

 μ = Sum of each x times P( X = x) = Σ x p(x)

**a.**

1. Substitute where p(x) is given in table – these values can be obtained by dividing the frequency by the total number of observations ie x=0 frequency 32,891/100,000- total number of observations- relative frequency
2. Sum up all the values of x corresponding to the sample
	1. μ = 1.0.
		1. Interpretation of μ indicates that over a period of time that number of consumers who favor Brand A will be equal to 1.o. The mean of a random variable provides the long run average of the variable or expected average outcome over many observations

c. Find the standard deviation

 1. The variance is equal to the [sum of x^2p(x)]- μ^2 .

Substitute the values of the unknown and compute to get the standard deviation

Compute the mean using the formula from part A , square the value of X corresponding to the sample and solve

μ =sumxp(x), μ = 0(p)0+ 1(p)1+2(p)2

Next square the values of X ie 0(p)0+ 1(p)+ 4(p)2+ 9(p)3 to (1)^2

Now plug in values of p(x) ie. 0(.32768)+ 1(.40960)+4(.20480) to - 1

This gives the variance,

The square root of the value gives the standard deviation.

**Example 5.4**

SOLUTION:

Empirical rule application μ ±2(standard deviation)

Calculate both the mean and standard deviation,

substitute the unknown values for the known values.

Calculate and complete the problem.

The interval for the problem show a range of -.788 to 2.788

includes the values=0, x=1,x=2.

Add the values of the relative frequency or probability to determine the interval

P(0)+p(1)+p(2)= .32768+.40960+.20480= .94208.

The value illustrates the empirical rule that 95% will lie within the 2(standard deviation) of the mean

**Example 5.8**

SOLUTION:

Summing binomial probabilities-

that three or more persons in the sample prefer Brand A.

Find the probability of 3 or more observations - when applying probability rule #1

probability of 3 mutually exclusive events is equal to the sum of their probabilities.

**Example 5.9**

SOLUTION:

Determine if the statement is true:

 p0+p1+p2+p3+p4+p5+p6 = .0010

Because the probability is very small the event is likely a rare event, not likely to happen if all other factors remain constant.

Homework #4

**Example 4.8** Find the probability that the sum of the numbers is equal to 7

 1. Record how many ways the dice could fall. Each di has 6 sides for a total of 36. The equal frequency of the dice falling is equal to 1/36.

2. Then compute the problem by identifying how many possible ways a 7 could be tossed with a pair of dice. P (6,1)+ P(2, 5)+P (5,2)+P(3,4)+P(4,3)+P(1,6) each is equal to 1/36 equaling 6/36 or 1/6

**Example 4.13** Suppose that during a given period of time you observe five defective units emerging in sequence form the production line

 A. What is the probability of observing a sequence of five consecutive defective units’ sample of 10?

Set up the problem and determine the unconditional probability which is 1/10

**Example 4.15** Suppose an experiment consists of tossing a pair of coins and observing the upper faces. Define the following events

A: Observe at least one head

B: Observe at least one tail

Use the additive law of probability

P( A or B or both)=P(A U B)= P(A)+P(B) –P(A and B)

1. P(A)= P(head)= 3/4
2. P(B)=P(tail)= 3/4

P(A) + P(B)- P(A and B)

The chance of tossing either a head or tail on each coin is ½

(3/4)+(3/4)-(1/2)= 1

**Example 5.2** Construct a relative frequency distribution for the 100,000 values of X

 1. To determine the values of relative frequency divide the frequency of each x value by the total number of observations

|  |  |  |
| --- | --- | --- |
| X | Frequency  | Relative frequency |
| 0 | 32891 | .32891 |
| 1 | 40929 | .40929 |
| 2 | 20473 | .20473 |
| 3 | 5104 | .05104 |
| 4 | 599 | .00599 |
| 5 | 4 | .00004 |
|  |  |  |
|  |  |  |

2. Show that the properties of a probability distribution are satisfied

 To determine sum up all values of X and determine whether distribution is between 0 and 1.

**Example 5.3** Computing the Mean and standard deviation

 Find the mean for example 5.2

Using the formula- find the expected value of X- You find the expected value of X by referring to table 5.1. X is the observations of X in the sample size of N=5

 μ = Sum of each x times P( X = x) = Σ x p(x)

**a.**

1. Substitute where p(x) is given in table – these values can be obtained by dividing the frequency by the total number of observations ie x=0 frequency 32,891/100,000- total number of observations- relative frequency
2. Sum up all the values of x corresponding to the sample

b. Interpret the value of μ - the value of the mean can by obtained by adding the values of x and dividing the sum by the sample size. For this example the μ = 1.0. Interpretation of μ indicates that over a period of time that number of consumers who favor Brand A will be equal to 1.o. The mean of a random variable provides the long run average of the variable or expected average outcome over many observations

c. Find the standard deviation of X

 1. The variance is equal to the [sum of x^2p(x)]- μ^2 . First you substitute the values of the unknown and compute to get the standard deviation. Compute the mean using the formula from part A , square the value of X corresponding to the sample and solve. Ie μ **=sumxp(x),** μ **= 0(p)0+ 1(p)1+2(p)2…………….**

 **Next square the values of X ie 0(p)0+ 1(p)+ 4(p)2+ 9(p)3…………- (1)^2**

**Now plug in values of p(x) ie. 0(.32768)+ 1(.40960)+4(.20480)…….- 1**

This gives the variance, finally complete the problem by square rooting the value gives the standard deviation.

**Example 5.4** Locate the interval using the empirical rule application

μ ±2(standard deviation)

 For the sample compute both the mean and standard deviation, then substitute the unknown values for the known values. Compute and complete the problem. The interval for the problem show a range of -.788 to 2.788 and includes the values=0, x=1,x=2. Now sum all the values of the relative frequency or probability to determine the interval the population value falls within. Ie

P(0)+p(1)+p(2)= .32768+.40960+.20480= .94208. The value shows agrees with the empirical rule that 95% will lie within the 2(standard deviation) of the mean

**Example 5.8**. Summing binomial probabilities- Refer to 5.7 find the probability that three or more persons in the sample prefer Brand A.

This problem is simple asking to find the probability of 3 or more observations, when applying probability rule #1 the probability of 3 mutually exclusive events is equal to the sum of their respective probabilities.

Homework #5

**Example 6.1**T compute a z score, we can use the formula Z = ( x - µ ) / σ. Given the values of x, µ and σ, plug into equation and solve. Z score measures the number of standard deviations that a point (x) away from the mean.

**Example 6.2**To calculate the area under the standard normal curve between a point x and the given value of Z, in this case 1.26, use the standard normal table and move down the left side of the table to 1.2 and across the top of table to .06 and the intersection of these tow values is the area. Result is 0.3962.

**Example 6.3**Since Z (-1.26 for this example) is negative, we know that it is to the left of the mean. Remember that the mean in the standard normal curve is zero. In the last example, we found the area for 1.26. Since the standard normal curve is symmetric, the area is the same. Result is 0.3962

**Example 6.4**For this example, the z score is 2. We follow the same procedure as before looking down the left column to 2 and across the top to 0. The intersection is 0.4772. Since we want from –z to z, this is 2 times z or 2(0.4772) =.9544.

**Example 6.5**Standard normal table is symmetrical. The total probability is 1.0, so based on symmetry area to the right of 0 and area to the left of zero are the same or .5. To find the probability that a normally distributed random variable will lie more than z (in this case 2) standard deviations we follow the procedure to find the area from 0 to z as we have done before using the table. Then subtract this value from .5, because this result is the area under the curve corresponding to that from 0 to the Z value. In this case, we use our area of 0.4772 and subtract from 0.5 and get 0.0228.

**Example 6.6**It helps to draw a diagram to visualize this. Since both z score values are positive, they are on the same side of the mean (the right) and the result would be the difference between the two values. Based on what we have been doing, this would be the difference between the two values we would get by looking up our z scores in the table. The two z values in this example are 1.6 and 1.2, which is the difference of 0.4452 and 0.3849, which is 0.0603.

**Example 6.7**Given a value, Zα = Z.10, we are asked to find the area. In this case zα = z.10 which is positive, so we seek the area of 0 to Z.10 and then we will look for that area in the chart to get thez. Once again we know the half of the standard normal curve has an area of .5. We take the difference of 0.5 and 0.1 to get 0.4. We then look into the body of the normal distribution table to find a z value. The corresponding value is not shown, but we can see that it is between .3997 (Z = 1.28) and .4015 (Z = 1.29). BY examining the difference from 0.4 to each of our values, we see that 0.3997 (Z = 1.28 is closer) so we choose that value. We could probably interpolate to find a better figure.

**Example 6.8**The problem gives us the following information
Average Salt Consumed per day by an American - 15 g (µ)
Actual physiological minimum daily requirement - 22 g
Amount of salt intake is normally distributed with a standard deviation of 5 g (σ)
We are given two values for x, let x1 = 14 and x2 = 22
We want to know what percentage of Americans consume between x1 and x2

Use the formula z = x - µ / σ for each value of x, because we have x, σ and µ. We get z1 = -.2 and z2 = 1.4. Once again using the standard normal table and finding that corresponding area we have A1=.0793 and A2=.4192. One Z value is negative and once is positive. We add the A values we get and the result is.0793 + .4192 = .4985. Based on this we state that 49.85% of Americans consume between 14 and 22 grams of salt per day.

**Example 6.9** with the information given in the previous example, we are asked to determine the probability that a randomly selected American consumes less than 1 g of salt per day. Use the formula z = ( x - µ ) / σ for this value of x, we get z=-2.8. By using the Standard Normal Probability Table, we find that .0026 is the result. Based on this we state that .26% of Americans consume less than 1 gram of salt per day.

**Example 6-15**We are told that production lines have lengths that can be modeled by a uniform probability distribution over the interval of .95 to 1.05 inches.

a. We are asked to calculate the mean and standard deviation of x. Since we are told this is a uniform probability distribution for the interval in questions, we can use the following equations: µ = (a + b) /2, and σ = (b – a )/12⅟₂. µ = 1 inch and σ=.029 inch.

b. We are also asked to graph the probability distribution and locate µ, µ ± σ, µ ± 2σ. The height of rectangle of this probability distribution plot is 1 / (b-a) = / (1.05-.95) = 10. All heights are the same. Once you plot you can see that all production line lengths of .95 to 1.05 inches are within two standard deviations of the mean.

Computer Assignment

I used the Standard (Normal) density program to test what result I would get if I plugged in a z value. We were to “*Enter your computed z-statistic, and then click the* ***Compute*** *button.”* From the Homework, example we were to calculate the Area above a z value. From Example 6.5, we calculated this for z = 2. From the homework, the answer was 0.0228. I plugged 2.0 into the program and 2P was .046. At first I was surprised because the number was not what I expected. Dividing 2P by 2 and I got .023 which is the rounded version of the homework correct response. I wasn’t sure what this meant or whether this was a coincidence. Next I used z value of 1.26. We know from the homework that the probability is .3962, but the result from the program was .208. Dividing this result by 2 and p=.104. This time I subtracted .3962 from .5 and got .1038 or .104 rounded. I was still not sure whether there is a trend, so I decided to try some more z values. For z=0, program returned 2p of 1.0 or a p=.5. From the standard normal curve, we know at z=0, which is at the mean, that one half of the area is to the right (and left). For z=1.4, area is .4192. Program returned a 2P value of .162 or p=.081. From the homework, result should be .4192. Subtracting this from .5 and we get .0808 or rounded .081

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Z Value | Result from Table | .5-Table result | Program Result, 2P | P |
| -2 | 0.4772 | 0.0228 | 0.046 | 0.023 |
| -1.5 | 0.4332 | 0.0668 | 0.134 | 0.067 |
| -1.26 | 0.3962 | 0.1038 | 0.208 | 0.104 |
| -1 | 0.3413 | 0.1587 | 0.317 | 0.1585 |
| 0 | 0 | 0.5 | 1 | 0.5 |
| 1 | 0.3413 | 0.1587 | 0.317 | 0.1585 |
| 1.26 | 0.3962 | 0.1038 | 0.208 | 0.104 |
| 1.5 | 0.4332 | 0.0668 | 0.134 | 0.067 |
| 2 | 0.4772 | 0.0228 | 0.046 | 0.023 |

It looks liked the program returns a value of twice the area to the right of the z value for z≥0.For z≤0, program returns a value twice the area to the left of the z value. Below is a chart of data I used

Homework #6

**Example 10.7**

SOLUTION STEPS:

1. First develop and null hypothesis (H0) and a alternate hypothesis (HA)
	1. H0 : Defendant is innocent
	2. HA : Defendant is guilty
2. Determine all possible outcomes:
	1. Four outcomes;
		1. reject or accept (Type I error) H0
		2. reject or accept (Type II error) HA
3. Decide which error needs to be avoided:
	1. Type I error accepting the H0 is the error the needs to be avoided the most
		1. Decrease alpha help eliminate the possibility of Type I error

**Example 10.8**

1. First develop and null hypothesis (H0) and a alternate hypothesis (HA)
	* 1. H0 : Defendant is innocent
		2. HA : Defendant is guilty
2. Obtain random sample from population
	1. The sample statistic, aka, test statistic is the basis for the decision
3. Determine the a test statistic
	1. The best guess about the population mean is usually 
4. Identify the rejection range of values that might lead us to reject the null

**Computer Implementation: n/a**

**Example 10.11**

The null hypothesis (H0) and a alternate hypothesis (HA) are stated below as:

* + - H0 : µ = 72
		- HA : µ > 72
1. If  = 110 it is unlikely the population µ would actually be 72.
2. Since HA is saying the µ is less than 72,  = 59 does not support HA
3. More information is needed to determine why  could be 73

**Example 10.12**

The null hypothesis (H0) and a alternate hypothesis (HA) are stated below as:

* + - H0 : µ = 72
		- HA : µ > 72
		- α = .05
* Determine the z-score of  to see if 73 is sufficiently greater than 72 to reject H0
	+ The sample mean and standard deviation are valid if the null is true
		- z =  - µ(72) / s / √n
			* z = 1.645 = critical value
* All values at or above 1.645 (standard deviations / critical value) are in the null rejection region

**Example 10.16**

The null hypothesis (H0), alternate hypothesis (HA), and significance value are stated as:

* + - H0 : µ = .5
		- HA : µ ≠ .5
		- α = .05
* Determine the rejection range by using the z-score of 
	+ The sample mean and standard deviation are valid if the null is true
		- z =  - µ(.5) / s / √n
			* z = -3.77 /2 = -1.96 (left tail) and 1.96 (right tail)
* The results a Type I error could occur with a probability of α = .05

**Example 11.10**

The null hypothesis (H0), alternate hypothesis (HA), and significance value are stated as:

* + - H0 : µ1 - µ2 = 0
		- HA : µ1 - µ2 ≠ 0
		- α = .01

Let µ1 = mean of stayers and Let µ2 = mean leavers

* Determine the rejection range by the z value
	+ Two tailed sample - Reject null if z is greater than z value divided by 2 (tails)
		- z value of .005 = 2.58; z less than 2.57 and z greater than -2.58
		- z = 1 – 2 / √s12/n1 + s22/n2 = 5.68
* Verify if the z value is outside the rejection range
	+ Z = 5.68 is in the rejection region – The mean of the stayers is significantly different than the leavers
		- The probability of Type I error is α = .01

**Example 11.12**

The null hypothesis (H0), alternate hypothesis (HA), and significance value are stated as:

* + - H0 : µ1 - µ2 = 0
		- HA : µ1 - µ2 > 0
		- α = .05
* Let µ1 = mean of type A and Let µ2 = mean of type B
* t distribution – sample is small; n = 10
	+ (n - 1 ) = (10-1) = 9 degrees of freedom
* t > t.05 = 1.833
* Determine the test statistic
	+ Calculate the difference (d), in duration of use, for the shoes and for each jogger (table 11.4)
	+ Calculate the mean difference and the standard deviation of the sample
* Use the results from computer programs like Excel (t-distribution) or ASP we find that:
	+ Mean difference = 2.6
	+ Standard deviation = √13.3778 = 3.66
		- t = d – D0 / sd / √n = 2.25
* The test statistic exceeds the t.05 = 1.833 critical value, there is sufficient evident to indicate the mean difference of A exceeds the mean for B

**Computer Implementation:**

The use of Excel’s t-test provided the same information as ASP. Less information was needed. I type the chart into Excel including the headings and all information was computed. Nevertheless, both programs came to the same conclusion: A mean exceeds the mean for B.

**Example 11.15**

The null hypothesis (H0), alternate hypothesis (HA), and significance value are stated as:

* + - H0 : $σ$2 = .01
		- HA : $σ$2 < .01
		- Normally distributed
		- Alpha = .05

1. Determine test statistic

* + Chi Square = (n-1)s2 / $σ$20

2. Find rejection range

* + Reject null if chi-square < 3.32511 (to the right) & (to the left) chi-square.95 =1.66

3. The test statistic chi-square = 1.66 and is less than 3.32511, therefore we reject the null and accept the alternate

**Example 11.17**

The null hypothesis (H0), alternate hypothesis (HA), and significance value are stated as:

* + - H0 : $σ$21 / $σ$22 = 1
		- HA : $σ$21 / $σ$22 ≠ 1
		- Alpha = .10

1. Determine test statistic

* + F = s21 / s22 = 2.27

2. Determine rejection region for a two tailed F test

* + Reject null if F > 3.79

3. Output concludes we should fail to reject the null with a certainty of alpha = .10

* + Data did not support a difference between the variances

To further support the claim the value of Type II error needs to be figured