**Math E-3 Assignment 7**

**If you use the Microsoft Word version to do the homework, make sure to view the pdf version to ensure that you have all the correct figures.**

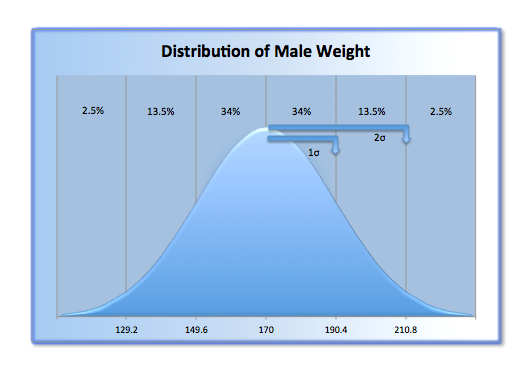
**Z scores**

**Please show your normal distribution curve. Refer to z score tables at the end of the homework assignment to answer problems 2 and 3.**

**Problems 1-3**

In a given population of males it is found that the mean weight is 170 lbs. Furthermore, the calculated standard deviation is 20.4 pounds. Assume the distribution of weights is normal.

1. Draw the curve of the normal distribution of male weights.



1. Use the Z score tables at the end of the assignment to calculate the percentage of males who have weights between 140lbs and 170lbs.



*According to Curve A from , the percentage of males who have weights between 140lbs and 170lbs is 43.32%.*

1. Use the Z score tables at the end of the assignment to calculate the percentage of males who have weights between 155lbs and 195lbs.

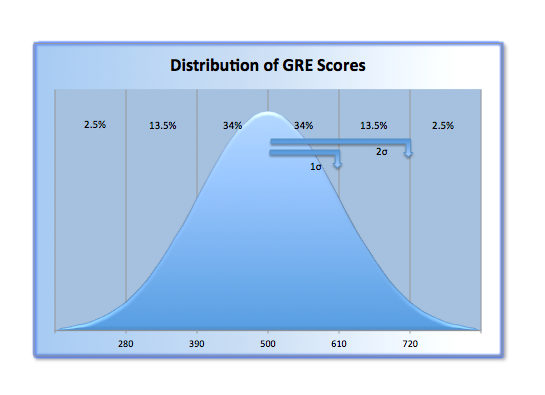


*Since the point 155lbs is to the left of the mean (170lbs), and the point 195lbs is to the right of the mean, the sum of both of these distances will yield the correct percentage of males who have weights between 155lbs and 195lbs, i.e. 64.29%.*

**Problems 4-6**

Suppose the scores on the Graduate Record Exam (GRE) are normally distributed with a mean of 500 and a standard deviation of 110.1

4) Draw curve of the distribution of GRE scores.



1. If the graduate school you are interested in attending requires a GRE score of 630 for admission, how many standard deviations above the mean do you need to score?



*According to the indicated value of Z, it would appear that the “Z number” gives us the distance and the “number of standard deviations” a given point is away from the mean value of a normally distributed density function graph. For further explanation consider the following:*



*With the lather being said, I would agree that it is possible for the “Z score table” to produce the “number of standard deviation’s” a point is from the mean average of a normally distributed graph.*

*Thus in order to gain admission to the chosen graduate school, one must obtain a GRE score of at least 630, which would be 1.2 standard deviations above the mean.*

1. If the admission's officer chooses an applicant at random, what is the probability of finding a student who scores a grade over 700 on the GRE? Use the Z score table at the end of the assignment to help you answer this.



*As indicated in the figure Curve B, the probability of an admission officer randomly choosing a student who scores a grade over 700 on GRE is 3.59%.*

*Note\* Max is left to be defined as the maximum possible score.*

1 Problem adapted from *Using and Understanding Mathematics*. Jeffrey Bennett, William Briggs. Addison, Wesley, Longman, 1998.

**Hypothesis Testing**

**For problems 7 through 12, assume all samples were randomly chosen even if not stated in the problem.**

**Make sure you follow the steps as outlined in class and the reading; be sure to include a diagram for each question, and remember to clearly state both the Null Hypothesis and the conclusion in each case.**

**Follow this format for full credit for problems 7, 8, 9, 10, and 12:**

**Step 1) State your Null Hypothesis – use words not just a percentage.**

**Step 2) Calculate the Standard Deviation.**

**Step 3) Draw your diagram with the mean and 1 and 2 standard deviations identified.**

**Step 4) Calculate (if necessary), state, and compare the observed percentage.**

**Step 5) Construct the proper sentence either rejecting or not rejecting the Null Hypothesis. Use the proper statistical language.**

**Step 6) Give an informal conclusion.**

**Problem 11 does not require a new hypothesis test.**

**Problem 7**

The Beautiful Body Cosmetics Company claims that its new wart cream dissolves 54% of all warts with one application. A scientist from a competing company is given the job of disproving this claim. She purchases a few jars of the product and does her own tests. If this scientist tries the cream on several randomly selected people (and randomly selected warts!), and finds that after applying the cream to 300 warts, 158 of the unattractive warts disappeared with one application. Perform a hypothesis test and determine what the scientist would conclude about Beautiful Body’s product.

**1. Null Hypothesis:**

Beautiful Body Cosmetics Company’s new wart cream dissolves exactly 54% of all warts with a single application. Let *P* represent this percentage, such that  implies that 46% of the time the cream is ineffective:



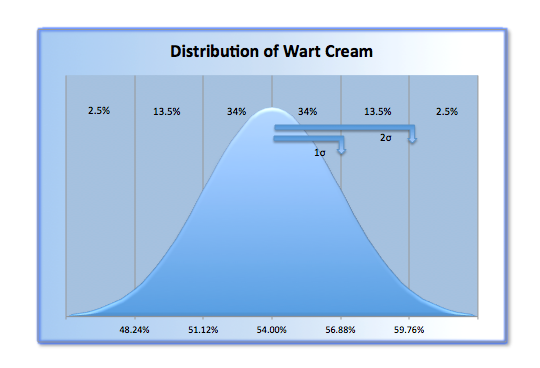
If this were true, then in a random sample of 300 warts, you would expect to get a 54% success rate. Our expected mean is 54% or, and our sample size is.

Now let us calculate the standard deviation of the data set:

****

Thus, the standard deviation is approximately 2.88%.

**3. Diagram**

****

After observing all 300 applications of the wart cream, we discovered that 158 of the trials were successful:

****

This is roughly equivalent to 52.67%. Now lets compare our observed success rate to the 54% success rate claimed by the company. As indicated by the graph, 52.67% falls within less than 1 standard deviation away from the mean, which puts our observations within the likely region:





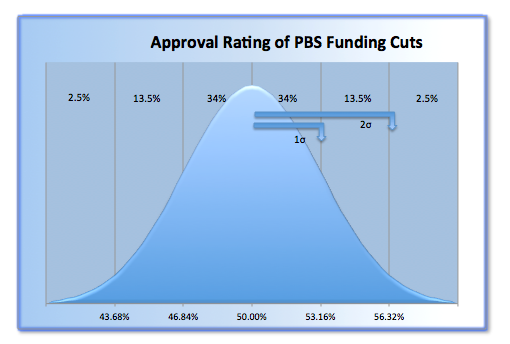
Since our observed percentage falls within the likely region, we cannot reject the company’s claim of a 54% success rate, at a 5% level of significance.

We conclude that it is possible (but not certain) that the company is accurate in their assessment.**Problem 8 (This was an issue a few years ago, when Trent Lott was in the Senate)**

Senator Lott claims that 50% of Americans want federal funding for PBS **cut.** You, a devout fan of “Barney” and “Lamb Chop” (sadly, Shari Lewis passed away in 1998), are deeply suspicious of this claim. So, you decide to take your own *unbiased* poll.You sample 250 people and find that 46% want to see the cuts in funding. Perform a hypothesis test, and give your conclusions.

 *50% of Americans want federal funding for PBS cut.*

****



When comparing our observation that 46% of the 250 people want to see cuts in funding, to that of Senator Lott’s claim of 50%, our observation falls within in the two standard deviation range, which is considered a likely region.

Since our observed percentage falls within the likely region, we cannot reject Senator Lott’s claim that 50% of Americans want federal funding for PBS cut, at a 5% level of significance.

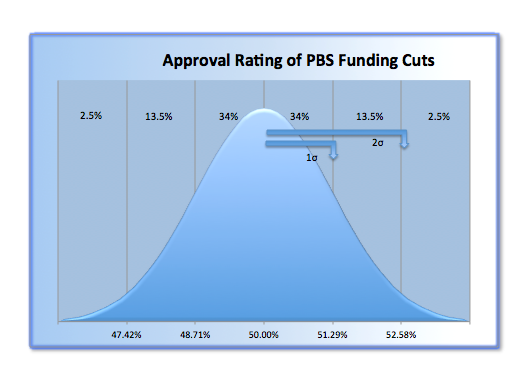
We conclude that it is possible (but not certain) that the senator is accurate in his assessment.

**Problem 9**

You decide totake another sample of people to find out for sure about the public’s opinion on cuts to PBS, this time of 1500 people. From this sample, 700 folks favor the cuts. Perform another hypothesis test. What is your conclusion this time? If your conclusion is different than in problem 8 above, explain *why* there is a difference.

 *50% of Americans want federal funding for PBS cut.*

****



Let us calculate our observed percentage of the 1500 people: 

When comparing our observation of 46.78% to Senator Lott’s claim of 50%, our observation does not fall within in the two standard deviation range. Thus our observation is considered unlikely.

Since our observed percentage falls in the unlikely region, we can reject Senator Lott’s claim that 50% of Americans want federal funding for PBS cut, at a 5% level of significance.

We conclude by rejecting Senator Lott’s claim that 50% of Americans want federal funding for PBS cut. The observed differences in the conclusions between our tests were due to a larger sample size. By increasing the sample size, we were able to reject Senator Lott’s claim.

**Problem 10**

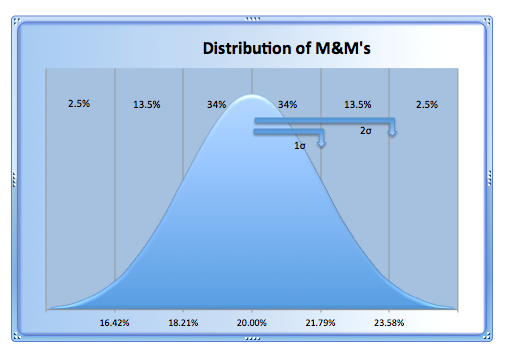
The Mars Company, maker of M&M’s, recently claimed that M&M’s were so much fun because they were a perfect rainbow, that is, each bag they made contained equal numbers of each of the five colors. An enterprising student decided to test this and bought several small bags selected at random, emptied them out and found that the number of M&M’s found for each color were:

|  |  |
| --- | --- |
| Brown: | 128 |
| Blue: | 87 |
| Green: | 95 |
| Orange: | 100 100 100 100 |
| Yellow: | 90 |

Perform a hypothesis test and state your conclusions about Mars’ claim.

 Each bag of M&M’s contains an equal number of the five colors.





By dividing the number of M&M’s in each color by the sample size we have:

Brown = 25.6%

Blue = 17.4%

Green = 19%

Orange = 20%

Yellow = 18%

Brown is **not** within two standard deviations of the mean, which is in the **unlikely** region.

Blue is within two standard deviations of the mean, which is in the likely region.

Green is within one standard deviation of the mean, which is in the likely region.

Orange is within one standard deviation of the mean, which is in the likely region.

Yellow is within two standard deviations of the mean, which is in the likely region.

Since our observed percentage of the Brown M&M’s falls within the unlikely region, we can reject the Mars Company’s claim that M&M’s are evenly distributed according to their color at a 5% level of significance.

We conclude that Mars Company was inaccurate in their assessment.

**Problem 11**

Refer to problem 10 above**.** If the colors were in equal numbers, what would be the probability that a random sample of 500 M&M’s will have the above proportion of brown ones? (Hint: Go back to the Normal Distribution Curve and look at the probabilities under the curve.) No need to perform another hypothesis test here.

This question is quite vague when it comes to defining the word “proportion”, in context to the interpretation of the meaning of the word proportion I propose two unique solutions.

(I think this the right interpretation) Solution 1: If we interpret the word “proportion”, and “colors were in equal in numbers” in the way such that the claim is true that M&M’s do indeed share an equal distribution of colors and we just took an unlikely random sample that happened to contain 25.6% brown M&M’s then:



Thus we could expect this unlikely event to exist with an expectation 0.1% of the time when conducting a random sample.

(However if not) Solution 2:

Let proportion be defined as the number of brown M&M’s in problem 10, such that there exist 128 brown M&M’s, or 25.6% of the 500 M&M’s are brown. Also it requires that all colors are uniformly distributed in the set M&M’s, (our sample space) then all other elements of M&M’s that are not Brown should give an expectation equal to . Since it is true that all M&M’s are equally distributed, we can expect to see an M&M that is not brown 80% of the time. But since in this solution we explicitly defined “proportion” in the following context:



Thus, if this were the case then the probability that 500 M&M’s would have the above proportion of brown ones (i.e. 128) would be 0% since it was asserted that there exists an equal distribution of M&M’s in our sample space.

**Problem 12**

A psychologist wants to test whether very young children are especially attracted to bright ‘crayon box’ colors. Thus, she places her young subjects in a playroom with three balls to choose from:

one is striped with black and white

one is made up of two assorted pastels shades

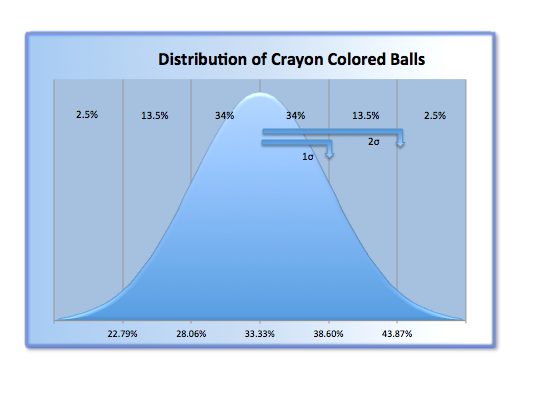
the last is striped with two bright primary colors.

The balls are identical in all other respects and their position is shuffled between trials so that it will not influence the results. From her observations she found that of the 80 toddlers she tested, 42 went for the brightly colored ball.

Formulate an appropriate null hypothesis as if you were the researcher, and perform a hypothesis test. (Remember that in hypothesis testing, you must have a proportion (percent) to work with. Think carefully about this percentage.) At the end, state your conclusion explicitly, i.e. do children tend to show a preference for bright colors?

It makes no difference to young children if the balls are bright ‘crayon box’ colors.





First lets calculate our observed percentage in which the toddlers chose the bright colored ball: .

When comparing our observations to that of our hypothesis, we determine that 52.5% is farther than two “standard deviation’s” away from the mean (33.33...%), this puts our observed percentage within the unlikely region.

Since our observed percentage of the toddlers choosing the bright ball falls within the unlikely region, we can reject our hypothesis that said it doesn’t make a difference to toddlers. Our findings are insured within a 5% level of significance.

We conclude that it is possible (but not certain) that toddlers are attracted to bright colored balls, thus we can reject our null hypothesis even without the help of the pastel shades!

**See below for info about a program I made analogously to this PSET to analyze Z-Tables by importing the table of values below automatically.**

INSTALLING

After downloading the folder, unzip the .zip file.

Within that directory you should see the following files:

*ztables - (executable)*

*ztables.c - (source code)*

*Z-Tables.txt - (txt file containing Z-Tables)*

*percent stuff.xlsx - (used to make graphs in addition to lecture notes)*

*readme.txt – (copy of this)*

To run the ztables, do the following on a Mac or Linux system:

Click on the magnifying glass (known as spotlight) in the upper right hand corner -> search for “terminal” and select, then click it.

Move the folder you just unzipped onto your desktop.

Click on the black screen called the terminal and type the following commands pressing return at the end of each of line:

*cd ~/Desktop/ztables*

*make ztables*

*./ztables*

You should have successfully executed the program, however you didn’t get to decide what it computed!

Note\* If you run “make” and nothing happens, don’t worry! You can still run the program, just you won’t be able to edit it yourself and personalize the code.

USAGE

ZTABLES

The usage goes like this:

*./ztables <flag> <args>*

To compute percentages using ztables with different intervals:

*./ztables –zt <START> <END> <MEAN> <STDDEV>*

Example: To have the program compute problem #2 and #3 execute the following:

Problem #2:

*./ztables -zt 140 170 170 20.4*

Problem #3

*./ztables -zt 160 195 170 20.4*

STANDARD DEVIATION

To compute the standard deviation using the new method execute the following command:

*./ztables –sd <P><N>*

Example:

*./ztables –sd 42 80*

Is equivalent to: 

**Thanks for reading!**Z -Scores\*

Z = *x bar*=mean

A B

x bar x xbar x

|  |  |  |
| --- | --- | --- |
| **Z** | **Area between the Mean and X (curve A)** | **Area beyond X (curve B)** |
| **.1** | .0398 | .4602 |
| **.2** | .0793 | .4207 |
| **.3** | .1179 | .3821 |
| **.4** | .1554 | .3446 |
| **.5** | .1915 | .3085 |
| **.6** | .2257 | .2743 |
| **.7** | .2580 | .2420 |
| **.8** | .2881 | .2119 |
| **.9** | .3159 | .1841 |
| **1** | .3413 | .1587 |
| **1.1** | .3643 | .1357 |
| **1.2** | .3849 | .1151 |
| **1.3** | .4032 | .0968 |
| **1.4** | .4192 | .0808 |
| **1.5** | .4332 | .0668 |
| **1.6** | .4452 | .0548 |
| **1.7** | .4554 | .0446 |
| **1.8** | .4641 | .0359 |
| **1.9** | .4713 | .0287 |
| **2** | .4772 | .0228 |
| **2.1** | .4821 | .0179 |
| **2.2** | .4861 | .0139 |
| **2.3** | .4893 | .0107 |
| **2.4** | .4918 | .0082 |
| **2.5** | .4838 | .0062 |
| **2.6** | .4953 | .0047 |
| **2.7** | .4965 | .0035 |
| **2.8** | .4974 | .0026 |
| **2.9** | .4981 | .0019 |
| **3** | .4987 | .0013 |
| **3.1** | .4990 | .0010 |
| **3.2** | .4993 | .0007 |
| **3.3** | .4995 | .0005 |
| **3.4** | .4997 | .0003 |
| **3.5** | .4998 | .0002 |
| **3.6** | .4998 | .0002 |

\*Adapted from "*Understanding Social Statistics*, " by Jane Fielding and Nigel Gilbert.