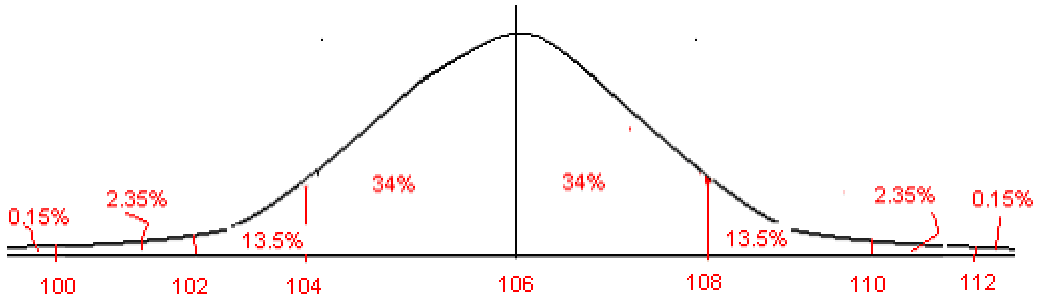
MAT 114 Problem Set I Key

1. A machine is used to put nails into boxes. It does so such that the actual number of nails in a box is normally distributed with a mean of 106 and a standard deviation of 2.

The first step is to draw the Normal curve with 8 regions, and to then label the 7 vertical dividing lines with the mean plus/minus one/two/three standard deviations:



a) What percentage of boxes?

i What percentage of boxes contain more than 104 nails? 84%
ii What percentage of boxes contain more than 110 nails? 2.5%
iii What percentage of boxes contain less than 108 nails? 84%
iv What percentage of boxes contain less than 100 nails? 0.15%
v What percentage of boxes contain between 102 and 112 nails? 97.35%
vi What percentage of boxes contain between 100 and 106 nails? 49.85%

These values come from adding the percentage values in the areas indicated in the question.

b) What is the z-score for a box containing

i. 101 nails? (101-106)/2 = -2.5
ii. 103 nails? (103-106)/2 = -1.5
iii. 107 nails? (107-106)/2 = 0.5

To compute a z-score subtract the given mean (in this problem 106) from the value given in the problem, and then divide by the given standard deviation (in this problem 2). The z-score is the number of standard deviations above the mean corresponding to the given value. (Negative values are interpreted as *below the mean.*)

c) What is the percentile for a box containing

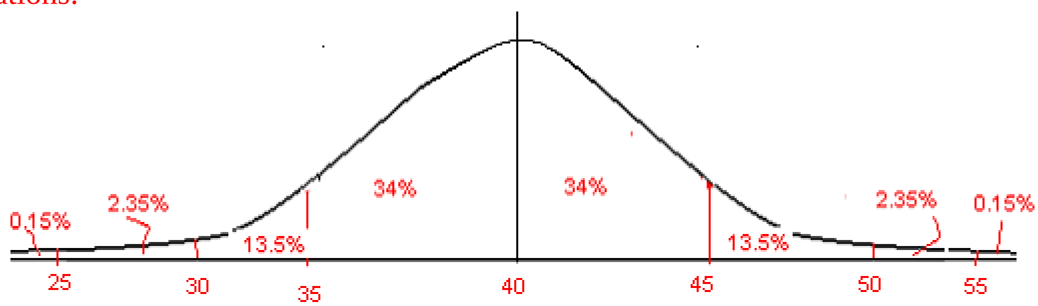
i. 101 nails ? Table value 0.0062. Percentile 0.62% rounded up to **1st**

ii. 103 nails ? Table value 0.0668. Percentile 6.68% rounded up to **7th**
iii. 107 nails? Table value 0.6915. Percentile 69.15% rounded up to **70th**

These answers use the **Percentile Z-Table** from Section 2A to convert the z-scores from part b into percentiles. The Table (in a simplified form) will be provided on the Second Test.

2. The heights of the people if the planet Ixx are normally distributed with a mean of 40 inches and a standard deviation of 5 inches. [They are a vertically diverse people.]

The first step is to draw the Normal curve with 8 regions, and to then label the 7 vertical dividing lines with the mean plus/minus one/two/three standard deviations:



a) What percentage of Ixxians are

i	over 30 inches tall?	97.5%
ii	over 45 inches tall?	16%
iii	under 50 inches tall?	97.5%
iv	under 40 inches tall?	50%
v	between 35 and 55 inches tall?	83.85%
vi	between 25 and 35 inches tall?	15.85%

These values come from adding the percentage values in the areas indicated in the question.

b) What is the z-score for an Ixxian of height

i.	32 inches?	[32-40]/5 = -1.6
ii.	47 inches	[47-40]/5 = 1.4
iii.	51 inches	[51-40]/5 = 2.2

To compute a z-score subtract the given mean (in this problem 40) from the value given in the problem, and then divide by the given standard deviation (in this problem 5). The z-score is the number of standard deviations above the mean corresponding to the given value. (Negative values are interpreted as *below the mean.*)

c) What is the percentile for an Ixxian of height

i. 32 inches? Table Value = 0.0548. Percentile 5.48 rounded up to 6th
ii. 47 inches? Table Value = 0.9192. Percentile 91.82 rounded up to 92nd
iii. 51 inches? Table Value = 0.9861. Percentile 98.61 rounded up to 99th

These answers use the **Percentile Z-Table** from Section 2E to convert the z-scores from part b into percentiles.

d) What is the height of an Ixxian who is at the

i.	25th percentile?	z-score roughly -0.7	Value =(-0.7)(5)+40 = 36.5"
ii.	40th percentile?	z-score roughly -0.25	Value =(-0.25)(5)+40 =38.75"
iii.	60th percentile?	z-score roughly 0.27	Value =(0.27)(5)+40 = 41.35"
iv.	86th percentile?	z-score roughly 1.08	Value =(1.08)(5)+40 = 45.4"

This time we need to first read from the table to get the z- score. The scores do not fall exactly on values on the tables, so it requires a little guesswork. Answers may therefore be different to mine; but they should be close.

To convert from z-score to actual value, reverse the process used in part b. Multiply by the standard deviation (in this case 5) and then add the mean (in this case 40)

3. All that remains of a bridge mix are 6 soft centers (3 mint creams, 2 orange creams and 1 lemon cream) and 4 hard centers (3 toffees and 1 nut).

a. Suppose that one is chosen at random. What is the probability that

c) It is a mint cream? 3/10
d) It is a nut? 1/10
e) It is a mint cream given that it is a soft center? 3/6
f) It is a nut given that it is a soft center? 0

For i and ii, the answer is just the number of things asked for divided by the total number of things. In iii and iv, we are told that the selection is definitely a soft center so we can ignore the hard centers. Of the soft centers, there are 6 in all of which 3 are mint creams and 0 are nuts

b. Are the following events independent?

c) mint and soft center? No
d) nut and soft center? No

The question basically asks whether knowing that the selection is a soft center changes the probability. The answer is yes in both cases. So the events are not independent because the answer is *dependent* on whether you know it is a soft center or otherwise.

c. Suppose that one is chosen at random, and that a second is later chosen at random without the first being replaced. What is the probability of getting

c) an orange cream followed by an orange cream? (2/10)(1/9)=2/90
d) a lemon cream followed by a lemon cream? (1/10)(0/9)=0
e) a hard center followed by a hard center? (4/10)(3/9)=12/90

The two probabilities (for the first and second selections) are worked out individually and then multiplied. In this case, *without replacement,* the two probabilities are different because on the second selection there are fewer things to pick from.

d. Suppose that one is chosen at random, identified and then put back into the bag. Suppose a second is then chosen at random. What is the probability of getting

c) an orange cream followed by an orange cream? (2/10)(2/10)=4/100
d) a lemon cream followed by a lemon cream? (1/10)(1/10)=1/100
e) a hard center followed by a hard center? (4/10)(4/10)=16/100

The two probabilities (for the first and second selections) are worked out individually and then multiplied. In this case, *with replacement,* the two probabilities are identical.

There is no need to multiply out the probabilities on c,d; for example to write (2/10)(1/9) as 2/90. However it is important that you communicate that they are indeed **multiplied.**

4. All that remains of a deck of cards are the 2, 3, 5, 7, 8 and J of diamonds; the 3 and 7 of hearts, the 5, Q and K if spades and the J of clubs. (12 cards in all.)

a. Suppose that one is chosen at random. What is the probability that

c) It is a diamond? 6/12
d) It is a 7? 2/12
e) It is a diamond given that it is a 7? 1/2
f) It is an 8 given that it is red? 1/8

For i and ii, the answer is just the number of things asked for divided by the total number of things. In iii and iv, the conditional information allows us to ignore many of the cards when computing the probabilities.

b. Are the following events independent?

c) diamond and 7? Yes
d) 8 and red? No

The question basically asks whether having more information changes the probability. The answer is No in the first case and Yes in the second. The events in i are independent because the answer does not *depend* on whether you are given the extra information or not. It is the same either way! In ii the extra information does change the probability and so the events are dependent.

c. Suppose that one is chosen at random, and that a second is later chosen at random without the first being replaced. What is the probability of getting

c) a 3 followed by a Q? (2/12)(1/11)=2/132
d) a red followed by a red? (8/12)(7/11)=56/132
e) a club followed by a club? (1/12)(0/11)=0
f)

The two probabilities (for the first and second selections) are worked out individually and then multiplied. In this case, *without replacement,* the two probabilities are different because on the second selection there are fewer things to pick from.

d. Suppose that one is chosen at random, identified and then put back into the bag. Suppose a second is then chosen at random. What is the probability of getting

c) a 3 followed by a Q? (2/12)(1/12) = 2/144
d) a red followed by a red? (8/12)(8/12)=64/144
e) a club followed by a club? (1/12)(1/12)=1/144

The two probabilities (for the first and second selections) are worked out individually and then multiplied. In this case, *with replacement,* the two probabilities are identical.

There is no need to multiply out the probabilities on c,d; for example to write (2/10)(1/9) as 2/90. However it is important that you communicate that they are indeed **multiplied.**

5. One hundred students from three residence halls were surveyed as to their favorite subject. The results obtained are given in the following table:

	Allen	Wilson	Sechrist	Total
Math	1	0	2	3
English	12	11	10	33
Psychology	21	15	28	64
Total	34	26	40	**100**

Suppose one of the surveyed students is selected at random.

a) What is the probability that he/she is from Allen? 34/100
b) What is the probability his/her favorite subject is English? 33/100

In a and b, the answer is just the number of things asked for divided by the total number of things. This requires reading form the chart, but is the same idea as in the previous problems.

c) What is the probability his/her favorite subject is English given that he/she is from Allen? 12/34
d) What is the probability that he/she is from Allen given that his/her favorite subject is English? 12/33

In c and d, the conditional information allows us to ignore many students when computing the probabilities.

e) Are the events "from Allen" and "favorite subject is English" independent?
No

The question basically asks whether having more information changes the probability. The answer is Yes. The events are dependent because the answer is *dependent* on whether you are given the extra information or not. Note that it doesn't matter how you treat 'English' and 'Allen'; in c 'Allen' is the extra information; in d 'English' is the extra information.

6. If a gambling game costs $2 up front to play and a player can win $5 with probability 0.1, $3 with probability 0.1 and gets the $2 with probability 0.2 (and $0 otherwise), what is the expected value of the game to the player?

$$\begin{aligned} \text{Expected Value} &= [(\$5)(0.1)+(\$3)(0.1)+(\$2)(0.2)+(\$0)(0.6)]-\$2 \\ &= [\$1.2]-\$2 \\ &= -\$0.80 \end{aligned}$$

So the player expects to lose 80 cents per play.

7. Two fair 6-sided dice are rolled, and the *player* scores the higher of the two values rolled.

a) Find the probabilities of the scores 1-6. Hint: Complete the following table that shows the scores of each possible roll. Some things have been done for you.

	Red Die						
		1	2	3	4	5	6
	1	1	2	3	4	5	6

Green Die	2	2	2	3	4	5	6
	3	3	3	3	4	5	6
	4	4	4	4	4	5	6
	5	5	5	5	5	5	6
	6	6	6	6	6	6	6

P(1)=1/36; P(2)=3/36; P(3)=5/36;
P(4)=7/36;P(5)=9/36;P(6)=11/36

b) In a game, the player $5 to play and then receives whatever his/her score is in dollars. What are the expected winnings of the player?

Expected Winnings
= [($6)(11/36)+($5)(9/36)+($4)(7/36)+($3)(5/36)+($2)(3/36)+($1)(1/36)]-$5
= [$161/36]-$5
= [$4.47222]-$5
= -$0.52777

So the player expects to lose about 5.28 cents per play.

8. All that is left of a bag of Skittles are 3 yellows, 2 pinks and 4 greens.

a) If two Skittles are removed without replacement, what is the probability of

i. Two yellows = (3/9)(2/8) = 6/72
ii. Two pinks = (2/9)(1/8) = 2/72
iii. Two greens = (4/9)(3/8) = 12/72

The two probabilities (for the first and second selections) are worked out individually and then multiplied. In this case, *without replacement,* the two probabilities are different because on the second selection there are fewer things to pick from.

b) Suppose it costs a player $1 to play a game in which two Skittles are drawn without replacement. If the player wins $5 if he/she draws two pinks, $3 if he/she draws two yellows and $2 if he/she draws two greens, what is the expected value of this game to the player

Expected Value = [($5)(2/72)+($3)(6/72)+($2)(12/72)] - $1
= [$52/72]-$1
= -$20/72
=-0.27777

So the player expects to lose about 28 cents per play.

9. a) The sample size of a poll is 500. The poll finds 37% in favor. Use the formula for the standard deviation of the sampling distribution $\sqrt{\frac{p(100-p)}{n}}$ to construct a 95% confidence interval for the percentage of the *population* who are in favor.

Standard Deviation = $\sqrt{\frac{37(100-37)}{500}} = \sqrt{\frac{(37)(63)}{500}} = \sqrt{4.662} = 2.16$.

A 95% Confidence Interval requires 2 standard deviations (The 68-95-99.7 Rule from Section 2E.) and so the margin of error is 2(2.16) = 4.32.

To get the confidence interval we must subtract 4.32 from 37 to get 32.68, and add 4.32 to 37 to get 41.32. The confidence interval is therefore (32.68, 41.32).

The interpretation is that although the survey yielded 37% of the sample in favor, we can be 95% confident that between 32.68% and 41.32% of the population are in favor. This comes from the ideas of probability theory and assumes, for example, that the sample was chosen in an unbiased manner.

a. Repeat part a) for a sample size of 50.

Standard Deviation = $\sqrt{\frac{37(100-37)}{50}} = \sqrt{\frac{(37)(63)}{50}} = \sqrt{46.62} = 6.83$.

A 95% Confidence Interval requires 2 standard deviations (The 68-95-99.7 Rule from Section 2E.) and so the margin of error is 2(6.83) = 13.66

To get the confidence interval we must subtract 13.66 from 37 to get 23.34, and add 13.66 to 37 to get 50.66. The confidence interval is therefore (23.34, 50.66).

Note that with the smaller sample size we get a much wider interval; the probability theory tells us we may have greater variation with a smaller sample – this has to do with the Law of Large Numbers in Section 2F.

b. Repeat part a) for a sample size of 5000.

Standard Deviation = $\sqrt{\frac{37(100-37)}{5000}} = \sqrt{\frac{(37)(63)}{5000}} = \sqrt{0.4662} = 0.68$.

A 95% Confidence Interval requires 2 standard deviations (The 68-95-99.7 Rule from Section 2E.) and so the margin of error is 2(0.68) = 1.36

To get the confidence interval we must subtract 1.36 from 37 to get 35.64, and add 1.36 to 37 to get 38.36. The confidence interval is therefore (35.64, 38.36). Note that we get a narrower interval with the larger sample. (Law of Large Numbers in 2F)

10. a) In January 2005, Time Magazine reported that President Bush had an approval rating of 53% obtained from a random sample of 1002 adults. Time put the margin of error at *approximately ±3%*. Find the actual value to 2 decimal places. (For more details see http://www.time.com/time/press_releases/article/0,8599,1018033,00.html)

Standard Deviation = $\sqrt{\frac{53(100-53)}{1002}} = \sqrt{\frac{(53)(47)}{1002}} = \sqrt{2.486} = 1.58$.

A 95% Confidence Interval requires 2 standard deviations (The 68-95-99.7 Rule from Section 2E.) and so the margin of error is 2(1.58) = 3.16.

b) Time also noted that the approval rating was up from 49% in December 2004. Assuming the same sample size (or thereabouts), explain why the rise from 49 to 53 need not be a good thing or the President.

With a margin of error of 3%, the December value could have been 49+3=52% while the January value could have been 53-3=50%. It is therefore possible that the President's rating had dropped by 2% rather than increased by 4% as shown with the two samples.

In fact things could be worse than this. There is a chance that through pure bad luck the samples weren't very representative of the population and the rating has fallen from (say, no calculation here, just speculation) 55% to 45% of the population. But it is just as likely that it has risen from 43% to 61%. We cannot be sure; the reasoning is inductive and not deductive; but the rising approval rating is probably good news for the President.

11. A random sample of 1000 boxes finds a (sample) mean of 106 nails per box. Given that the standard deviation is known to be 2 nails, give a 95% confidence interval for the (population) mean number of nails.

The Central Limit Theorem for means (Section 2F) is similar to that of proportions. It says that for a 95% confidence interval, the mean will lie within $\frac{2\sigma}{\sqrt{n}}$ of $\bar{x}$. Here

$\frac{2\sigma}{\sqrt{n}} = \frac{2(2)}{\sqrt{1000}} = \frac{4}{31.62} = 0.13$ and $\bar{x} = 106$. So the 95% confidence interval is given by (106-0.13, 106+0.13)=(105.87,106.13) In this problem the standard deviation is small and the sample size large that we have a very narrow confidence interval.

This problem is essentially the same as #1 and #2 except that we are given σ explicitly whereas in problems involving proportions the formula for the standard deviation $\sigma = \sqrt{p(100-p)}$ appears within the formula.

Note that on the Second Test you will be given the formula for proportions as in #1 but will need to know the formula for the mean.

12. A tire company finds that the (sample) mean number of miles of a tire is 20,478. This is obtained from a sample size of 250, and the standard deviation is 168. Give a 95% confidence interval for the true mean number of miles for a tire.

The Central Limit Theorem for means (Section 2F) is similar to that of proportions. It says that for a 95% confidence interval, the mean will lie within $\frac{2\sigma}{\sqrt{n}}$ of $\bar{x}$. Here

$\frac{2\sigma}{\sqrt{n}} = \frac{2(168)}{\sqrt{250}} = \frac{336}{15.81} = 21.25$ and $\bar{x} = 20478$. So the 95% confidence interval is given by (20478-21.25, 20478+21.25)=(20456.75, 20499.25)

13. a) There are 86 people in the treatment group in a medical test. At the end of the study period, 43% say that they feel that their condition has improved. Find a 95% confidence interval for the percentage of the population that will be improved by the drug.

This is a proportion problem and we need the formula provided in #1.

Standard Deviation = $\sqrt{\frac{43(100-43)}{86}} = \sqrt{\frac{(43)(57)}{86}} = \sqrt{28.5} = 5.39$.

A 95% Confidence Interval requires 2 standard deviations (The 68-95-99.7 Rule from Section 2E.) and so the margin of error is 2(5.39) = 10.78

To get the confidence interval we compute 43±10.78 get 32.22 and 53.78. The confidence interval is therefore (32.22, 53.78).

a. There are 68 people in the placebo group in a medical test. At the end of the study period, 36% say that they feel that their condition has improved. Find a

95% confidence interval for the percentage of the population for whom the placebo effect will make them feel improved by the drug.

Standard Deviation = $\sqrt{\frac{36(100-36)}{68}} = \sqrt{\frac{(36)(64)}{68}} = \sqrt{33.88} = 5.82$.

A 95% Confidence Interval requires 2 standard deviations (The 68-95-99.7 Rule from Section 2E.) and so the margin of error is 2(5.82) = 11.64

To get the confidence interval we compute 36±11.64 to get 24.36 and 47.64. The confidence interval is therefore (24.36, 47.64).

b. Do the results of this study lead appear to lead to the conclusion that the drug is beneficial? Comment on what the confidence intervals seem to imply.

The results themselves (43% versus 36%) present a case, but the 95% confidence intervals have a lot of overlap. We cannot really be sure the drug is effective.