Victoria Haley
IST 772 Homework 2

Exercise 1.

Fair coin flip landed on heads 4 times out of 9.
R code for 100,000 trials: table(rbinom(n=100000, size=9, prob=0.5))
Result:

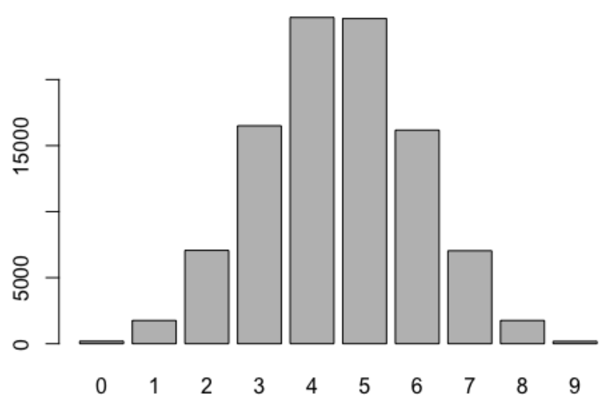
```
> table(rbinom(n=100000,size=9,prob=0.5))

  0     1     2     3     4     5     6     7     8     9
213  1788  7165 16238 24892 24540 16199  7007  1758   200
```

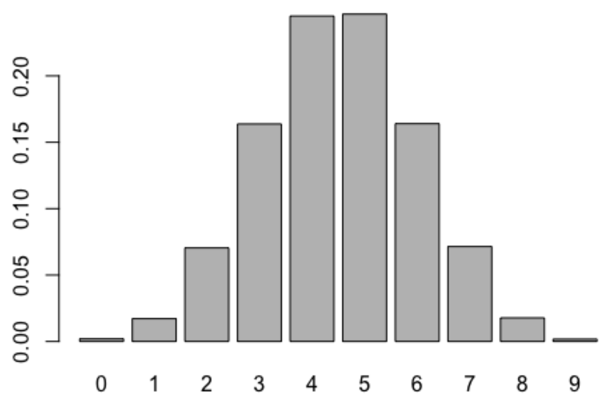
The results of the larger trial in R shows that the number of times that a flipped, fair coin is more likely to land heads-up between 4 and 5 times out 9 flips (events). The trial that I conducted with a real coin also matches up with the results of the larger trial. Seeing that 4 heads-up flips out of 9 occurred slightly more, I will use that to determine probability. So, using the above results, a trial of 100,000 fair-coin flips with 9 events will result in 4/9 coins landing heads-up 24,892 times for a probability of 24.892% ((24,892/100000)*100)

Exercise 2.

R-code for bar plot of results above: barplot(table(rbinom(n=100000, size=9, prob=0.5)))



R-code for bar plot of probability of results above: barplot(table(rbinom(n=100000, size=9, prob=0.5))/100000)



Bar plot 1 shows the distribution of the number of times each event occurred throughout all 100,000 trials. Bar plot 2 shows the relative frequency of a certain event happening throughout a "universe" of trials of coin flips. Both of these plots show that flipping a coin 9 times will likely result in landing heads-up 4 or 5 times out of the 9 flips. Both graphs show a normal distribution, with 2 symmetric tails that gradually curve up towards a peak in the middle. I believe that the graphs took on this shape because there are many different variations in the coin flip and not every coin flip was exactly the same, and the variations cancelled each other out and lead to the average number of heads-up landings to fall in the middle. This is also why the center of the graph is where it is, because the occurrences of heads-up landings less than 4 and greater than 5 fall outside of the mean.

Problem 6.

Table 1:

	Pass	Fail	Row totals
	80	20	100
Column totals	80	20	100
	Pass	Fail	Row totals
High-school student	33	17	50
College student	47	3	50
Column totals	80	20	100

The one additional piece of information provided (only 3 college students failed) was enough to complete the table because only logic and basic math skills were needed at that point. Knowing that there are 50 high-schoolers and 50 college students and that only 80 out of the 100 students passed, I was able to deduce that if 3 college students failed, 17 high-schoolers failed (3+17 =20). From there, I was also able to determine that 47 college students passed (3+47 =

50) and subtracted 47 from 80 to determine the number of high-schoolers that passed (80-47 =33).

Table 2:

	Pass	Fail	Row totals
High-school student	33/100 = 0.33	17/100 = 0.17	0.50
College student	47/100 = 0.47	3/100 = 0.03	0.50
Column totlas	0.80	0.20	1

The pass rates for high-school students is 0.33 or 33%, meaning that 33% of high-school students will pass the test.

Problem 7.

	Pass	Fail	Row totals
No default risk	87,266	5996	93,262
Default risk	6,669	69	6,738
Column totals	93,935	6,065	100,000

The percentage of customers that both pass the test and do not have their homes repossessed is 87.3% (87,266/100,000)

I was able to complete the table with the little information given in the problem by first figuring out how many failed screenings there were. Knowing that 5996 of the homeowners with no risk of defaulting failed the test, and that there were 6065 total fails, I was able to determine that 69 of the failed screenings were of homeowners with risk of defaulting (6065-5996). To determine the number of passed screenings, I took the clue from the beginning of the problem (71 out of 100,000 homes repossessed yearly) and applied that to the pass column total and found that 6,669 of the screenings that passed had risk of defaulting (93935*(71/100000)). Lastly, I subtracted 6,669 from 93,935 to find the last cell (93935 – 6669 = 87266). To check my answers, I added the rows and column totals to make sure that they equaled 100,000.

Problem 8.

Table from #7

	Pass	Fail	Row totals
No default risk	87,266	5996	93,262
Default risk	6,669	69	6,738
Column totals	93,935	6,065	100,000

Info from above table:

Customer who has no default risk and passes the test = 87.266%

Customer who has no default risk and fails the test = 5.996%

Customer who has default risk and passes the test = 6.669%

Customer who has default risk and fails the test = 0.069%

Probability table

	Pass	Fail	Row totals
No default risk	0.87266	0.05996	0.93262
Default risk	0.06669	0.00069	0.06738
Column totals	0.93935	0.06065	1

Probability table with isolated failed screening

	Fail
No default risk	0.05996
Default risk	0.00069
Column totals	0.06065

Probability table with isolated failed screening and normalized probabilities

	Fail	Normalized fails
No default risk	0.05996	0.9886232
Default risk	0.00069	0.01137675
Margin	0.06065	1

The probability of a new customer who fails the screening having their home repossessed is 1.14%.

I came to this number by taking the table from #7 and converting the occurrences into probabilities (Probability table). Then, I isolated the table so that only the failed screenings were listed (Probability table with isolated failed screening.) the last step was to normalize the failed screenings (raw fails probabilities/column margin)*100 in order to find the probability, in percent, of a new customer who fails the screening having their home repossessed.

Math: (raw fails probability of default risk/column margin)*100

(0.00069/0.0605) = 0.01137675

0.01137675 * 100 = 1.137675