Answers to Practice for Midterm 3 - Intro Stats

1. First, we need to figure out on how many of the trials we expect to have 4 heads. Each trial is Binomial with $n = 5, p = 1/2, k = 4$. Hence, the probability of getting exactly 4 heads is $\binom{5}{4}(1/2)^4(1/2) = 5/32$. We now do a hypothesis test on the box that has 5/32 as the fraction of 1's and 27/32 as the fraction of 0's. We are drawing 10 times. Hence, $EV_{sum} = (10)\cdot(5/32) = 50/32 = 1.5625$ and $SE^+_{sum} = \sqrt{10}\sqrt{\frac{10}{9}}\sqrt{(5/32)(27/32)} \approx 1.21$. Since we are drawing less than 25 times, the appropriate hypothesis test is the one-sample t-test. We have $t_9 = \frac{3-1.5625}{1.21} \approx 1.19$. Consulting the t-table with 9 degrees of freedom we have a p-value between 25% and 10%. Hence, we accept H_0 and conclude that the coins are fair.

2. a) We need to the create the box. Based on the given probabilities 400 of the 20 year-olds and 2100 of the 30 year-olds own cars. Hence, our box has 2500 people who own cars and 2500 people who do not. Hence, our approximate box, based on our sample, is: 0,1, since it is equally likely that we pick someone with a car as we are that we pick someone without a car. So we expect 2500 people in the room to own cars. i.e., $EV_{sum} = 2500$, but we need the SE. We are drawing without replacement from the entire population of 20 and 30 year-olds. Hence, we should use the correction factor, but we don't know the population size so we assume $CF \approx 1$. We have $SE_{sum} \approx \sqrt{5000}\sqrt{(.5)(.5)} \approx 35.36$. Thus our C.I. is 2500 ± 106.

b) If we were to pick 5000 people many times and find the 99.7% C.I. for each of these samples, about 99.7% of the C.I.'s should cover the actual number of people in the room who own a car.

3. a) The box has 5 tickets of \$2 and 8 tickets of -\$1. We are drawing 100 times. Hence, $EV_{sum} = (100)(2/13) = 15.38$ and $SE_{sum} = \sqrt{100}\sqrt{\frac{28}{13} - (2/13)^2} \approx 1.46$. Hence, we expect to win \$15.38, give or take \$1.46.

b) Our box has 2/13 as the fraction of 1's and 11/13 as the fraction of 0's. We are drawing 400 times. Hence $EV_{\%} = (2/13)\cdot 100\% \approx 15.38\%$ and $SE_{\%} = \frac{\sqrt{(2/13)(11/13)}}{\sqrt{400}} \cdot 100\% \approx 1.8\%$. Hence, we expect to win 15.38% of the time, give or take 1.8%.

c) We expect to get 20 blues and got 11. Using the same box as in (b), we have $SE_{sum} = \sqrt{130}\sqrt{(2/13)(11/13)} \approx 4.11$. We have a z-stat of $z = \frac{11-20}{4.11} \approx 2.2$. Using the Normal chart we have a p-value of about 1.4%. Since highly statistically significant mean $\alpha = 1\%$, we do not have such evidence. Hence, we conclude that the observed number of blues is not different from expected.

d) First, we create the Expected frequency column:

Color	Observed frequency	Expected frequency
Red	48	50
Black	51	50
Blue	11	20
Green	20	10
	130	130

Our χ^2 statistics is

$$\chi^2_3 = \frac{(48-50)^2}{50} + \frac{(51-50)^2}{50} + \frac{(11-20)^2}{20} + \frac{(20-10)^2}{10} = 14.15$$

Consulting the χ^2 table we have a p-value of less than 1%. Hence, we would conclude that the wheel is not fair.

4. We need to calculate the expected table. From the observed table we have the following:

	A	B	C	Total
2009	20	30	14	$64 = 32\%$
2010	16	34	4	$54 = 27\%$
2011	24	10	6	$40 = 20\%$
2012	30	6	6	$42 = 21\%$
Total	90	80	30	200

We then determine the expected table:

	A	B	C	Total
2009	28.8	25.6	9.6	$64 = 32\%$
2010	24.3	21.6	8.1	$54 = 27\%$
2011	18	16	6	$40 = 20\%$
2012	18.9	16.8	6.3	$42 = 21\%$
Total	90	80	30	200

We have $(4-1)(3-1) = 6$ degrees of freedom, so we calculate the χ^2 statistic:

$$\chi^2_6 = \frac{(20-28.8)^2}{28.8} + \frac{(16-24.3)^2}{24.3} + \frac{(24-18)^2}{18} + \frac{(30-18.9)^2}{18.9} + \frac{(30-25.6)^2}{25.6} + \frac{(34-21.6)^2}{21.6}$$
$$+\frac{(10-16)^2}{16} + \frac{(6-16.8)^2}{16.8} + \frac{(14-9.6)^2}{9.6} + \frac{(4-8.1)^2}{8.1} + \frac{(6-6)^2}{6} + \frac{(6-6.3)^2}{6.3},$$

which gives

$$\chi^2_6 = 35.2.$$

Consulting the χ^2 chart with 6 degrees of freedom, we see that our p-value is much less than 1%. Hence, we conclude that class year and highest grade are related.

5. We are testing $H_0 : r = 0$ against $H_A : r > 0$. The test statistic is a $t_{50-2} = t_{48}$ statistic, which we approximate with a z-stat (Normal curve is approximately a t-curve is the degrees of freedom are at least 25). Hence, we use

$$z \approx t_{48} = \frac{.29\sqrt{48}}{\sqrt{1-(.29)^2}} \approx 2.099$$

Using the Normal curve, the area to the left of 2.1 is $\frac{100-96.43}{2} = 1.785\%$. Since our p-value is greater than the significance level (1%), we do *not* reject H_0. We conclude that, based on this data, there is no positive correlation between parental income and SAT scores.