DAWSON COLLEGE

MATHEMATICS DEPARTMENT

Probability & Statistics

201-BZS-05

Thursday, May 23, 2019 - 14:00-17:00

Instructor: Rodney Acteson

Name:______________________________

Student ID# ________________________

Section:_______________________

- This exam contains 14 problems on 15 pages. Check to see if any pages are missing.
- Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page, and please indicate that you have done so.
- Give the work in full. Write and arrange your exercise in a legible and orderly manner.
- For Hypothesis Testing: Always state the null and alternate hypothesis, decision rule and conclusion.
- You are only permitted to use the Sharp EL531XG or Sharp EL-531X calculator.
- This examination booklet must be returned intact.
- Tables and Formula Sheets are provided separately.

1. [9 Marks] Consider the following data set:

0.11	0.20	0.37	0.46	0.65	0.69	0.70	0.71	0.72	0.78
0.89	1.14	1.18	1.34	1.38	1.38 /	1.50	1.53	1.81	1.88
1.99	2.05	2.24	2.32	2.36	2.54	2.72	2.74	2.74	2.87
2.87 /	3.26	3.32 /	4.47	5.05 /	6.06	6.08	7.28 /	8.87	8.98

a. Find the median, mode, first quartile and 40^{th} percentile for the data set.
b. Construct a frequency distribution table for the data, using 6 classes.
c. The actual sample mean is 2.51. Estimate the sample mean from the frequency distribution table in part b.

a) $n = 40$

location of median : $\frac{n+1}{2} = \frac{41}{2} = 20.5 \rightarrow$ median $= \frac{1.88 + 1.99}{2}$

$= \boxed{1.935}$

$\boxed{\text{No mode}}$

$Q_1 = \frac{.78 + .89}{2} = \boxed{0.835}$

$P_{40} =$ 40% of 40 $= .4(40) = 16$ $\quad P_{40} = \frac{1.38 + 1.50}{2} = \boxed{1.44}$

b) range $= 8.98 - 0.11 = 8.87$

class width $\approx \frac{8.87}{6} \approx 1.48$ (use 1.5)

Class	boundaries	f	Mark	f(m)
1	-0.005 – 1.495	16	.745	11.92
2	1.495 – 2.995	15	2.245	33.675
3	2.995 – 3.495	2	3.745	7.49
4	3.495 – 5.995	2	5.245	10.49
5	5.495 – 7.495	3	6.745	20.235
6	7.495 – 8.995	2	8.245	16.49
				100.3

B)

c) $\bar{X} \approx \frac{\sum f m_i}{n}$

$\approx \frac{100.3}{40}$

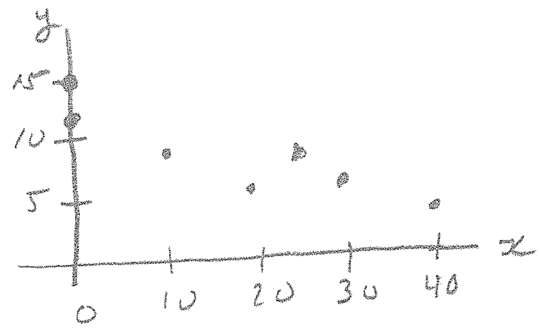
$\boxed{\bar{X} \approx 2.5075}$

2. [6 Marks] A survey was taken at a university campus. Students were asked: "How many hours per week are you employed?" and "How many credit hours are you currently registered for?"

Hours Employed (x)	10	20	0	30	40	25	0
Credit hours (y)	9	6	15	6	3	9	12

a. Draw a scatter diagram of these data.
b. What is the correlation coefficient?
c. Find the line of best fit.
d. Predict the number of credit hours for a student working 15 hours per week?

a)



b) correlation coefficient

$\boxed{r \approx -0,91}$

c) line of best fit

$y = ax + b$

$\boxed{y = 12,88 - 0,242x}$

d) $x = 15$ hours

$y = 12,88 - 0,242(15) = \boxed{9,26}$ credit hours.

3. [6 Marks] The distributions of two independent random variables X and Y are given below:

$X = x$	$P(X = x)$
1	0.1
2	0.2
3	0.3
4	0.4

$Y = y$	$P(Y = y)$
1	1/4
2	1/4
3	1/4
4	1/4

a. Calculate the mean and variance of X.

b. Calculate the probability $P(X \le 2 \text{ and } Y < 4)$.

x	$P(x)$	$x P(x)$	$x^2 P(x)$
1	.1	.1	.1
2	.2	.4	.8
3	.3	.9	2.7
4	.4	1.6	6.4
TOTA		3	10

a) $\mu = E(x) = \Sigma x P(x) = \boxed{3}$

$\sigma^2 = E(x^2) - (E(x))^2 = 10 - 3^2 = \boxed{1}$

b) Since X and Y are independent

$P(X \le 2 \text{ and } Y < 4) = P(X \le 2)\, P(Y < 4)$

$= (.3)\,(.75)$

$= \boxed{.225}$

4. [6 Marks] Consider the experiment of tossing 2 fair dice. Events A and B are defined as follows:

A : {The sum of the numbers showing is odd}

B : {The sum of the numbers showing is 9, 11 or 12}

a. Calculate $P(A \text{ or } B)$.

b. Are the events A and B independent? Justify your answer.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$P(A) = \frac{1}{2}$

$P(B) = \frac{7}{36}$

$P(A \text{ and } B) = \frac{6}{36} = \frac{1}{6}$

a) $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

$= \frac{18}{36} + \frac{7}{36} - \frac{6}{36} = \boxed{\frac{19}{36}} \approx .5278$

b) A and B are independent, if $P(A) = P(A|B)$

$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{6/36}{7/36} = \boxed{6/7 \approx .857}$

$P(A) = \frac{1}{2} \neq P(A|B) = 6/7$

$\therefore$ the events are dependent.

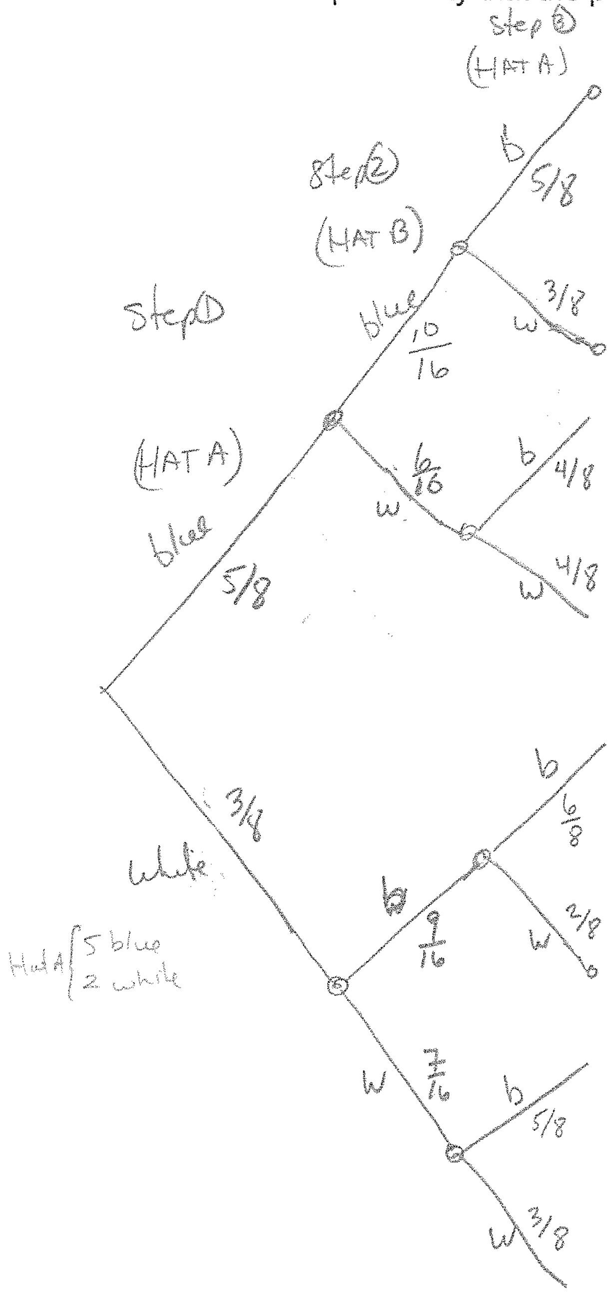
5. [6 Marks] Suppose

Hat A contains 5 blue and 3 white poker chips and
Hat B contains 9 blue and 6 white poker chips.

Consider the following 3 step process:

Step 1: Randomly select a poker chip from Hat A and place it in Hat B
Step 2: Randomly select a poker chip from Hat B and place it in Hat A
Step 3: Randomly select a poker chip from Hat A

What is the probability that the poker chip selected in Step 3 is blue?



$P(b\,b\,\boxed{b}) = \left(\frac{5}{8}\right)\left(\frac{10}{16}\right)\left(\frac{5}{8}\right)$

$bbw = \left(\frac{5}{8}\right)\left(\frac{10}{16}\right)\left(\frac{3}{8}\right)$

$bw\boxed{b} = \left(\frac{5}{8}\right)\left(\frac{6}{16}\right)\left(\frac{4}{8}\right)$

$bww = \left(\frac{5}{8}\right)\left(\frac{6}{16}\right)\left(\frac{4}{8}\right)$

$wb\boxed{b} = \left(\frac{3}{8}\right)\left(\frac{9}{16}\right)\left(\frac{6}{8}\right)$

$wbw = \left(\frac{3}{8}\right)\left(\frac{9}{16}\right)\left(\frac{2}{8}\right)$

$ww\boxed{b} = \left(\frac{3}{8}\right)\left(\frac{7}{16}\right)\left(\frac{5}{8}\right)$

$www = \left(\frac{3}{8}\right)\left(\frac{7}{16}\right)\left(\frac{3}{8}\right)$

$P(\text{blue on step 3}) = \left(\frac{5}{8}\right)\left(\frac{10}{16}\right)\left(\frac{5}{8}\right) + \left(\frac{5}{8}\right)\left(\frac{6}{16}\right)\left(\frac{4}{8}\right) + \left(\frac{3}{8}\right)\left(\frac{9}{16}\right)\left(\frac{6}{8}\right) + \left(\frac{3}{8}\right)\left(\frac{7}{16}\right)\left(\frac{5}{8}\right)$

≈ 0.622

6. [6 Marks] How many different orderings of the word "REGRESSIONS"
   a. are there?
   b. are there, if it must begin and end with the letter R and the 3 S's must always be together?
   c. are there, if the word "ION" must appear in each order?

a) $\dfrac{11!}{\underset{\text{"R"}}{2!}\,\underset{\text{"E"}}{2!}\,\underset{\text{"S"}}{3!}} = \boxed{1663200}$

b) $\boxed{R}$ $\boxed{SSS}$ 6 other letters $\boxed{R}$

$\dfrac{7!}{\underset{\text{"E"}}{2!}} = \boxed{2520}$

c) $\boxed{ION}$ 8 other letters

$\dfrac{9!}{\underset{\text{'R'}}{2!}\,\underset{\text{'E'}}{2!}\,\underset{\text{'S'}}{3!}} = \boxed{15120}$

7. [9 Marks] The probability of randomly selecting a defective product off an assembly line is 5%. If 10 products are randomly selected, what is the probability that

a. at most one product is defective?
b. the last product selected is the 3^{rd} defective one found?
c. the first defective product selected is the 6^{th} product selected?

a) $Bin(p=.05, n=10)$

P(at most one defective product) $= P(x=0) + P(x=1)$

$= {}_{10}C_0 (.05)^0 (.95)^{10} + {}_{10}C_1 (.05)^1 (.95)^9$

$= 0.5987 + .31512$

$= \boxed{0.9139}$

b) P(last one selected is 3^{rd} defective)

$= B(x=2, p=.05, n=9) \times p$

$\approx 0.06285 \times .05$

$\approx 3.143 \times 10^{-3} = \boxed{0.003143}$

c) P(first defective product is the 6^{th} producted selected)

$= (.95)^5 (.05)$

$= \boxed{0.0387}$

8. [8 Marks] Health Canada reported that 20% of preschoolers lack the required immunization. Let x be the number of preschoolers who lack the required immunization in a group of 64 students.

   a. What is the expected value and variance of x?
   b. What is the probability that exactly 10 preschoolers lack the required immunization?
   c. Use the Normal distribution to approximate the probability that there are at most 10 preschoolers without immunization.

a) $X \sim Bin(p = .2, n = 64)$

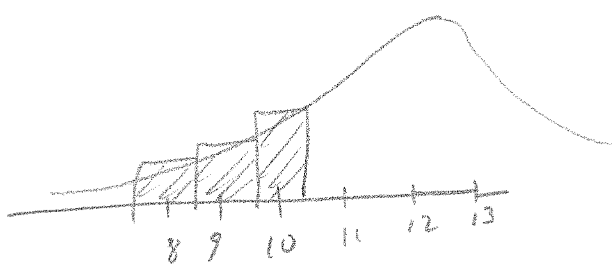
$E(x) = np = (.2)(64) = \boxed{12.8}$

$\sigma^2 = npq = (.2)(64)(.8) = \boxed{10.24} \rightarrow \sigma = 3.2$

b) $P(X = 10) = {}_{64}C_{10} (.20)^{10} (.8)^{54}$

$\approx \boxed{0.0907}$

c) $Bin(X \leq 10) \approx P(X < 10.5 \mid \mu = 12.8, \sigma = 3.2)$



$z = \frac{x - \mu}{\sigma} = \frac{10.5 - 12.8}{3.2} \approx -0.72$

$Bin(X \leq 10) \approx P(z < -0.72) = .5 - .2642$

$\approx \boxed{0.2358}$

9. [6 Marks] Consider the following probability density function:

$$f(x) = \begin{cases} \frac{2}{x^3} & \text{if } x \geq 1 \\ 0 & \text{if } x < 1 \end{cases}$$

a. Calculate the expected value of x.
b. Calculate the probability that x is greater than its mean.

a) $E(x) = \int_1^\infty x\left(\frac{2}{x^3}\right)dx = \int_1^\infty 2x^{-2}dx = \frac{2x^{-1}}{-1}\Big|_1^\infty$

$= \frac{-2}{x}\Big|_1^\infty = \frac{-2}{\infty} - \left(\frac{-2}{1}\right) = \boxed{2}$

b) $P(x \geq 2) = \int_2^\infty \frac{2}{x^3}dx = \int_2^\infty 2x^{-3}dx = \frac{-1}{x^2}\Big|_2^\infty$

$= \frac{-1}{\infty} + \frac{1}{2^2} = 1/4$

10. [6 Marks] On average, 14.8 people approach the help desk at a shopping mall every hour. Let X be defined as the number of people who approach the help desk during any hour.

a. What is the probability distribution function for X?

b. What is the probability that exactly 10 people approach the help desk in the next hour?

c. What is the probability that 5 or more people approach the help desk in the next half hour?

a) $X \sim$ Poisson distribution ($\lambda = 14.8$ people / 1 hour)

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

b) $$P(x=10) \quad \frac{14.8^{10} e^{-14.8}}{10!} \approx \boxed{0.0519}$$

c) $P(X \geq 5 \mid \lambda = 7.4$ people / ½ hour$)$

$$= 1 - P(X \leq 4)$$

$$= 1 - \left(\frac{7.4^0 e^{-7.4}}{0!} + \frac{7.4 e^{-7.4}}{1!} + \frac{7.4^2 e^{-7.4}}{2!} + \frac{7.4^3 e^{-7.4}}{3!} + \frac{7.4^4 e^{-7.4}}{4!} \right)$$

$$= 1 - e^{-7.4} \left(1 + 7.4 + \frac{7.4^2}{2} + \frac{7.4^3}{6} + \frac{7.4^4}{24} \right)$$

$$= \boxed{0.8605}$$

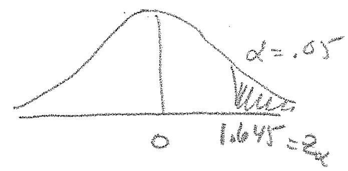
11. [9 Marks] An article in a local newspaper stated that at most 24% of all those who gamble at the local casino come away winners. A random sample of 127 gamblers leaving the casino were asked if they had won, 38 said yes.

   a. Does the data support the newspaper article? Use a significance level of 5% to conduct a test of hypothesis. State the hypotheses, decision rule and conclusion.
   b. How large of a sample size is required to estimate the true percentage of winners to within 1% with 95% confidence, if no preliminary estimate is available?
   c. Calculate the power of the test if the true percentage of winners is 26%.

a) $H_0: p \le .24$

$H_a: p > .24$

$\hat{p} = \frac{x}{n} = \frac{38}{127} \approx 0.299$



$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{38}{127} - .24}{\sqrt{\frac{.24(.76)}{127}}} \approx \underline{1.56}$$

Since $Z_0 = 1.56 < Z_\alpha = 1.645$ then we fail to $\underline{\text{reject } H_0}$.

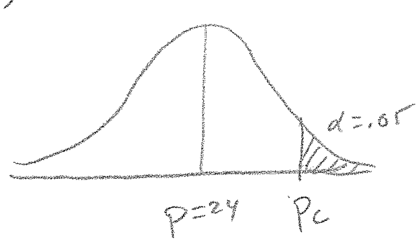
The data supports the newspaper article.

b) If no preliminary estimate is available we use $p = \frac{1}{2}$

$$E = Z_{\alpha/2}\sqrt{\frac{pq}{n}} \rightarrow n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 pq$$

$$= \left(\frac{1.96}{.01}\right)^2 \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \boxed{9604}$$

c)



$$P_c = .24 + 1.645\sqrt{\frac{.24(.76)}{127}}$$

$\approx \underline{.302}$

$$\beta = P(\hat{p} < .302 \mid p = .26)$$

$$= P\left(z < \frac{.302 - .26}{\sqrt{\frac{.26(.74)}{127}}}\right) = P(z < 1.08)$$

$= .5 + .3599$

$= .8599$

Power $= 1 - \beta = 1 - .8599$

$\approx \boxed{0.14}$

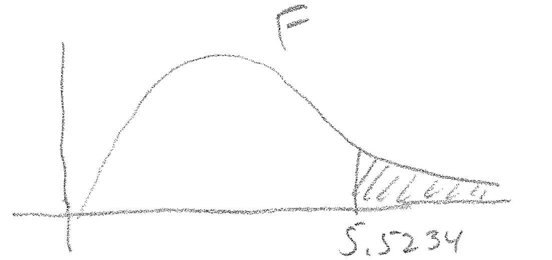
12. [9 Marks] A random sample of 10 adult male African Bush elephants was selected. The sample had a mean weight of 6.0 tons with a standard deviation of 1.2 tons. While a randomly selected sample of 7 adult male Indian elephants had a mean weight and standard deviation of 5.4 tons and 0.8 tons, respectively. Assume that the populations are normally distributed and have equal variances.

a. Is there sufficient evidence to support the claim that the population variances are equal? (Use $\alpha = 0.05$)

b. Is there sufficient evidence to support the claim that the average weight of an adult male African Bush elephant is more than that of an adult male Indian elephant? (Use $\alpha = 0.05$)

a) $H_0: \sigma_a^2 = \sigma_b^2$ $\qquad H_0: \frac{\sigma_A^2}{\sigma_B^2} = 1$

$H_a: \sigma_a^2 \neq \sigma_b^2$ $\qquad H_a: \frac{\sigma_A^2}{\sigma_B^2} \neq 1$



$$F^* = \frac{S_1^2}{S_2^2} = \frac{(1.2)^2}{(.8)^2} = 2.25$$

$F_{.025}(V_1 = n_1 - 1, V_2 = n_2 - 1)$

$F_{0.025}(V_1 = 9, V_2 = 6) = 5.5234$

Since $F^* = 2.25 < 5.5234$ then we fail to reject H_0.

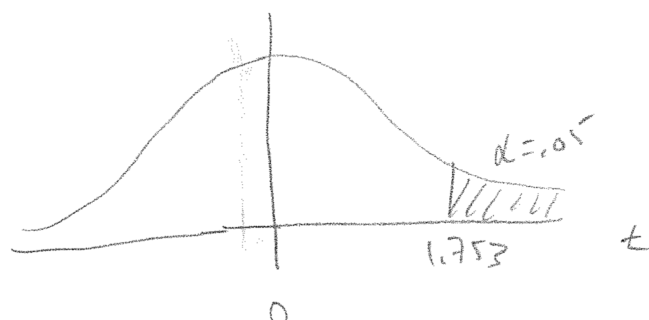
there is insufficient evidence to support the claim that the variances are unequal.

b) $H_0: \mu_A = \mu_I$

$H_a: \mu_A > \mu_I$

$$t^* = \frac{6.0 - 5.4}{\sqrt{1.12\left(\frac{1}{10} + \frac{1}{7}\right)}} \approx 1.15$$

$$t^* = \frac{\bar{x}_A - \bar{x}_I - 0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$



$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$= \frac{(10-1)(1.2)^2 + (7-1)(.8)^2}{10+7-2}$$

$$\approx 1.12$$

$t_{.05}(df = 15) = 1.753$

Since $t^* = 1.15 < 1.753$ there is insufficient evidence to support the claim that the average weight of an African elephant is more than the Indian elephant

13. [7 Marks] Five radar machines were placed on residential streets throughout a local borough where the posted speed limit is 50 km/hr. The average speed of motorists between the hours of 6 a.m. and 9 a.m. and between the hours of 9 a.m and noon are given below:

	Radar Machine	1	2	3	4	5
period 1:	6 a.m. to 9 a.m	62	74	61	67	53
period 2:	9 a.m to noon	55	60	57	58	51
	$d_i = p_1 - p_2$	7	14	4	9	2

$\Sigma d_i = 36$

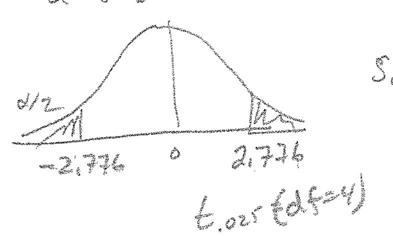
$\Sigma d_i^2 = 346$

Is there a significance difference between the motorist's average speed during the two time periods? (Use a ~~1~~% level of significance.) 0,05

$H_0: \mu_d = 0$

$H_a: \mu_d \neq 0$

$\alpha = .05$



$\bar{d} = \frac{\Sigma d}{n} = \frac{36}{5} = 7.2$

$s_d = \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n-1}}$

$= \sqrt{\frac{346 - \frac{(36)^2}{5}}{4}}$

$= \sqrt{21.7}$

$= \boxed{4.6583}$

$t^* = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$

$= \frac{7.2 - 0}{4.6853/\sqrt{5}}$

$t^* = \boxed{3.46}$

Since $t^* = 3.46 > 2.776$ then we reject H_0.

there is a significant difference between the two time periods.

14. [7 Marks] Conduct an appropriate test to determine whether a die used at the casino is a fair die. The die was rolled 300 times and the results are shown below. (use $\alpha = 0.05$)

Possible Results	1	2	3	4	5	6
Observed Results	40	37	59	63	53	49

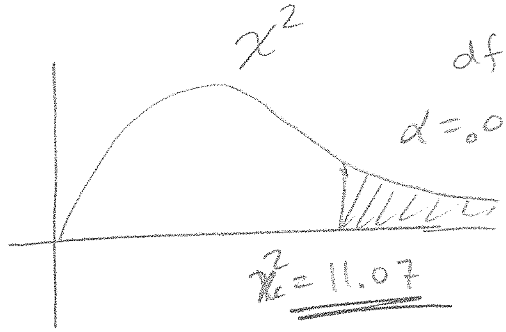
H_0: $p_i = 1/6$, for $i = 1, 2, 3, 4, 5, 6$ $\qquad E_i = np_i \ \forall i$

H_a: not as in H_0 $\qquad E_i = 300(\frac{1}{6}) = \underline{50}$

$$\chi^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(40-50)^2}{50} + \frac{(37-50)^2}{50} + \frac{(59-50)^2}{50} + \frac{(63-50)^2}{50} + \frac{(53-50)^2}{50} + \frac{(49-50)^2}{50}$$

$$\boxed{\chi^2 = 10.58}$$



$df = k-1 = 5$

$\alpha = .05$

Since $\chi^2 = 10.58 < \chi^2_c = 11.07$ then we fail to reject H_0.