AP Statistics Name _________________

Review Ch. 12 Chi-Square & Linear

1. Each fruit fly litter consists of 100 fruit flies. I expect half to have red eyes. I have four samples and here is the following number of fruit flies with red eyes: 60, 80, 20, and 10. What is my chi-squared value?

a. 18 b. 32 c. 50 d. 70 e. 110

2. Which of the following is **not** true of the $c^2$ probability density function?

a. For small degrees of freedom, the curve has right skewness.
b. As the degrees of freedom increase, the curve approaches a normal curve.
c. $c^2$ is defined only for positive values of any given variable.
d. The total area under the $c^2$ is equal to one.
e. The $c^2$ is only used for numerical data sets.

3. There are four surgical methods currently being used to place various medical implants in patients. Patients will be monitored and their pain level recorded as severe, moderate or mild for 1 week after the implant. What are the degrees of freedom for these four methods and pain levels?

a. 2 b. 3 c. 4 d. 6 e. 12

Questions 4-7 refer to the following set of data.

A group of AP statistics students wanted to see if all bags of M&Ms had the same proportion of colors no matter which type of M&M product they purchased. To test this, the students randomly sampled king-size bags of each type and recorded their findings.

| | Red | Blue | Yellow | Green | Orange | Brown |
|---|---|---|---|---|---|---|
| Plain | 20 | 18 | 15 | 10 | 14 | 12 |
| Peanut | 8 | 6 | 8 | 25 | 5 | 7 |
| Almond | 7 | 11 | 10 | 12 | 10 | 9 |

4. How would the expected count for the number of green peanut M&Ms be determined?

a. $\frac{25}{47}$ b. $\frac{25*47}{47}$ c. $\frac{59*47}{59}$ d. $\frac{59*47}{207}$ e. $\frac{25*47}{207}$

5. What are the correct degrees of freedom for this table?

a. 18 b. 12 c. 10 d. 8 e. 4

6. For this table of M&Ms what would be the appropriate set of hypotheses?

   a. Ho: There is no association between the type of M&M and colors in the bag.
      Ha: There is an association between the type of M&M and colors in the bag.

   b. Ho: There is an association between the type of M&M and colors in the bag.
      Ha: There is no association between the type of M&M and colors in the bag.

   c. Ho: The proportions of M&Ms are not the same for all colors.
      Ha: The proportions of M&Ms are the same for all colors.

   d. Ho: The proportions of red M&Ms are the same no matter the type of M&M.
      Ha: The proportions of red M&Ms are not the same no matter the type of M&M.

   e. Ho: There is no difference in the proportions of M&Ms by color.
      Ha: There is a difference in the proportions of M&Ms by color.

7. What would be a valid conclusion based on the table of M&M values?

   a. With a $c^2 = 22.9$ and df = 10, there is insufficient evidence to reject the null claim that the proportion of M&Ms is the same for all colors.

   b. With a $c^2 = 22.9$ and df = 10, there is strong evidence to reject the claim that the proportion of M&Ms is different for each color and claim that the proportions are the same.

   c. With a $c^2 = 22.9$ and df = 10, there is strong evidence to reject the null. In other words, there is a difference in the proportion of M&Ms by color.

   d. With a $c^2 = 22.9$ and df = 10, there is strong evidence to reject the null. In other words, there is absolutely no association between the color and type of M&Ms.

   e. With a $c^2 = 22.9$ and df = 10, there is insufficient evidence to reject the null. In other words, there is no relationship between the colors and the type of M&Ms.

8. A cupcake store believes that half of the cupcakes sold are chocolate. The remaining half sold seems to be equally divided between vanilla and cherry. If an order has been placed for 300 cupcakes, how many should they expect to be requested in cherry?

   a. 300 b. 200 c. 150 d. 100 e. 75

9. A random sample of patrons of a local mall was surveyed regarding a public smoking area outside one of the mall entrances. The patrons were asked if they are agitated by the idea of a public smoking area for the mall and the mall management is curious if the smoking section equally agitates men and women. The results are recorded below. Given the table, what would be an appropriate set of hypotheses for this data?

**Public Smokers**

| | Agitated | Not Agitated |
|---|---|---|
| Males | 28 | 57 |
| Females | 39 | 31 |

a. Ho: Gender and agitation due to public smoking are not independent.
   Ha: Gender and agitation due to public smoking are independent.

b. Ho: Gender and agitation due to public smoking are independent.
   Ha: Gender and agitation due to public smoking are not independent.

c. Ho: Knowing people are agitated by smoking indicates their gender.
   Ha: Knowing people are agitated by smoking gives no indication of their gender.

d. Ho: There is no difference in gender and agitation by public smoking.
   Ha: There is a difference in gender and agitation by public smoking.

e. Ho: There is an association between gender and agitation by public smoking.
   Ha: There is no association between gender and agitation by public smoking.

10. A local bagel shop makes six types of bagels and eight types of cream cheese toppings. Assuming all the types of bagels and cream cheese are part of a study, how many degrees of freedom would this table have?

a. 7 b. 12 c. 14 d. 35 e. 48

11. In a recent opinion poll, likely voters were asked if they would continue to support health care reform if the cost per taxpaying citizen were increased by $800 per year. They also recorded the person's political preference.

| | Democrat | Republican | Independent | Libertarian |
|---|---|---|---|---|
| Support | 125 | 87 | 99 | 48 |
| Don't Support | 75 | 113 | 101 | 52 |

Is there evidence at a significance level of 0.01 that political preference is independent of one's support for the proposed increase?

a. Since P = 0.001, there is strong evidence that political preference and support are independent of each other. In other words, knowing their political preference gives no insight into their support of the increase.

b. Since P = 0.001, there is strong evidence that political preference and support are not independent of each other. In other words, knowing their political preference gives insight into their support of the increase.

c. Since P = 0.001, there is strong evidence that political preference and support are independent of each other. Since they are independent, there is no association between the two categories.

d. Since P = 0.001, there is insufficient evidence that political preference and support are independent of each other. Since they are not independent, there is no association between the two categories.

e. Since P = 0.001, there is insufficient evidence that political preference and support are not independent of each other. In this case, knowing their political preference would not help in knowing their support.

12. Which is not true for a chi-square homogeneity test?

I. Data is categorical in nature.
II. df is calculated using k-1, where k = the number of categories.
III. Chi-square distribution is right skewed.

a. I, II and III    b. I and II    c. II and III    d. I only    e. II only

13. Two different cereal companies are competing for the larger market share of three types of specialty cereals. A random sample of 300 potential customers sampled the cereals and selected their favorite and the results were recorded. Based on the data shown, what is the appropriate null hypothesis and is there a difference in selection by brand at $\alpha = 0.05$?

| | Cereal 1 | Cereal 2 | Cereal 3 |
|---|---|---|---|
| Company A | 51 | 48 | 47 |
| Company B | 43 | 65 | 46 |

a. Ho: There is no difference in cereals liked for either company brand. There is no difference in cereal chosen by brands as $P > \alpha$. ($0.22 > 0.05$)

b. Ho: There is no difference in cereals liked for either company brand. There is a difference in cereal chosen by brands as $P > \alpha$. ($0.22 > 0.05$)

c. Ho: There is no difference in cereals liked for either company brand. There is no difference in cereal chosen by brands as $P < \alpha$. ($0.22 < 0.05$)

d. Ho: There is a difference in cereals liked for either company brand. There is no difference in cereal chosen by brands as $P > \alpha$. ($0.22 > 0.05$)

e. Ho: There is a difference in cereals liked for either company brand. There is no difference in cereal chosen by brands as $P > \alpha$. ($0.22 > 0.05$)

Questions 14 – 15 refer to the following set of data.

A movie theater recorded the number of items sold for different types of films on family night.

| | Soda | Popcorn | Candy |
|---|---|---|---|
| Children's Movie | 70 | 83 | 47 |
| Action Movie | 61 | 49 | 20 |

14. If the manager wanted to see if there might be an association between the type of movie and snacks sold, which type test would be appropriate?

a. Chi-square goodness of fit
b. Chi-square test for homogeneity
c. Chi-square test for independence
d. Two-sample proportions test
e. Multiple sample proportions test

15. Is there evidence that the type of movie and snacks are not independent?

a. With $\chi^2 = 5.67$, df = 2, there is strong evidence at $\alpha = 0.10$ but not at $\alpha = 0.05$.
b. With $\chi^2 = 5.67$, df = 2, there is insufficient evidence at $\alpha = 0.10$.
c. With $\chi^2 = 5.67$, df = 2, there is strong evidence at $\alpha = 0.05$ but not at $\alpha = 0.10$.
d. With $\chi^2 = 5.67$, df = 5, there is insufficient evidence at $\alpha = 0.10$.
e. With $\chi^2 = 5.67$, df = 5, there is strong evidence at $\alpha = 0.10$ but not at $\alpha = 0.05$.

16. A geneticist hypothesizes that half of a given population will have brown eyes and the remaining half will be split evenly between blue and green-eyed people. In a random sample of 60 people from this population, the individuals are distributed as shown in the table above. What is the value of the $\chi^2$ statistic for the goodness of fit test on these data?

| Brown Eyes | Green Eyes | Blue Eyes |
|---|---|---|
| 34 | 15 | 11 |

(A) Less than 1
(B) At least 1, but less than 10
(C) At least 10, but less than 20
(D) At least 20, but less than 50
(E) At least 50

17. In an experiment, two different species of flowers were crossbred. The resulting flowers from this crossbreeding experiment were classified, by color of flower and stigma, into one of four groups, as shown in the table below.

| Flower Type Resulting from Crossbreeding | Number of Flowers Observed with These Colors |
|---|---|
| I: Magenta lower with green stigma | 115 |
| II: Magenta flower with red stigma | 49 |
| III: Red flower with green stigma | 32 |
| IV: Red flower with red stigma | 21 |

A biologist expected that the ratio of 9:3:3:1 for the flower types I:II:III:IV, respectively, would result from this crossbreeding experiment. From the data above, a $\chi^2$ value of approximately 8.04 was computed. Are the observed results <u>inconsistent</u> with the expected ratio at the 5 percent level of significance?
(A) Yes, because the computed $\chi^2$ value is greater than the critical value.
(B) Yes, because the computed $\chi^2$ value is less than the critical value.
(C) No, because the computed $\chi^2$ value is less than the critical value.
(D) No, because the computed $\chi^2$ value is greater than the critical value.
(E) It cannot be determined because some of the expected counts are not large enough to use th $\chi^2$ test.

18. Sophomore, junior and senior students at a high school will be surveyed regarding a potential increase in the extracurricular student activities fee. There are three possible responses to the survey question – agree with the increase, do not agree with the increase, or no opinion. A chi-square test will be conducted to determine whether the response to the question is independent of the class in which the student is a member. How many degrees of freedom should the chi-square test have?

(A) 9 (B) 6 (C) 4 (D) 2 (E) 1

19. Each person in a random sample of adults indicated his or her favorite color. The results, shown in the table below, are reported by age group of the respondents.

| | Red | Green | Blue | Other | Total |
|---|---|---|---|---|---|
| Under 30 | 20 | 42 | 16 | 36 | 114 |
| 30-50 | 24 | 35 | 24 | 25 | 108 |
| Over 50 | 25 | 22 | 35 | 10 | 92 |
| Total | 69 | 99 | 75 | 71 | 314 |

If choice of color is independent of age group, which of the following expressions is equal to the expected number of respondents who are aged 30 to 50, inclusive, and prefer green?

(A) $\frac{(99)(108)}{314}$ (B) $\frac{(69)(108)}{314}$ (C) $\frac{(35)(99)}{108}$ (D) $\frac{(35)(108)}{314}$ (E) $\frac{(99)(35)}{314}$

20. In order to plan its next advertising campaign, the Trendy Motor Vehicle Company is investigating whether the type of vehicle and the color of vehicle are related. Each person in a random sample of size 275 selected from the company's mailing list was classified according to the type (car or truck) and the color of vehicle he or she drove. The data are shown in the table below.

| | | Vehicle Color | | | | |
|---|---|---|---|---|---|---|
| | | Red | Black | White | Tan | Green |
| Vehicle Type | Car | 35 | 23 | 41 | 21 | 12 |
| | Truck | 27 | 55 | 39 | 12 | 10 |

Which of the following procedures would be most appropriate to use for investigating whether there is a relationship between vehicle type and color?
(A) A two-sample t-test
(B) A two-sample z-test
(C) A matched pairs t-test
(D) A chi-square goodness-of-fit test
(E) A chi-square test of independence

*Questions 21 – 27 refer to the following information:*
Data on x = number of powerboat registrations in Florida and y = number of manatee deaths in Florida for 31 randomly selected years was used to produce the following computer output. You can assume that all conditions needed for inference are met.

**Regression Analysis: Manatees versus Registrations**

| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | -40.641 | 5.840 | -6.96 | 0.000 |
| Registrations | 0.125006 | 0.007867 | 15.89 | 0.000 |

S = 7.87985 R-Sq = 89.7% R-Sq(adj) = 89.3%

Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 1 | 15677 | 15677 | 252.48 | 0.000 |
| Residual Error | 29 | 1801 | 62 | | |
| Total | 30 | 17477 | | | |

21. The estimate for the population regression line is

(a) predicted manatees = -40.641 + 5.84(registrations)
(b) predicted manatees = 0.125 + 0.007867(registrations)
(c) predicted manatees = 5.840 + 0.007867(registrations)
(d) predicted manatees = -40.641 + 0.125006(registrations)
(e) predicted manatees = 7.87985 + 0.125006(registrations)

22. The standard error of the slope is shown to be 0.0079. Interpret this value in context.

(a) The standard deviation of the number of powerboat registrations is 0.0079.
(b) The standard deviation of the number of manatees killed is 0.0079.
(c) If many samples were taken and the slope of each least squares regression line was recorded, the estimated standard deviation of these slopes is 0.0079
(d) If many samples were taken and the slopes of each least squares regression line were recorded, the difference in the estimated change in number of manatees killed per powerboat registration and the true change in number of manatees killed per powerboat registration is on average 0.0079.
(e) The distance between the true change in number of manatees killed per powerboat registration and the sample change in number of manatees killed per powerboat registration is 0.0079.

23. Which of the following is a confidence interval for the mean change in number of manatee deaths associated with an increase of 1 powerboat registration?

(a) -40.641 ± t*(5.84) (b) 0.125 ± t*(0.0079) (c) 0.125 ± t*(15.89)
(d) 0.125 ± t*(5.840) (e) 0.125 ± t*(15.89)

24. The value of the t test statistic for testing is

(a) 7.87985 (b) -6.96 (c) 15.89 (d) 252.48 (e) 89.3

25. What is the value of the degrees of freedom that would be used in determining the P=value in a hypothesis test of $H_o$: $\beta = 0$?

(a) 1 (b) 28 (c) 29 (d) 30 (e) 32

26. In a test of Ho: $\beta = 0$ versus Ha: $\beta \neq 0$ with a significance level of 0.05, the decision would be to

(a) Accept $H_o$ (b) fail to reject $H_a$ (c) fail to reject $H_o$ (d) reject $H_o$ (e) reject $H_a$

27. The estimated standard deviation of the residuals is

(a) 5.84 (b) 7.87985 (c) 0.0079 (d) 1801 (e) 89.7

*Questions 28 – 33 refer to the following information:*
Each person in a random sample of adults was asked to estimate the average number minutes of television watched per day and their yearly salary. The following data resulted.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Minutes of TV Watched | 36 | 57 | 60 | 62 | 70 | 76 | 101 |
| Annual Salary (in thousands) | 89 | 45 | 30 | 50 | 55 | 67 | 30 |

The data was used to produce the following regression output.

```
The regression equation is
Salary = 94.4 - 0.638 TV


Predictor      Coef   SE Coef      T      P
Constant      94.40     25.60   3.69  0.014
TV          -0.6382    0.3736  -1.71  0.148


S = 18.2023    R-Sq = 36.9%    R-Sq(adj) = 24.2%
```

The researcher wished to determine if these data provide evidence of a linear relationship between number of minutes spent watching television each day and annual salary. You can assume that all conditions required for inference are met.

28. The appropriate hypothesis are

(a) $H_o$: $\beta = 0$; $H_a$: $\beta \neq 0$ (b) $H_o$: $b = 0$; $H_a$: $b \neq 0$ (c) $H_o$: $\beta \neq 0$; $H_a$: $\beta = 0$
(d) $H_o$: $\beta = 0$; $H_a$: $\beta < 0$ (e) $H_o$: $\beta = 0$; $H_a$: $\beta > 0$

29. What is the value of the correlation coefficient for these data?

(a) -0.369 (b) -0.607 (c) 0.148 (d) 0.369 (e) 0.607

30. What proportion of variability in salary can be explained by the linear relationship between TV watched and salary?

(a) 0 (b) 0.148 (c) 0.369 (d) 0.374 (e) 18.202

31. The p-value for the test of $H_o$: $\beta = 0$ versus $H_a$: $\beta \neq 0$ is 0.148. A correct interpretation of this value in context is

(a) The probability that we committed a Type I error is 0.149.
(b) The probability that we committed a Type I or Type II error is 0.149.
(c) If we were to take many samples of people and ask their TV viewing habits and their salary, 14.9% of the tests would yield different results.
(d) If minutes watching TV and salary are not linearly related, we would get samples like this one or more extreme 14.9% of the time just by chance.
(e) If minutes watching TV and salary are associated, we would detect the relationship only 14.9% of the time.

32. Using a significance level of 0.05, the correct conclusion for the test of $H_o$: $\beta = 0$ versus $H_a$: $\beta \neq 0$ in context is

(a) Since the p-value is greater than the significance level, we accept the null hypothesis that minutes watching TV and salary are linearly related.
(b) Since the p-value is greater than the significance level, we reject the null hypothesis that minutes watching TV and salary are not linearly related.
(c) Since the p-value is greater than the significance level, we fail to reject the null hypothesis that minutes watching TV and salary are not linearly related.
(d) Since the p-value is greater than the significance level, we fail to reject the null hypothesis that minutes watching TV and salary are linearly related.
(e) Since the p-value is greater than the significance level, we have evidence that minutes watching TV and salary are not linearly related.

33. Which of the following is a correct description of a Type II error in context?

(a) We conclude that minutes watching TV and salary are not linearly related, when, in fact, they are linearly related.
(b) We conclude that minutes watching TV and salary are linearly related, when, in fact, they are not linearly related.
(c) We conclude that minutes watching TV and salary are not linearly related, when, in fact, they are not linearly related.
(d) We conclude that minutes watching TV and salary are linearly related, when, in fact, they are linearly related.
(e) A Type II error happens 2% of the time when we run similar test on similar samples.

**Free Response**

34. After getting trounced by your little brother in a children's game, you suspect the die he gave you to roll, may be unfair. To check, you roll it 60 times, recording the number of times each face appears. Do these results cast doubt on the die's fairness?

| Face | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Count | 11 | 7 | 9 | 15 | 12 | 6 |

Run the appropriate Chi-Square Test at the $\alpha$ = 0.05 level.

35. Many popular businesses are franchises – think McDonald's. The owner of a local franchise benefits from the brand recognition, national advertising, and detailed guidelines provided by the franchise chain. In return, he or she pays fees to the franchise firm and agrees to follow its policies. The relationship between the local owner and the franchise firm is spelled out in a detailed contract.

One clause that the contract may or may not contain is the entrepreneur's right to an exclusive territory. This means that the new outlet will be the only representative of the same chain in a specified territory and will not have to compete with other outlets of the same chain. How does the presence of an exclusive-territory clause in the contract relate to the survival of the business?

A study designed to address this question collected data from a random sample of 170 new franchise firms. Two categorical variables were measured for each franchisor. First, the franchisor was classified as successful or not based on whether or not it was still offering franchises as of a certain date. Second, the contract each franchisor offered to franchisees was classified according to whether or not there was an exclusive-territory clause. Here are the count data, arranged in a two-way table.

| | **Exclusive Territory** | | |
|---|---|---|---|
| **Success** | **Yes** | **No** | **Total** |
| Yes | 108 | 15 | 123 |
| No | 34 | 13 | 47 |
| **Total** | **142** | **28** | **170** |
| | | | |

Do these data provide convincing evidence of an association between an exclusive territory clause and business survival? Carry out an appropriate test at the $\alpha$ = 0.01 level.

36. A paper in the *American Journal of Sports Medicine* compared collegiate soccer player, athletes in sports other than soccer, and a group of students who were not involved in collegiate sports with respect to history of head injuries. The following table is the result of classifying each student in independently selected samples of 91 soccer players, 96 non-soccer athletes and 53 non-athletes according to the number of previous concussions the student reported on a medical history questionnaire.

Number of Concussions

| | 0 concussions | 1 concussion | 2 concussions | 3 or more concussions | Total |
|---|---|---|---|---|---|
| Soccer Players | 45 | 25 | 11 | 10 | 91 |
| Non-Soccer Athletes | 68 | 15 | 8 | 5 | 96 |
| Non-Athletes | 45 | 5 | 3 | 0 | 53 |
| Total | 158 | 45 | 22 | 15 | 240 |

Does there appear to be a difference in the number of concussions for the three different groups of students? At the 0.05 significance level, perform the appropriate test.

37. Each person in a random sample of 42 students at a large university was asked how much they paid per month for housing and how far from campus they lived (in miles). The resulting data was used to produce the following regression output. You can assume that any conditions needed for inference are met.

The regression equation is
Housing cost = 452 – 9.25(Distance)

| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | 452.1 | 178.1 | 2.54 | 0.039 |
| Distance | -9.2471 | 0.2145 | -43.11 | 0.000 |

s = 221.098 R-Sq = 99.6% R-Sq(adj) = 99.6%

a. Is there convincing evidence of a linear relationship between the distance from campus and cost of housing? (Assume all conditions are met. State hypotheses, t, p-value and conclusion only)

b. Construct a 95% confidence interval to estimate the slope of the regression line.
(Assume all conditions are met. State interval and conclusion only)

Answers

| | | | | | |
|---|---|---|---|---|---|
| 1. d | 2. e | 3. d | 4. d | 5. c | 6. e |
| 7. c | 8. e | 9. b | 10. d | 11. b | 12. e |
| 13. a | 14. c | 15. a | 16. b | 17. b | 18. c |
| 19. a | 20. e | 21. d | 22. d | 23. b | 24. c |
| 25. c | 26. d | 27. b | 28. a | 29. b | 30. c |
| 31. d | 32. c | 33. a | | | |

34. **State:**
Ho: All faces of the die have p = 1/6 or the die is fair.
Ha: At least one face of the die $p \neq 1/6$ or the die is not fair.
$\alpha = 0.05$

**Plan:** Run a chi-square goodness of fit test.
Random: Each roll of a die is random.
10%: 60 < .1(all die rolls)
Large Counts: The expected count for each face of the die is 10, which is at least 5.

**Do:** $\chi^2 = 5.6$, df = 5. P-value is greater than .25, with technology P = .347

**Conclude:** Since P > .05, we fail to reject the null. We do not have enough evidence to say the die is not fair (or not enough evidence to say that at least one $p \neq 1/6$)

35. **State:**
Ho: There is no association between territory type and success or not in the population of franchises
Ha: There is an association between territory type and success or not in the population of franchises
$\alpha = 0.01$

**Plan:** Run a chi-square test for association
Random: This is a random sample of franchises.
10%: 170 is less than 10% of all franchises in United States.
Large Counts: The lowest expected count is 7.74, which is at least 5.

**Do:** $\chi^2 = 5.911$, df = 1 and P-value is .01 < P < .02 or with technology P = .0150

**Conclude:** Since P > .01, we fail to reject the null hypothesis. We do not have enough evidence to conclude that there is an association between whether franchises have an exclusive territory on not and whether they are successful or not.

36. **State:**
Ho: The distribution of concussions is the same for the three different groups of students.
Ha: The distribution of concussions is different for the three different groups of students.
$\alpha = 0.05$

**Plan:** Run a chi-square test for homogeneity
Random: Three independent randomly selected samples.
10%: 91< .1 (all soccer players), 96 < .1 (all non-soccer athletes), 53 < .1(all non-athletes)
Large Counts: All expected counts are NOT at least 5. Proceed with caution.

**Do:** $\chi^2 = 22.029$, df = 6, and P-value is .0025 < P < .001 or with technology P = .0012

**Conclude:** Since P-value < α (.0012 < 0.05) we reject Ho. We have convincing evidence to say that the distribution of concussions is different for the three different groups of students.

Follow Up Analysis: (You don't have to do this unless asked) The four cells that contributed most to $\chi^2$ were Soccer, no concussions, Soccer, 1 concussion, Non-athlete, 3 or more concussions and Soccer, 3 or more concussions. Soccer, no concussions and Non-athlete, 3 or more concussions were less than expected. Soccer, 1 concussion and Soccer 3 or more concussions, were both more than expected.

37a. State: Ho: β = 0; Ha: $\beta \neq 0$ β is population slope of regression line between distance from campus and monthly cost of housing

Plan: Conditions are met, stated in problem

Do: t = -43.11, P-value = 0.000

Conclude: Since P is less than any reasonable significance level, we reject the null hypothesis. There is sufficient evidence that there is a linear relationship between distance from campus and monthly cost of housing.

b.

$$-9.2471 \pm 2.021(0.2145)$$
$$-9.2471 \pm 0.4335$$
$$(-9.681, -8.8136)$$

We are 95% confident that the interval -9.681 to -8.8136 captures the true slope of the regression line between cost of housing and how far from campus a student lived (miles).