

MULTIPLE CHOICE Write the letter corresponding to the best answer in the given blank.

- _____ 1. A survey is conducted on a group of 90 students to see if there is an association between gender (either male or female) and favorite flavor of ice cream (there were 3 choices in the survey: chocolate, vanilla, or strawberry). If χ^2 is 8.26, which of the following intervals contains the p-value?
- A) $p < 0.01$ B) $0.01 < p < 0.02$ C) $0.02 < p < 0.025$ D) $0.025 < p < 0.05$ E) $p > 0.05$
- _____ 2. For the survey described in the previous problem, the totals for each category of response are summarized in the following table:

	Chocolate	Vanilla	Strawberry	Totals
Male				50
Female				40
Totals	37	22	41	90

The expected number of females who prefer vanilla flavor ice cream is:

- A) $\frac{(50)(22)}{90}$ B) $\frac{(40)(22)}{90}$ C) $\frac{90}{(40)(22)}$ D) $\frac{40}{(22)(90)}$ E) $\frac{22}{(40)(90)}$
- _____ 3. For the scenario in the previous two problems: To see if there is statistical evidence that gender is related to preference in ice cream flavor, which of the following is the correct null hypothesis?
- A) the distribution of proportions for ice cream preference is the same for males and females.
 B) the distribution of proportions for ice cream preference is not the same for males and females.
 C) there is an association between gender and preference in ice cream flavor.
 D) there is no association between gender and preference in ice cream flavor.
 E) the distribution of proportions of ice cream preference matches the advertised relative distribution.
- _____ 4. Pig dice are little plastic pig-shaped figures that can be tossed and rolled just like dice! One of the six positions that pig dice may land on is called “razorback”, in which the pig lands on its back with all 4 legs in the air. Based on empirical data, the probability that a pig will land on this position is about 20%.

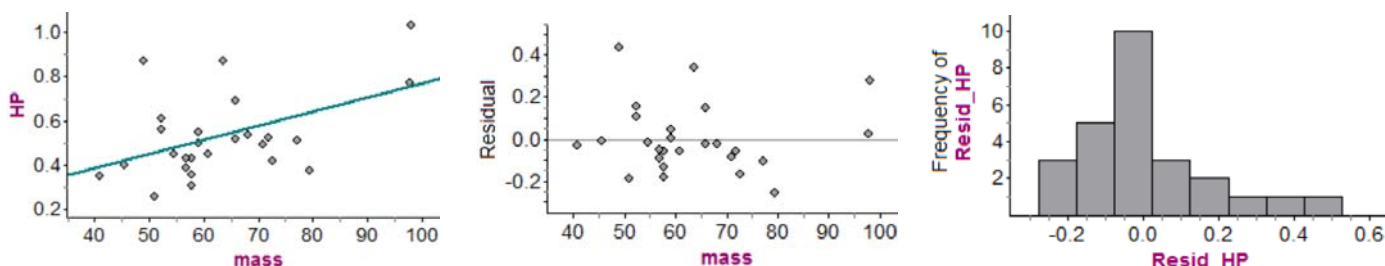
In a game of chance, you get to roll a pig dice a certain number of times, and if more than 50% of your rolls land “razorback”, then you win a large prize! From the following three choices, you get the pick how many times to roll the pig dice:

$n=10$, $n=20$, and $n=100$,

Which value of n should the player choose in order to maximize the probability of winning a prize?

- A) $n=10$ only B) $n=20$ only C) $n=100$ only D) all three probabilities are equal
- _____ 5. The distribution of SAT Math scores of students taking Statistics at a very large university is *skewed right* with a mean of 625 and a standard deviation of 44.5. If random samples of 100 students are repeatedly taken, which give the sampling model, which statement best describes this sampling model of sample mean scores?
- A) Approximately normal with a mean of 625 and standard deviation of 44.5.
 B) Approximately normal with a mean of 625 and standard deviation of 4.45.
 C) Skewed to the right with a mean of 625 and standard deviation of 44.5.
 D) Skewed to the right with a mean of 625 and standard deviation of 4.45.
 E) No conclusion can be drawn since the population is not normally distributed.

During a “horsepower lab” activity, a group of twenty six AP Statistics ran up a flight of stairs and measured their estimated horsepower, based on their mass (and also the time required to make it up the stairs). The information from the regression analysis is shown below. The response variable in this analysis is “horsepower”.



Predictor	Coefficient	S.E. of Coeff	T-ratio	Probability
Mass (kg)	0.0064	0.0024	2.658	0.0138
Constant	0.1224	0.1551	0.789	0.4377
S = 0.167		R-sq = 22.7%		R-sq adj = 19.5%

6. Based on the computer output, what is the least squares regression equation in context?
- predicted horsepower = $0.1551 + 0.0024(\text{mass})$
 - predicted horsepower = $0.0064 + 0.1224(\text{mass})$
 - predicted horsepower = $0.0064 + 0.0024(\text{mass})$
 - predicted horsepower = $0.1224 + 0.0064(\text{mass})$
 - predicted horsepower = $0.1224 + 0.1551(\text{mass})$
7. In the computer output, the coefficient for “Constant” is 0.1224. Which of the following is a correct interpretation for this value?
- If a student has a mass of zero kilograms, then they are predicted to generate 0.1224 horsepower.
 - If a student makes it up the stairs in zero seconds, they are predicted to generate 0.1224 horsepower.
 - If a student has a mass of 0.1551 kilograms, the model predicts they will generate 0.1224 horsepower.
 - A student with 0.1224 kilograms in mass is predicted to generate zero horsepower.
8. At the 5% significance level, which of the following is a correct conclusion about whether there is sufficient evidence of a linear association between mass and horsepower? Assume that the conditions for inference have been checked and verified.
- $t = 2.658$, $p = 0.0138$. Since $p < \alpha$, there is evidence of a linear association between mass and horsepower.
 - $t = 2.658$, $p = 0.0138$. Since $p < \alpha$, we lack evidence of a linear association between mass and horsepower.
 - $t = 0.789$, $p = 0.4377$. Since $p > \alpha$, there is evidence of a linear association between mass and horsepower.
 - $t = 0.789$, $p = 0.4377$. Since $p > \alpha$, we lack evidence of a linear association between mass and horsepower.
9. What is the proportion of the variation in horsepower that can be explained by the person’s mass?
- 0.167
 - 0.227
 - 0.195
 - 0.476
 - 0.442
10. Which of the following is a correct interpretation for “s = 0.167”?
- 0.167 is the typical distance from the mean horsepower for these students.
 - 0.167 is the probability of correctly rejecting a false null hypothesis.
 - 0.167 is the probability of observing a mean horsepower at least as extreme as the one observed given that the null hypothesis is true.
 - 0.167 is the typical difference between the observed and predicted horsepower for these students.
 - 0.167 is the typical difference between the observed and actual horsepower for these students.

An administrator at a large university is interested in determining whether the residential status of a student is associated with level of participation in extracurricular activities. Residential status is categorized as on campus for students living in university housing and off campus otherwise. A simple random sample of 100 students in the university was taken, and each student was asked the following two questions.

- Are you an on campus student or an off campus student?
- In how many extracurricular activities do you participate?

The responses of the 100 students are summarized in the frequency table shown.

Level of Participation in Extracurricular Activities	Residential Status		Total
	On campus	Off campus	
No activities	9 (12.87)	30 (26.13)	39
One activity	17 (13.86)	25 (28.14)	42
Two or more activities	7 (6)	12 (13)	19
Total	33	67	100

(Expected counts are shown in parentheses)

The administrator wishes to perform a test to see if there is statistical evidence of an association between residential status and level of participation in extracurricular activities among the students at the university.

- ____ 11. Are the conditions for the appropriate chi-square inference procedure met for these data?
- Yes. We have a random sample of 100 students from this university, and all of the observed counts are at least five (5).
 - Yes. We have a random sample of 100 students from this university, and all of the expected counts are at least five (5).
 - Yes. We have a random sample of 100 students from this university, and the sample size is greater than 30.
 - No, because we would required a sample of more than 10% of the students attending this university.
 - No, because this is a survey, and the students were not randomly assigned to to live on-campus or off-campus.
- ____ 12. For this test: Chi-Sq = 2.925, P-Value = 0.23. At $\alpha = 0.05$, which type of error might have been made?
- Since they would reject H_0 , this might be a type I error.
 - Since they would reject H_0 , this might be a type II error.
 - Since they would fail to reject H_0 , this might be a type I error.
 - Since they would fail to reject H_0 , this might be a type II error.
 - Depending on the location of the university, this could be either a type I or a type II error.
- ____ 13. Which of the following represents a type II error in the context of this problem?
- The administrator would incorrectly conclude that there is an association between residential status and participation in extracurricular activities.
 - The administrator would incorrectly conclude that there is no association between residential status and participation in extracurricular activities.
 - The administrator would correctly conclude that there is an association between residential status and participation in extracurricular activities.
 - The administrator would correctly conclude that there is no association between association between residential status and participation in extracurricular activities.

14. A taxicab company in a large city charges passengers a flat fee to enter a cab plus an additional fee per mile. There is also a charge for time spent stopped in traffic. The company wants to develop a new method for determining fares based on mileage and a flat fee only, not on time spent stopped in traffic. A random sample of 10 recent cab fares was selected, and the distance, in miles, and the fare, in dollars, were recorded. A regression model was fit to the data, and the output, scatterplot, and residual plot are given below.

Variable	Coefficient	Std. Err.	T-stat	Prob.
Intercept	4.296	0.298	14.400	<0.0001
Mileage	1.229	0.166	7.418	<0.0001

- a) Construct and interpret a 95% confidence interval for the slope of the least-squares regression line. Assume that the conditions for inference are met.
- b) A recent newspaper article claims that the cost of riding a taxi in this large city comes out to about \$1.70 per mile travelled. Based on your interval in part (a), is this claim of \$1.70 per mile a reasonable value? Explain.

15. Young magic students at a large school of magic are sorted by a “sorting hat” into one of four houses, which serve as the learning and living communities for its students. A random sample of 86 magicians was taken from all students entering the school in the year 1971, and a separate random sample of 72 magicians was taken from all students entering the school in year 1991. For the students in these random samples, the distribution of the number of students observed in the four houses in each year is shown in the following table:

	House			
	G	H	R	S
1971	22	13	30	21
1991	26	21	16	9

- a) Among students entering the school, is there evidence of a difference in the relative distribution of students among the four houses between 1971 and 1991? Name the type of test that should be conducted, along with the appropriate degrees of freedom.
- b) State the appropriate hypotheses for the test named in part (a), but do NOT perform the test.

16. The table below shows the political party registration by gender of 500 randomly selected registered voters in Franklin Township.

PARTY REGISTRATION—FRANKLIN TOWNSHIP

	Party W	Party X	Party Y
Female	60	120	120
Male	28	124	48

Among registered voters in Franklin Township, is there evidence of an association between gender and party registration? Give appropriate statistical evidence to support your conclusion.

17. The department of parks and recreation of a certain city conducts summer programs for residents of its five districts. The summer programs include operating and maintaining community swimming pools in each of the districts as well as offering sports and recreational programs for school-age children, young adults, and older adults.

The proportion of total households in each of the districts was determined for this study, and the proportions are displayed in the middle column in the table below. For instance, 32% of the city's households are located in district A, 12% of the city's households are located in district B, and so on.

A group of statisticians collected data from a simple random sample this past summer, and the number of households that participated in summer programs from each district is shown in the right-hand column below.

District	Proportion of Total Households	Number of Households that Participated in Summer Programs
A	0.32	100
B	0.12	35
C	0.10	40
D	0.27	45
E	0.19	51
Total	1.000	271

- a) The department of parks and recreation expect the number of households that participate in each district to be proportional to the percentage of the city's households that are located in each district. Are the data consistent with this expectation? Conduct an appropriate statistical test to support your conclusion.
Assume the conditions for inference are met.

If you need more room for your work in part (a), use the space below.

- b) Relative to the proportion of total households, in which district is participation in summer programs the lowest? Explain.