AP REVIEW PACKET #3 INSTRUCTIONS – Due by Friday 3/22 7:50 – 40 points

1) Read Topic 4: Exploring Bivariate Data and Topic 5: Exploring Categorical Data, pp 105-138 & pp 145-160, in the Barron's book. It is highly suggested that you go over the questions at the end of the each topic as some will appear on your quiz next week.

2) Answer the multiple choice questions in this packet, these will be graded for accuracy. Show all work and then copy your answers into the blanks below.

16. _____ 17. _____ 18. _____ 19. _____ 20. _____

21. _____ 22. _____ 23. _____ 24. _____ 25. _____

3) Answer the Free-Response Questions from the free-response packet.

Optional Bonus Activity: Go to http://learner.org/resources/series65.html. Watch programs 4& 5 & complete the worksheet.

Guided Reading Questions for Packet #3:
Topic 4:

1) When analyzing the overall pattern in a scatterplot, it is also important to note _____ and _____.

2) _____ does not imply _____!

3) Correlation does not distinguish between which variable is _____ _____ and which is _____ _____.

4) It can be shown that r^2, called the _____ of _____, is the ratio of _____

_____.

5) It is reasonable, intuitive, and correct that the best-fitting line will pass through _____.

6) Be sure to _____ _____ and _____ _____ _____.

7) The mean of the residuals is always _____.

8) Influential scores are scores whose _____ would _____ _____ the regression line.

9) Choose 4 important points to remember from the summary:

1.

2.

3.

4.

Topic 5:

1) The relationship between two categorical variables may be displayed in a _____-_____ _____table.

2) Marginal distributions do not describe or measure the _____ _____ the two categorical variables.

3) Conditional relative frequencies can be displayed either with _____ of _____or by a _____ _____ _____.

4) Copy the Summary points below:

-
-
-
-
-

Multiple Choice Questions:

16. A baseball coach wants to compare the number of hits by two groups of batters each using a different type of bat. Which type of graphical display would NOT be appropriate?

    (A) Parallel boxplots

    (B) Dotplots drawn on the same scale

    (C) Back-to-back stemplots

    (D) Histograms drawn on the same scale

    (E) Scatterplot

17. A distribution of scores has a mean of 60 and a standard deviation of 18. If each score is doubled, and then 5 is subtracted from that result, what will be the mean and standard deviation of the new scores?

    (A) mean = 115, standard deviation = 31

    (B) mean = 115, standard deviation = 36

    (C) mean = 120, standard deviation = 6

    (D) mean = 120, standard deviation = 31

    (E) mean = 120, standard deviation = 36

18. A hypothesis test is conducted with respect to the mean weight (in ounces) of potato chip bags from a certain manufacturer. The test's hypotheses are H_0: $\mu = 0.8$ and H_a: $\mu \neq 0.8$. Which confidence interval below would support the conclusion that there is insufficient evidence to reject the null hypothesis at the $\alpha = 0.03$ level of significance?

(A) The 97% confidence interval for the mean weight in ounces of potato chips is (0.765, 0.823).

(B) The 94% confidence interval for the mean weight in ounces of potato chips is (0.765, 0.823).

(C) The 97% confidence interval for the mean weight in ounces of potato chips is (0.725, 0.783).

(D) The 94% confidence interval for the mean weight in ounces of potato chips is (0.725, 0.783).

(E) We cannot conclude anything with a confidence interval unless we have the actual data set to construct the interval.

19. A new medication has been developed to cure a certain disease. The disease progresses in three stages, stages I, II, and III, each progressively worse than the one before it. Ninety volunteers are gathered to test the new medication, 30 in each of the three stages of the disease. The medication will be administered to subjects daily in one of three dosages: 100 mg for each subject in stage I of the disease, 200 mg to each subject in stage II, and 400 mg to each subject in stage III. After 8 weeks, the proportion of subjects cured of the disease will be recorded. Why is this NOT a good experimental design?

I. Because experiments of this type should only use one dosage level of medication.

II. Because disease stage is potentially confounded with dosage level.

III. Because the experiment lacks a control group.

(A) I only

(B) II only

(C) I and II only

(D) II and III only

(E) I, II, and III

20. Sparkles, a small tart candy, comes in four colors. Their website states that there are twice as many red as each of the other three colors. A random sample of 80 candies was taken and the colors were distributed as follows:

Red	Yellow	Green	Blue
37	15	18	10

The χ^2 test statistic for the goodness of fit test is

(A) $\chi^2 < 1$

(B) $1 \leq \chi^2 < 4$

(C) $4 \leq \chi^2 < 15$

(D) $15 \leq \chi^2 < 75$

(E) $75 < \chi^2$

21. A certain variety of table grapes has fruit diameters that are distributed normally with mean 13 mm and standard deviation 2 mm. Approximately what proportion of have diameters between 12 mm and 16 mm?
(A) 0.134 (B) 0.378 (C) 0.500 (D) 0.625 (E) 0.683

22. A tire manufacturer is testing a new tread design for its light-truck tires. The previous design had a mean tread life of 47,500 miles. Tires with the new design are manufactured and tested on a variety of light trucks. Which of the following is the correct pair of hypotheses to test the assertion that the new tread design has a longer life than the old design?

(A) $H_0: \mu < 47{,}500$, $H_a: \mu = 47{,}500$
(B) $H_0: \mu = 47{,}500$, $H_a: \mu \neq 47{,}500$
(C) $H_0: \mu = 47{,}500$, $H_a: \mu < 47{,}500$
(D) $H_0: \mu = 47{,}500$, $H_a: \mu > 47{,}500$
(E) $H_0: \mu > 47{,}500$, $H_a: \mu \leq 47{,}500$

23. A study was done to explore a link between a particular medication prescribed to pregnant women and a certain medical condition in newborns. Records of 952 recent newborns and their mothers were examined. The table shows the results of the study. Which of the following best describes the association between mothers-to be taking the medication and the presence of the condition in newborns?

		Mother took medication?		
		Yes	No	Total
Newborn has condition ?	Yes	21	245	266
	No	57	629	686
	Total	78	874	952

(A) There appears to be no association since the condition was present in newborns of mothers that either took or did not take the medication.

(B) There appears to be no association since the condition was present in about the same proportion of newborns of mothers that either took or did not take the medication.

(C) There appears to be no association because more newborns of mothers who did not take the medication had the condition than newborns of mothers who did take the medication.

(D) There appears to be no association because more newborns did not have the condition than those who did.

(E) There appears to be an association because the condition was present in newborns of mothers who took the drug.

24. The Sunday edition of the newspaper has 585,320 readers. Sixty-three percent of the readers are men. It is known that about 12% of the women and 23% of the men that read this newspaper will read the book review section. If a random sample of 200 readers is taken, what is the expected number of people that will read the book review section?
(A) 23 (B) 24 (C) 38 (D) 46 (E) 70

25. A restaurant recognizes customers who go there on their birthdays with a free piece of cake and the singing of a song. On one particular day, a sever noticed that 10 out of the 187 customers at the restaurant celebrated birthdays that day. Having taken statistics, the server did a significance test to determine if the true proportion of people with birthdays on that day is significantly different from what is expected. In a test of $H_0: p = \frac{1}{365}$ versus $H_A: p \neq \frac{1}{365}$, the test statistic was 13.27 and the P-value was 3.43×10^{-40}. Which of the following is true?

(A) There is a significantly different proportion of people born on that day than one would expect.
(B) The test is invalid; a sample size of 187 is far too small to do inference for a proportion.
(C) The test is inappropriate; a confidence interval should have been done instead.
(D) The test was incorrectly performed. The alternative hypothesis should have been $H_A: p > \frac{1}{365}$.
(E) Any inference is invalid. The sample of restaurant customers is not representative of the population.

Free Response Questions:

1. Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.

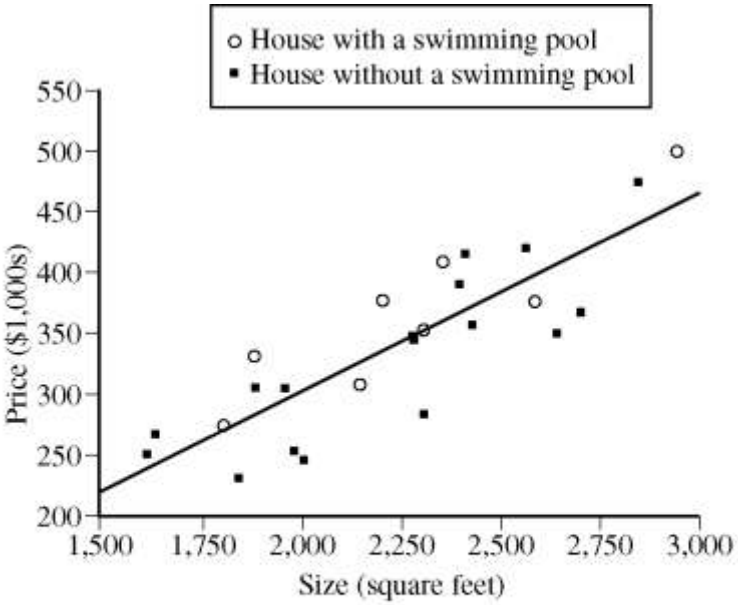
(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.

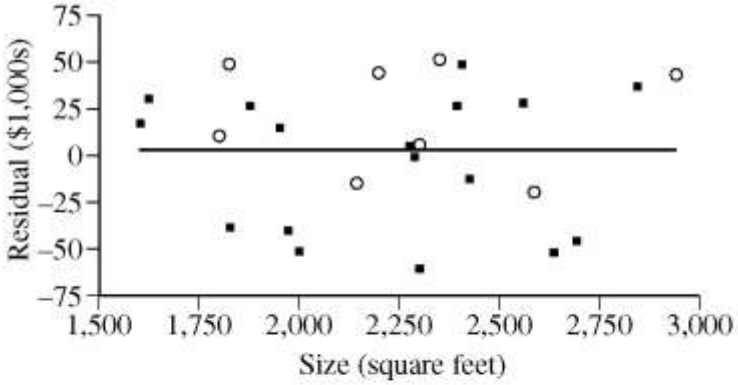
(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?

(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.

2. A real estate agent is interested in developing a model to estimate the prices of houses in a particular part of a large city. She takes a random sample of 25 recent sales and, for each house, records the price (in thousands of dollars), the size of the house (in square feet), and whether or not the house has a swimming pool. This information, along with regression output for a linear model using size to predict price, is shown below and on the next page.

Price ($1,000s)	Size (square feet)	Pool	Residual ($1,000s)
274	1,799	yes	6
330	1,875	yes	49
307	2,145	yes	–18
376	2,200	yes	42
352	2,300	yes	1
409	2,350	yes	50
375	2,589	yes	–23
498	2,943	yes	42
248	1,600	no	13
265	1,623	no	26
228	1,829	no	–45
303	1,875	no	22
303	1,950	no	10
251	1,975	no	–46
244	2,000	no	–57
347	2,274	no	1
345	2,279	no	–2
282	2,300	no	–69
389	2,392	no	23
413	2,410	no	44
353	2,428	no	–19
419	2,560	no	26
348	2,639	no	–58
365	2,701	no	–52
474	2,849	no	33





Linear Fit

Price = –28.144 + 0.165 Size

Summary of Fit

RSquare 0.722

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	–28.144	48.259	–0.58	0.5654
Size	0.165	0.0213	7.72	<.0001

(a) Interpret the slope of the least squares regression study.

(b) The second house in the table has a residual of Interpret this residual value in the context of the study.

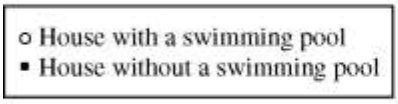
The real estate agent is interested in investigating the effect of having a swimming pool on the price of a house.

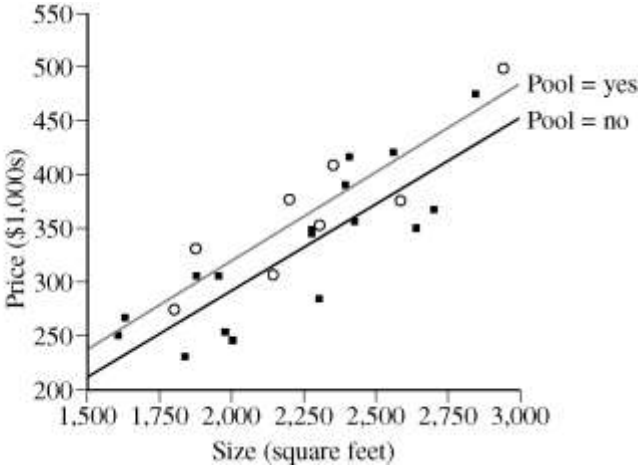
(c) Use the residuals from all 25 houses to estimate how much greater the price for a house with a swimming pool would be, on average, than the price for a house of the same size without a swimming pool.

To further investigate the effect of having a swimming pool on the price of a house, the real estate agent creates two regression models, one for houses with a swimming pool and one for houses without a swimming pool. Regression output for these two models is shown below.

Linear Fit (Pool = yes)
Price = −11.602 + 0.166 size

Linear Fit (Pool = no)
Price = −27.382 + 0.160 size





(d) The conditions for inference have been checked and verified, and a 95 percent confidence interval for the true difference in the two slopes is (−0.099, 0.110). Based on this interval, is there a significant difference in the two slopes? Explain your answer.

(e) Use the regression model for houses with a swimming pool and the regression model for houses without a swimming pool to estimate how much greater the price for a house with a swimming pool would be than the price for a house of the same size without a swimming pool. How does this estimate compare with your result from part (c) ?

3. The summary statistics for the number of inches of rainfall in Los Angeles for 117 years, beginning in 1877 are shown below.

N	MEAN	MEDIAN	TRMEAN	STDEV	SE MEAN
117	14.941	13.070	14.416	6.747	0.624

MIN	MAX	Q1	Q3
4.850	38.180	9.680	19.250

(a) Describe a procedure that uses these summary statistics to determine whether there are outliers.

(b) Are there outliers in these data? _____

Justify your answer based on the procedure that you described in part (a).

(c) The news media reported that in a particular year, there were only 10 inches of rainfall. Use the information provided to comment on this reported statement.

4) The table below shows the political party registration by gender of all 500 registered voters in Franklin Township.

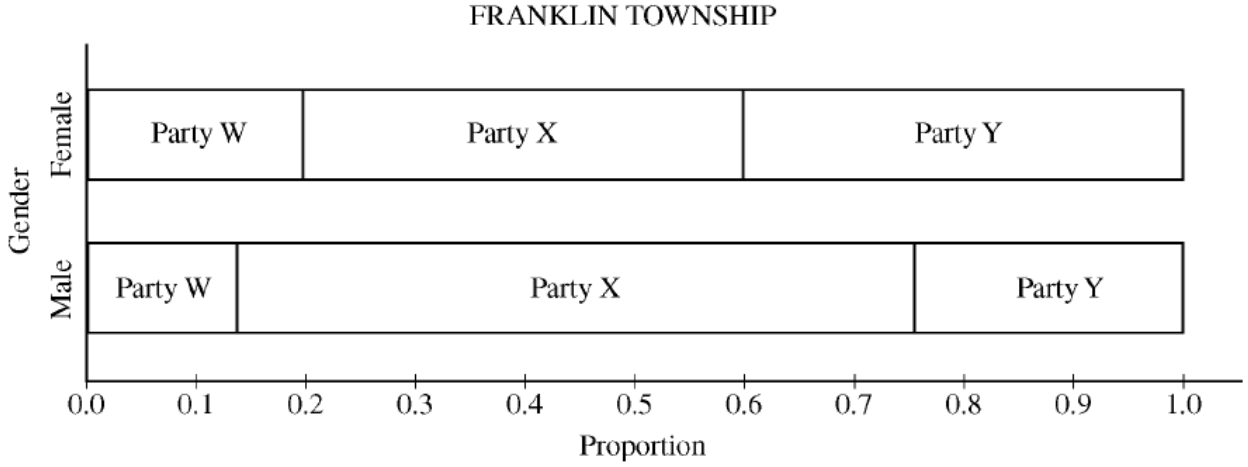
PARTY REGISTRATION–FRANKLIN TOWNSHIP

	Party W	Party X	Party Y	Total
Female	60	120	120	300
Male	28	124	48	200
Total	88	244	168	500

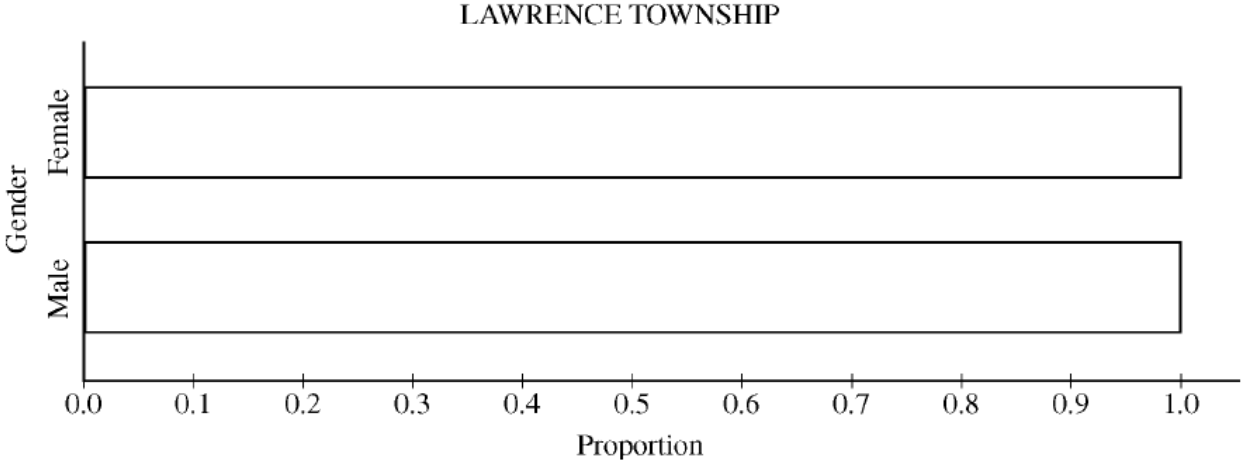
(a) Given that a randomly selected registered voter is a male, what is the probability that he is registered for Party Y?

(b) Among the registered voters of Franklin Township, are the events “is a male” and “is registered for Party Y” independent? Justify your answer based on probabilities calculated from the table above.

(c) One way to display the data in the table is to use a segmented bar graph. The following segmented bar graph, constructed from the data in the party registration–Franklin Township table, shows party-registration distributions for males and females in Franklin Township.



In Lawrence Township, the proportions of all registered voters for Parties W, X, and Y are the same as for Franklin Township, and party registration is independent of gender. Complete the graph below to show the distributions of party registration by gender in Lawrence Township.



5) A simple random sample of adults living in a suburb of a large city was selected. The age and annual income of each adult in the sample were recorded. The resulting data are summarized in the table below.

	Annual Income			
Age Category	$25,000-$35,000	$35,001-$50,000	Over $50,000	Total
21-30	8	15	27	50
31-45	22	32	35	89
46-60	12	14	27	53
Over 60	5	3	7	15
Total	47	64	96	207

(a) What is the probability that a person chosen at random from those in this sample will be in the 31-45 age category.

(b) What is the probability that a person chosen at random from those in this sample whose incomes are over $50,000 will be in the 31-45 age category?

(c) Based on your answers to parts (a) and (b), is annual income independent of age category for those in this sample? Explain.

Name _____ Date _____ Per____

6. A rural county hospital offers several health services. The hospital administrators conducted a poll to determine whether the residents' satisfaction with the available services depends on their gender. A random sample of 1,000 adult county residents was selected. The gender of each respondent was recorded and each was asked whether he or she was satisfied with the services offered by the hospital. The resulting data are:

	Male	Female	Total
Satisfied	384	416	800
Not Satisfied	80	120	200
Total	464	536	1,000

(a) Using a significance level of 0.05, conduct an appropriate test to determine if, for adult residents of this county, there is an association between gender and whether or not they were satisfied with services offered by the hospital.

(b) Is $\frac{800}{1,000}$ a reasonable estimate for the proportion of all adult county residents who are satisfied with the services offered by this hospital? Explain why or why not.

Name _____ Date _____ Per_____

Bonus Questions: **NORMAL DISTRIBUTIONS**

1. What type of curves are used to illustrate the U.S. (population in 1930 and 2075)? _____
2. What is the area under each of these curves? _____
3. What point divides a density curve into two equal areas? _____
4. What measure is the point at which the curve would balance? _____
5. In which direction is the mean pulled in a skewed distribution? _____
6. If the mean of a normal curved is changed, what happens to the curve? _____
7. If the standard deviation of normal curve is changed, what happens to the curve? _____
8. What example in the video illustrates standard deviation changing over time? _____
9. Using the standard (z) curves, who has the highest batting average of all time? _____

NORMAL CALCULATIONS

1. What is the area under a density curve?
2. What is the mean of a standard normal distribution?
3. What example is used to illustrate industry's use of normal calculations?
4. What example is used to illustrate the medical community's use of z-scores?
5. What are some of the uses of the army anthropological study of the typical soldier?
6. What z-score has a 95% of the population below it? _____
7. What type of plot indicates if a distribution is normal? _____
8. If data are normal, what pattern does the plot named in #7 have? _____