**AP Stats** Name: Key

**Review for the AP Exam - Chapters 14 & 15** Hour: ______

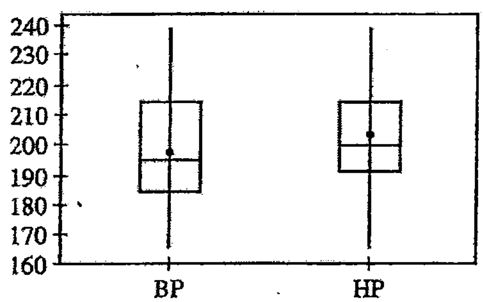
| Student | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Height | 65 | 72 | 64 | 68 | 65 | 70 | 61 | 73 | 69 | 70 |
| Arm Span | 67 | 71 | 60 | 69 | 60 | 65 | 58 | 74 | 70 | 67 |

1. The table above shows the height, in inches, and the arm span, in inches, for 10 randomly selected high school students. Which of the following significance tests should be used to determine whether a linear relationship exists between height and arm span, provided the assumptions of the tests are met?

(A) Two-sample z-test
(B) Two-sample t-test
(C) Chi-square test of independence
(D) Chi-square goodness-of-fit test
(E) t-test for slope of regression line

2. In a study of the weights of college athletes, player weights for a random sample of baseball players (BP) and for an independent random sample of hockey players (HP) were compared. The computer output shown below gives the results of a test of $H_0: \mu_{BP} = \mu_{HP}$ versus $H_a: \mu_{BP} < \mu_{HP}$.

BOXPLOTS OF BP AND HP
(means are indicated by solid circles)




Two Sample T-Test and Confidence Interval

Two sample T for BP vs HP

| | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| BP | 37 | 197.8 | 19.4 | 3.2 |
| HP | 34 | 204.2 | 17.7 | 3.0 |

95% CI for $\mu$ BP − $\mu$ HP: (−15.1, 2.4)
T-Test $\mu$ BP = $\mu$ HP (vs <): T = −1.45 P = 0.076 DF = 69

Which of the following is the best conclusion that can be drawn from the analysis?

(A) The mean weight of baseball players is not significantly lower than the mean weight of hockey players at the 0.05 level.
(B) The mean weight of baseball players is not significantly lower than the mean weight of hockey players at the 0.10 level.
(C) The mean weight of baseball players is not significantly higher than the mean weight of hockey players at the 0.10 level.
(D) The mean weight of baseball players is significantly lower than the mean weight of hockey players at the 0.05 level.
(E) The mean weight of baseball players is significantly different from the mean weight of hockey players at the 0.05 level.

3. The manager of a factory wants to compare the mean number of units assembled per employee in a week for two new assembly techniques. Two hundred employees from the factory are randomly selected and each is randomly assigned to one of the two techniques. After teaching 100 employees one technique and 100 employees the other technique, the manager records the number of units each of the employees assembles in one week. Which of the following would be the most appropriate inferential statistical test in this situation?

(A) One-sample *z*-test (B) Two-sample *z*-test (C) Paired *t*-test
(D) Chi-square goodness-of-fit test (E) One-sample *t*-test

(2013 #4)

4. The Behavioral Risk Factor Surveillance System is an ongoing health survey system that tracks health conditions and risk behaviors in the United States. In one of their studies, a random sample of 8,866 adults answered the question "Do you consume five or more servings of fruits and vegetables per day?" The data are summarized by response and by age-group in the frequency table below.

| | | | | Expected yes | Expected no |
|---|---|---|---|---|---|
| 18–34 | 231 | 741 | 972 | 240.204 | 731.796 |
| 35–54 | 669 | 2,242 | 2,911 | 719.378 | 2191.622 |
| 55 or older | 1,291 | 3,692 | 4,983 | 1231.418 | 3751.582 |
| Total | 2,191 | 6,675 | 8,866 | | |

Do the data provide convincing statistical evidence that there is an association between age-group and whether or not a person consumes five or more servings of fruits and vegetables per day for adults in the United States?

① $H_0$: There is no association between age-group and whether or not a person eats $\geq 5$ servings of fruits/veggies per day for all US adults

$H_a$: There is an association

② $\chi^2$ test for association

③ SRS is satisfied. All expected counts, stated in the table above, are $\geq 5$

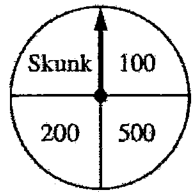
④ $\chi^2 = \frac{(231 - 240.204)^2}{240.204} + \ldots + \frac{(3692 - 3751.582)^2}{3751.582} = 8.983$

$df = (3-1)(2-1) = 2 \quad p = .011$

⑤ Since $p = .011 < \alpha = .05$, we reject $H_0$. There is convincing evidence of an association between age-group and consumption of fruits and vegetables for all US adults.

(2003B #5)

5. Contestants on a game show spin a wheel like the one shown in the figure below. Each of the four outcomes on this wheel is equally likely and outcomes are independent from one spin to the next.




- The contestant spins the wheel.
- If the result is a skunk, no money is won and the contestant's turn is finished.
- If the result is a number, the corresponding amount in dollars is won. The contestant can then stop with those winnings or can choose to spin again, and his or her turn continues.
- If the contestant spins again and the result is a skunk, all of the money earned on that turn is lost and the turn ends.
- The contestant may continue adding to his or her winnings until he or she chooses to stop or until a spin results in a skunk.

(a) What is the probability that the result will be a number on all of the first three spins of the wheel?

$\frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{27}{64} = .422$

(b) Suppose a contestant has earned \$800 on his or her first three spins and chooses to spin the wheel again. What is the expected value of his or her total winnings for the four spins?

$0\left(\frac{1}{4}\right) + 900\left(\frac{1}{4}\right) + 1300\left(\frac{1}{4}\right) + 1000\left(\frac{1}{4}\right) = \$800$

(c) A contestant who lost at this game alleges that the wheel is not fair. In order to check on the fairness of the wheel, the data in the table below were collected for 100 spins of this wheel.

| Result | Skunk | \$100 | \$200 | \$500 |
|---|---|---|---|---|
| Frequency | 33 | 21 | 20 | 26 |

Based on the data on the previous page, can you conclude that the four outcomes on this wheel are not equally likely? Give appropriate statistical evidence to support your answer.

① $H_0$: The 4 outcomes are equally likely.
$H_a$: The outcomes are not equally likely.

② $\chi^2$ GOF test

③ SRS is assumed as random sample of all spins. All expected values are 25 > 5.

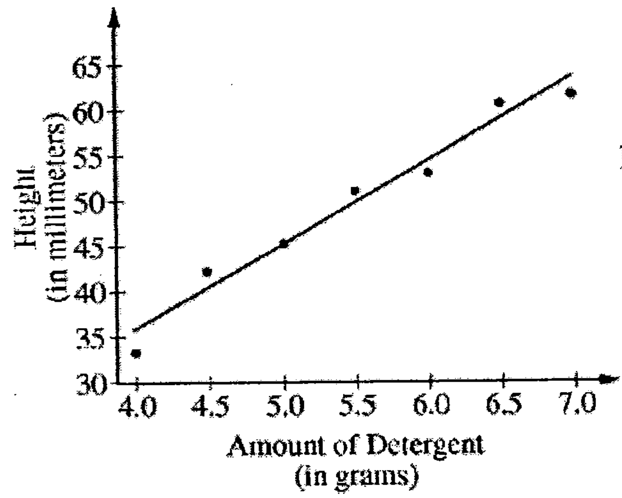
④ $\chi^2 = \frac{(33-25)^2}{25} + \dots \frac{(26-25)^2}{25} = 4.24$ $df = 4-1=3$ $p = .237$

⑤ Since .237 > $\alpha$ = .05 we fail to reject $H_0$. There is not sufficient evidence that the 4 outcomes are not equally likely.

(2006 #2)

6. A manufacturer of dish detergent believes the height of soapsuds in the dishpan depends on the amount of detergent used. A study of the suds' heights for a new dish detergent was conducted. Seven pans of water were prepared. All pans were of the same size and type and contained the same amount of water. The temperature of the water was the same for each pan. An amount of dish detergent was assigned at random to each pan, and that amount of detergent was added to the pan. Then the water in the dishpan was agitated for a set amount of time, and the height of the resulting suds was measured.

A plot of the data and the computer output from fitting a least squares regression line to the data are shown below.




| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | −2.679 | 4.222 | −0.63 | 0.554 |
| Amount | 9.5000 | 0.7553 | 12.58 | 0.000 |

S = 1.99821  R–Sq = 96.9%  R–Sq(adj) = 96.3%

(a) Write the equation of the fitted regression line. Define any variables used in this equation.

$\hat{y} = -2.679 + 9.500x$

$\hat{y}$ = predicted height of soapsuds (mm)

$x$ = amount of detergent (g)

(b) Note that $s = 1.99821$ in the computer output. Interpret this value in the context of this study.

1.99821 is the standard deviation of the residuals. It measures the average amount that the observed soapsuds heights vary from the predicted heights.

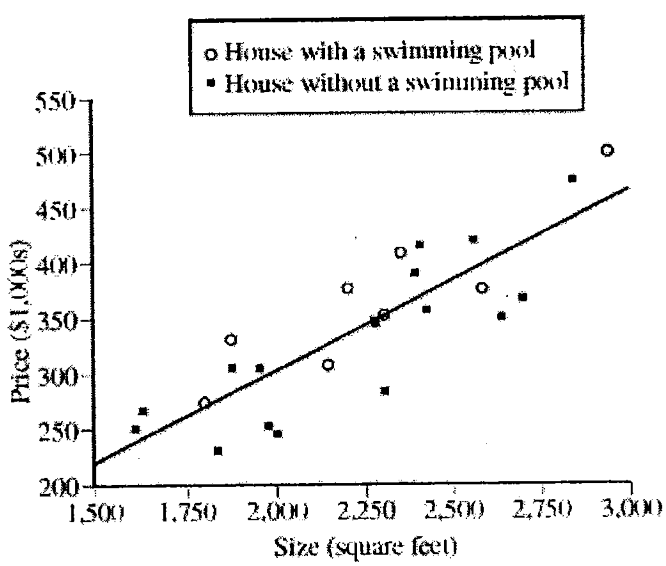
(c) Identify and interpret the standard error of the slope.

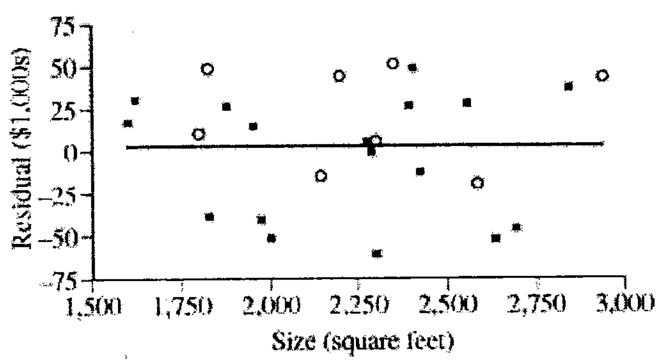
$SE_{b_1} = .7553$ This represents the standard deviation of the sampling distribution of the slope of the LSR line used to predict the height of the soapsuds in relation to the amount of detergent used. Thus estimating the variability of the sample slopes between experiments.

(2010B #6)

7. 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 line in the context of the study.

Slope = .165, meaning for each additional $ft^2$, the predicted cost of the house increases by .165 (\$165) on average.

(b) The second house in the table has a residual of 49. Interpret this residual value in the context of the study.

The positive residual means the house sold for \$49,000 more than the LSR line predicted.

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.

Pool $(\overline{\text{residual}}) = \frac{6+49+\ldots+42}{8} = \$18.625$ thousand

No pool $(\overline{\text{residual}}) = \frac{13+26+\ldots33}{17} = \$-8.824$ thousand

Difference = $18.625 - (-8.824) = 27.449$

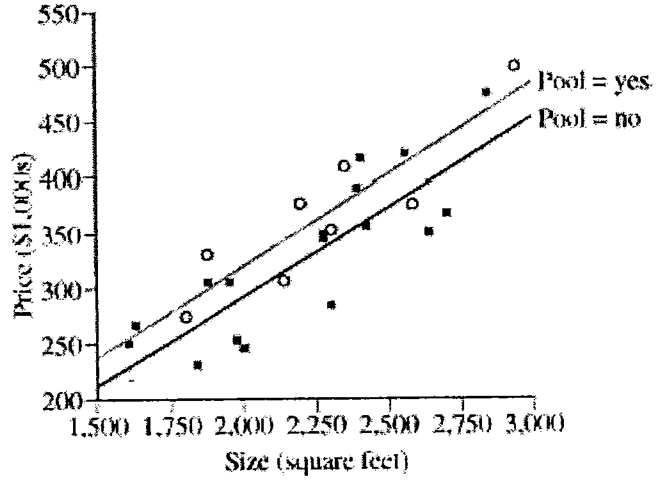
For two houses the same size, the house with the pool will sell for \$27,449 more than the house without a pool, on average.

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

o House with a swimming pool
▪ House without a swimming pool




(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.

Since 0 is in the 95% CI, it is likely the difference in the two slopes is 0, signifying no difference exists.

(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)?

mean size of houses w/ a pool = 2275.125 $ft^2$

" w/o a pool = 2216.706 $ft^2$

$$\text{Avg size} = \frac{2275.125 + 2216.706}{2} = 2245.915\ ft^2$$

$\widehat{\text{price}}$ w/ pool = $-11.602 + .166(2245.915) \approx 361.220$

= \$361,220

$\widehat{\text{price}}$ w/o pool = $-27.382 + .160(2245.915) = 331.964$

= \$331,964

Difference in price = 361,220 − 331,964 = \$29,256

This difference is similar to that found in part (c) = 27,449.