

## Review 1 Answer Key

1. Will be discussed in class

2.

Poverty

- $H_0$ : Poverty percentages are the same for the geographic regions in the United States.  
 $H_A$ : Poverty percentages differ by geographic region in the United States.
- degrees of freedom =  $(4 - 1)(2 - 1) = 3$
- $\frac{100}{4000}(484) = 121$
- Since the  $P$ -value of 0.0029 is so low, there is strong evidence that poverty percentages differ by geographic location in the United States. We will reject the null hypothesis and conclude that poverty is associated with geographic location in the United States. Specifically, the South contains a greater proportion of people below the poverty level than expected.

3.

Student progress

- $H_0$ : There is no association between Math and Reading CTBS scores.  $\beta_1 = 0$   
 $H_A$ : There is an association between Math and Reading CTBS scores.  $\beta_1 \neq 0$
- \* Straight Enough Condition: There is no obvious bend in the scatterplot.  
\* Independence Condition: The residuals show no clear pattern.  
\* Does the Plot Thicken? Condition: The residual plot shows reasonably consistent spread.  
\* Nearly Normal condition: A histogram of the residuals is unimodal and roughly symmetric.
- The  $P$ -value is very small, so we reject the null hypothesis. There is strong evidence of a positive association between CTBS scores Math and Reading.
- A 95% confidence interval for  $\beta_1$  is:  $\beta_1 \pm t_{18}^* \times SE(b_1) = 0.866 \pm 2.101(0.1045)$   
or (0.646, 1.086)
- We are 95% confident that the Reading CTBS score will be higher, on average, between 0.646 and 1.086 points for each additional CTBS point scored on the Math CTBS test.

4.

Car colors

$H_0$ : The distribution of the colors of cars at a NASCAR raceway is the same as the distribution of colors preferred by sports car enthusiasts.

$H_A$ : The distribution of the colors of cars at a NASCAR raceway is different than the distribution of colors preferred by sports car enthusiasts.

- \* Counted data condition: We have counts of car in 6 color categories.
- \* Randomization Condition: We have a random sample of cars at a NASCAR raceway.
- \* Expected cell frequency condition: The expected counts are below. These expected values are all at least 5, so the condition is satisfied.

| Expected Color %      | Red – 19% | Silver – 16.2% | Black – 14.7% | Green – 14.1% | White – 14% | Other – 22% |
|-----------------------|-----------|----------------|---------------|---------------|-------------|-------------|
| Expected Color counts | 47.5      | 40.5           | 36.75         | 35.25         | 35          | 55          |
| Observed Color Counts | 45        | 42             | 34            | 40            | 39          | 50          |

Under these conditions, the sampling distribution of the test statistics is  $\chi^2$  with  $6 - 1 = 5$  degrees of freedom. We will do a chi-square goodness-of-fit test.

$$\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp} = \frac{(45 - 47.5)^2}{47.5} + \dots + \frac{(50 - 55)^2}{55} = 1.945$$

$$P = P(\chi^2 > 1.945) = 0.8567$$

The  $P$ -value of 0.8567 is high, so we fail to reject the null hypothesis. There is no evidence that color preferences of cars at the NASCAR raceways differ from color preferences of sports car enthusiasts.

