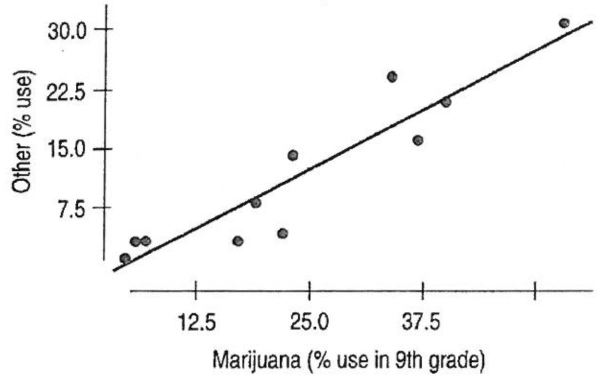
AP Statistics
Unit 08 – HW #3
Inference for Regression

Name Key
Period______

1. *The European School Study Project on Alcohol and other Drugs*, published in 1995, investigated the use of marijuana and other drugs. Data from 11 countries are summarized in the following scatterplot and regression analysis. They show the association between the percentage of a country's ninth graders who report having smoked marijuana and who have used other drugs such as LSD, amphetamines, and cocaine.



Dependent variable is: Other
R-squared = 87.3%
s = 3.853 with 11 − 2 = 9 degrees of freedom

Variable	Coefficient	SE(Coeff)	t-ratio	P-value
Intercept	−3.06780	2.204.	−1.39	0.1974
Marijuana	0.615003	0.0784	7.85	<0.0001

a. Explain in context what the regression says.

predicted % other drug users = −3.06780 + 0.615003(% marijuana users)

the percentage of ninth graders in these countries who have used other drugs is expected to increase 0.615% for each 1% increase in the percentage of ninth graders who have used marijuana.

b. State the hypothesis about the slope (both numerically and in words) that describes how use of marijuana is associated with other drugs.

H_0: there is no (linear) relationship between marijuana use and use of other drugs. $\beta = 0$

H_A: there is a relationship. $\beta \neq 0$

c. Assuming that the assumptions for inference are satisfied, perform a hypothesis test and state your conclusion in context.

test statistic = 7.85
p-value < 0.0001
df = 9

Because our p-value < 0.0001 is less than our significance level $\alpha = 0.05$, we reject the null. There is convincing evidence that there is a relationship (positive, as shown by the graph) between marijuana use and use of other drugs.

d. Explain what R-squared means in context.

87.3% of the variance in % of other drugs used can be accounted for by the LSRL of % other drugs used on marijuana use %. (87.3% of marijuana drug use accounts for other drug use).

e. Do these results indicate that marijuana use leads to the use of harder drugs? Explain.

The use of other drugs is associated with marijuana use, but we cannot infer causation. There may be other variables (lurking/confounding).

2. Does a person's cholesterol level tend to change with age? Data collected from 1406 adults aged 45-62 produced the regression analysis shown. Assuming that the data satisfy the conditions for inference, examine the association between age and cholesterol level.

Dependent variable is: Chol
s = 46.16

Variable	Coefficient	SE(Coeff)	t-ratio	P-value
Intercept	194.232	~~13.55~~	~~14.3~~	~~≤0.0001~~
Age	0.771639	0.2574	3.00	0.0056

a. State the appropriate hypothesis for the slope.

State:

$H_0: \beta = 0$ there is no association between age and cholesterol level.

$H_A: \beta \neq 0$ there is an association between age and cholesterol level.

$\alpha = 0.05$

b. Test your hypothesis and state your conclusion in the proper context. Assume conditions are met.

Plan: t-test for slope of a regression line

Do:

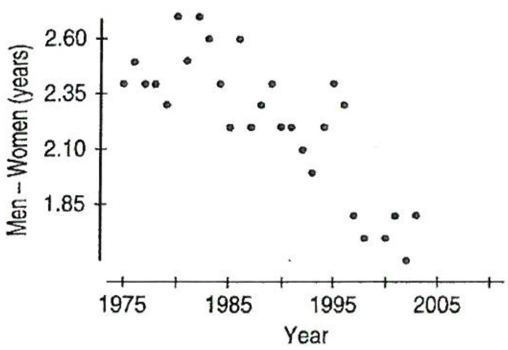
test statistic = 3.00

df = 1404

p-value = 0.0056

Conclude: Because our p-value = 0.0056 is less than our significance level $\alpha = 0.05$, we reject the null. There is convincing evidence that there is an association between age and cholesterol level, or that the slope β of the true regression line relating age and cholesterol level is not equal to 0.

3. The scatterplot suggests a decrease in the difference in ages at first marriage for men and women since 1975. We want to examine the regression to see if this decrease is significant.



Dependent variable is: Men – Women
R squared = 65.6%
s = 0.1869 with 28 – 2 = 26 degrees of freedom

Variable	Coefficient	SE(Coeff)	t-ratio	P-value
Intercept	61.8067	8.468	7.30	≤0.0001
Year	−0.02996	0.0043	−7.04	≤0.0001

a. Write appropriate hypotheses.

State:
$\alpha = 0.05$

$H_0: \beta = 0$ the difference in age between men and women at first marriage has not been decreasing since 1975.

$H_A: \beta < 0$ the difference in age between men and women at first marriage has been decreasing since 1975.

b. Test the hypothesis and state your conclusion about the trend in age at first marriage.

DO:

test statistic = −7.04
df = 26
p-value ≤ $\frac{0.0001}{2}$ (one-sided)

*Assume conditions are met.

Conclude: Because our p-value is less than our significance level $\alpha = 0.05$, we reject the null. There is convincing evidence that the difference in age between men and women at first marriage has been decreasing since 1975.

Plan: Because our conditions are met, we will perform a t-test for slope β.

c. Based on the analysis of marriage ages since 1975, give a 95% confidence interval for the rate at which the age gap is closing. Explain what your confidence interval means.

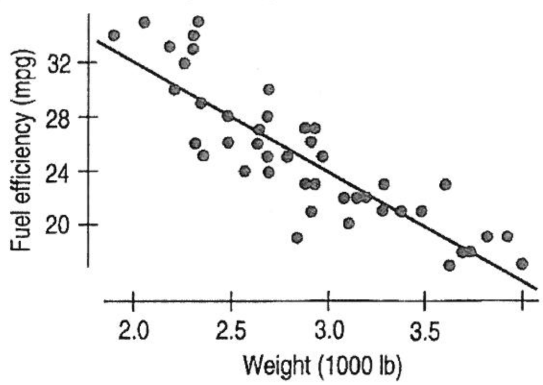
Plan: Because our conditions are met, we will use a t-interval for slope β.

Do: $-0.02996 \pm 2.056(0.0043) = (-0.03880, -0.02112)$

Conclude: We are 95% confident that the interval from -0.03880 to -0.02112 captures the slope β of the true regression line relating the difference in age between men and women at first marriage to year.

State: We want to estimate the slope β of the population regression line relating the difference in age between men and women at first marriage to year with 95% confidence.

4. A consumer organization has reported test data for 50 car models. We will examine the association between the weight of the car (in thousands of pounds) and the fuel efficiency (in miles per gallon).



Variable	Count	Mean	StdDev
MPG	50	25.0200	4.83394
wt/1000	50	2.88780	0.511656

Dependent variable is: MPG
R-squared = 75.6%
s = 2.413 with 50 − 2 = 48 df

Variable	Coefficient	SE(Coeff)	t-ratio	P-value
Intercept	48.7393	1.976	24.7	≤0.0001
Weight	−8.21362	0.6738	−12.2	≤0.0001

a. Is there strong evidence of an association between the weight of a car and its gas mileage? Write the appropriate hypothesis.

State:

$H_0: \beta = 0$ There is no (linear) relationship between the weight of a car and its gas mileage.

$H_A: \beta \neq 0$ There is a relationship.

$\alpha = 0.05$

b. Test your hypothesis and state your conclusion. * Assume conditions are met.

Plan: t-test for slope β of a regression line

DO: test statistic = -12.2
df = 48
p-value ≤ 0.0001

Conclude: Because our p-value is less than our significance level $\alpha = 0.05$, we reject the null. There is convincing evidence that the slope of the population regression line relating weight of a car to its gas mileage is not zero.

c. Create a 95% confidence interval for the slope of the regression line.

State: we want to estimate the slope β of the population regression line relating weight of a car to its gas mileage with 95% confidence.

Plan: Because our conditions are met, we will use a t-interval for slope β of a regression line.

DO: $-8.21362 \pm 2.0106(0.6738) = (-9.5684, -6.8589)$

Conclude: we are 95% confident that the interval from -9.5684 to -6.8589 captures the slope β of the population regression line relating weight of a car to its gas mileage.

d. Explain in this context what your confidence interval means.