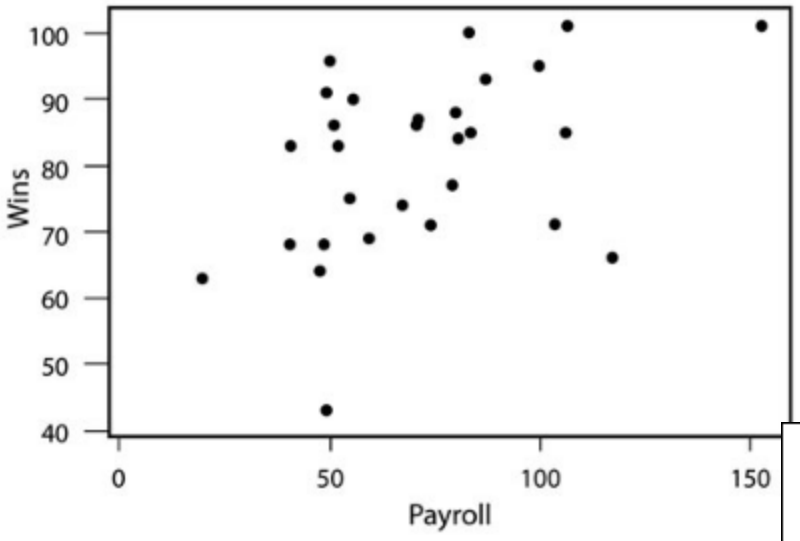
TEST Linear Regressions

Name _____

Answer all questions completely and with justifications.

1) Major League Baseball payrolls and winning percentage data were collected on the opening day payrolls (in millions of dollars) and the number of wins during the season for a random sample of Major League Baseball teams in the early 2000s. A scatterplot of the data is displayed at left, together with partial computer output from a linear regression calculation.



Predictor	Coef	SE Coef	T	P
Constant	66.901	6.261	10.68	0.000
Payroll	0.19803	0.08221	2.41	0.023

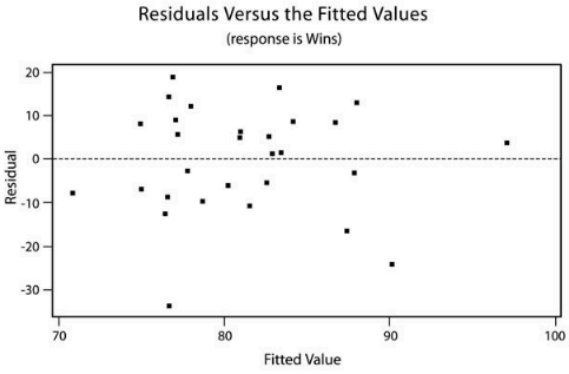
S = 12.38 R-Sq = 17.2% R-Sq (adj) = 14.2%

a) Determine the value of the correlation coefficient and interpret its meaning in the context of the problem.
r = 0.415 there is a moderate positive linear relationship.

b) Give the equation of the least squares regression line. Be sure to define any **variables** that you use in your equation.

$\hat{y} = 66.9 + 0.198x$ yy is the number wins in a season during the early 2000's and x is the opening day payroll in millions

c) A residual plot resulting from the linear regression is shown below. Estimate the payroll and the number of wins for the team with the most negative residual.



41 wins and a payroll of about $50 million dollars

d) Is there statistically significant evidence of a positive linear relationship between payroll and games won for Major League Baseball teams? Justify your answer.

$H_0\beta = 0$ versus $H_a:\beta > 0$, where β is the slope of the least squares regression line between payroll and wins for MLB. $\alpha = 0.05$

t test for β.
$t = 2.4$. With df = 28, the P value is 0.024.

Random: The data come from a random sample.
Normal:There is enough data to assure normality $n \geq 30$
Linear: The scatter plot is apparently linear .
Equal Scatter: The scatter plot shows approximately equal scatter
Independent: 30*10< the seasons in the population.

Because the P value of 0.024 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence of a positive linear relationship between payroll and wins for MLB

2) 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 tak**en 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 **constants** used in the equation.

$\hat{y} = 0.137 + 0.240x$ **0.137 amperes is the expected electrical output if wind velocity is 0**
Typically, there is an 0.24 increase in amperes/miles per hour of wind speed.
Typically, for every increase of 1 mph in wind speed, the electrical production increases .24 amperes

(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. 25→ .136+.24(25) 15→ -.136+.24(15)

10(0.24) = 2.4 amperes

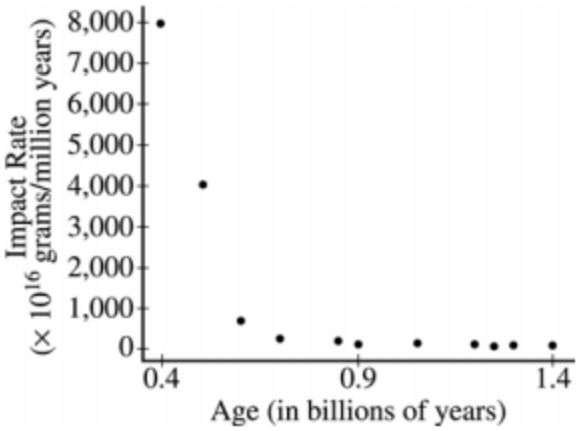
(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity? coefficient of determination

87.3% of variation in electricity production is explained by its linear relationship with wind velocity.

(d) Interpret the value of *s* in context of the problem.

s = 0.237 means that typically the amount of observed amperes is 0.237 away from the expected amount

3. The Earth's Moon has many impact craters that were created when the inner solar system was subjected to heavy bombardment of small celestial bodies. Scientists studied 11 impact craters on the Moon to determine whether there was any relationship between the age of the craters (based on radioactive dating of lunar rocks) and the impact rate (as deduced from the density of the craters). The data are displayed in the scatterplot below.



a) Describe the nature of the relationship between impact and age.

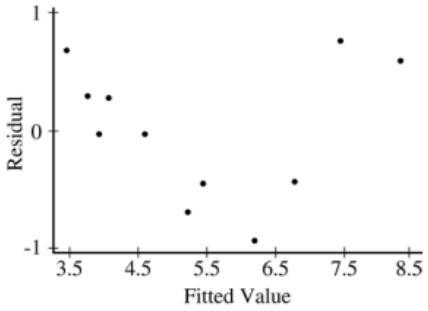
relationship between impact and age is strong, negative and non-linear.

Regression Equation: ln(rate) = 4.82 – 3.92 ln(age)

Predictor	Coef	SE Coef	T	P
Constant	4.8247	0.1931	24.98	0.000
ln(age)	–3.9232	0.4514	–8.69	0.000

S = 0.5977 R-Sq = 89.4% R-Sq (adj) = 88.2%

Prior to fitting a linear regression model, the researchers transformed both impact rate and age by using logarithms. The following computer output and residual plot were produced.



(b) Interpret the value of r^2.

89.4% of the variation in the rate of impact can be predicted by age.

by the linear regresssion

(c) Comment on the appropriateness of this linear regression for modeling the relationship between the ransformed variables.

It is a better model than the original, but there still is a pattern in the residual plot, indicating that there is a better equation to model the data

~~(d) since this model uses ln(y) and ln(x) is it a power model or an exponential one?~~

~~small lecture on regression~~