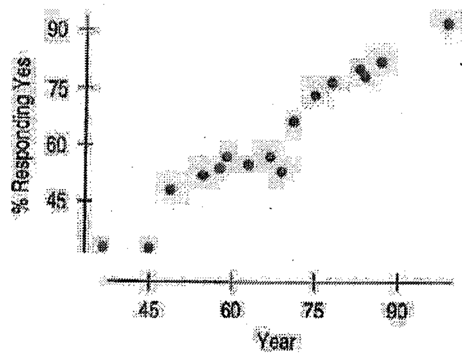
AP Stats Ch 27 - Required Practice **Name:** solutions **Per:** ____

1. Ms. President? In Chapter 7 we saw data collected by the Gallup organization. They have, over six decades, periodically asked the following question:

If your party nominated a generally well-qualified person for president who happened to be a woman, would you vote for that person?

Here is a scatterplot of the percentage answering "yes" vs. the year of the century (37 = 1937):



In Chapter 7 we could describe the relationship only in general terms. Now we can learn more. Here is the regression analysis:

Dependent variable is: Yes
R-squared = 94.2%
s = 4.274 with 16 − 2 = 14 degrees of freedom

Variable	Coefficient	SE(Coeff)	t-ratio	P-value
Intercept	−5.58269	4.582	−1.22	0.2432
Year	0.999373	0.0661	15.1	<0.0001

a) Explain in words and numbers what the regression says.
b) State the hypothesis about the slope (both numerically and in words) that describes how voters' thoughts have changed about voting for a woman.
c) Assuming that the assumptions for inference are satisfied, perform the hypothesis test and state your conclusion. Be sure to state it in terms of voters' opinions.
d) Explain what the R-squared in this regression means.

a) LSRL: $\hat{y} = -5.58269 + .999373x$ x: year
y: % saying yes

Main thing to explain is the slope:
For every added year the % of people who say they will vote for a woman president increases by about 1%, on average.

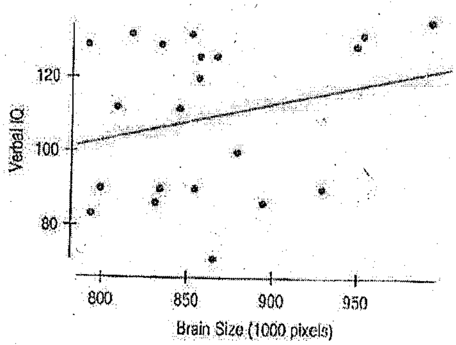
b) Ho: $\beta = 0$ There is no association between % saying yes and year.
Ha: $\beta > 0$ There is a positive association between % saying yes and year.

c) $t = 15.1$ w/ df = 14 : tcdf(15.1, 999, 14) = $2.3 \times 10^{-10} \approx 0$
LB, UB, df

with $\alpha = .05$, p-value ≈ 0 is low so we reject Ho.
We do have sufficient statistical evidence to conclude that there is a positive association between % saying yes and year.

d) $r^2 = .942$
About 94% of the variation in percentage of people saying yes to voting for a woman president is explained by the LSRL which relates % voting yes to year.

16. **Brain size.** Does your IQ depend on the size of your brain? A group of female college students took a test that measured their verbal IQs and also underwent an MRI scan to measure the size of their brains (in 1000s of pixels). The scatterplot and regression analysis are shown, and the assumptions for inference were satisfied.



Dependent variable is: IQ_Verbal
R-squared = 6.5%

Variable	Coefficient	SE(Coeff)
Intercept	24.1835	76.38
Size	0.098842	0.0884

$b = .098842$

$s_b = .0884$

a) Test an appropriate hypothesis about the association between brain size and IQ.

b) State your conclusion about the strength of this association.

a) $H_0: \beta = 0$ No linear association between brain size and verbal IQ.

$H_A: \beta \neq 0$ Is a linear association between brain size and verbal IQ.

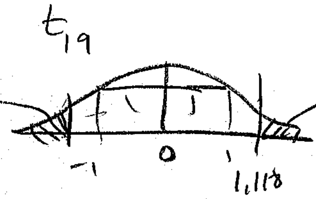
$$t = \frac{b - b_0}{s_b} = \frac{b - 0}{s_b} = \frac{.098842}{.0884} = 1.118$$

$n = 21$ (count the dots)

$df = n - 2 = 19$

2 sided

also .13875



tcdf(1.118, 999, 19) = .13875

p-value = 2(.13875) = $\boxed{.2775}$

With $\alpha = .05$, p-value = .2775 is high so we fail to reject H_0.
We do not have sufficient statistical evidence to conclude there is a linear association between brain size and verbal IQ.

b) $r^2 = .065$

Only 6.5% of the variation in verbal IQ can be explained by the LSRL which relates verbal IQ to brain size.

In this sample, slope is .0988 so there is a very weak positive association. But this is not strong enough to infer that there is any association at all in the population of all students.

6. **Used cars.** Classified ads in a newspaper offered several used Toyota Corollas for sale. Listed below are the ages of the cars and the advertised prices.

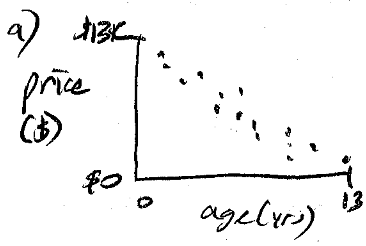
Age (yr)	Prices Advertised ($)
1	12,995; 10,950
2	10,495
3	10,995; 10,995
4	6,995; 7,990
5	8,700; 6,995
6	5,990; 4,995
9	3,200; 2,250; 3,995
11	2,900; 2,995
13	1,750

a) Make a scatterplot for these data.
b) Do you think a linear model is appropriate? Explain.
c) Find the equation of the regression line.
d) Check the residuals to see if the conditions for inference are met.

e) State hypotheses for determining if there is an association between age and price, and conduct a hypothesis test on the slope of the LSRL. Use the p-value to write a conclusion statement in context.
f) Now use the data to create a 95% confidence interval. Explain what the interval means in context.
g) Explain the meaning of '95% confidence' (the confidence level) in context.

$\hat{y} = a + bx$

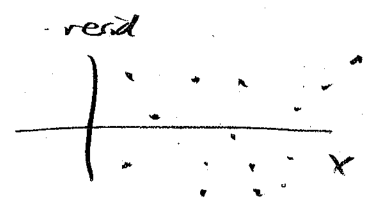
a)

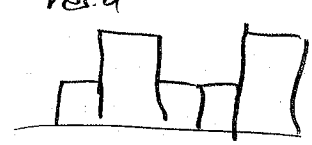


b) yes, scatterplot appears linear

c) $\widehat{price} = 12319.59 - 924(age)$

d) – straight enough? yes ✓
– indep?/fanning? no pattern in residuals ✓
– NN residuals? close enough ✓





e) Ho: $\beta = 0$ There is no association between price and age of car.
Ha: $\beta \neq 0$ There is a linear association between price and age of car.

LinRegTTest w/ $\beta, \rho \neq 0$ $\mu_y = \alpha + \beta x$

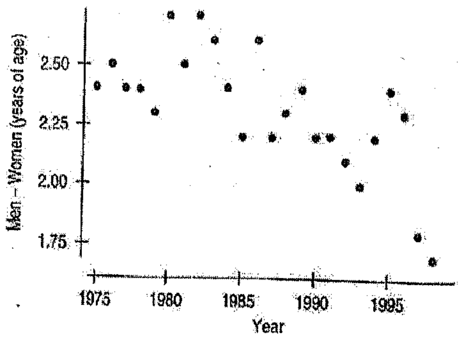
With $\alpha = .05$, $p = 1 \cdot 10^{-8}$ is low so we reject Ho. We do have sufficient statistical evidence to conclude there is a linear assoc. b/w price and age of car.

$\widehat{price} = 12319.59 - 924(age)$
$t = -11.2287$
$p = 1.066 \cdot 10^{-8}$
$df = 15$

f) LinRegTInt w/ C-level = .95 : $(-1099, -748.6)$
we are 95% confident that for all corollas, the price will decrease between $748.60 and $1099.00 for every additional year of age, on average.

g) If we were to repeat this by taking different samples of cars and finding confidence intervals for the slopes of the LSRLs 95% of those confidence intervals would capture the true slope, β, of the LSRL for all Toyota Corollas.

5. Marriage age. 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.



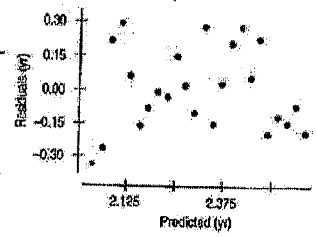
Dependent variable is: Men – Women
R-squared = 46.3%
s = 0.1866 with 24 – 2 = 22 degrees of freedom

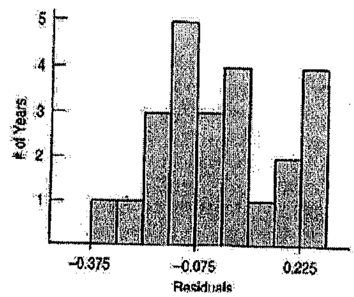
Variable	Coefficient	SE(Coeff)	t-ratio	P-value
Intercept	49.9021	10.93	4.56	0.0002
Year	–0.023957	0.0055	–4.35	0.0003 ← 2 sided

p-value 1 sided $= .00015$

a) Write appropriate hypotheses.

b) Here are the residuals plot and a histogram of the residuals. Do you think the conditions for inference are satisfied? Explain.





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

a) $H_0: \beta = 0$ (there is no change in marriage age diff.)

$H_A: \beta < 0$ (marriage age difference is decreasing)

b) straight enough: scatterplot ✓
indep: no resid pattern ✓
fanning? no ✓
Nearly Normal: hist, red good ✓

t test$(-999, -4.35, 22)$

c) $\widehat{\text{men-women}} = 49.9 - .024(\text{yr})$, $t = -4.35$, $p = .00015$, reject H_0.
there is evidence of a decreasing difference in marriage age.

7. Marriage age, again. Based on the analysis of marriage ages since 1975 given in Exercise 5, give a 95% confidence interval for the rate at which the age gap is closing. Clearly explain what your confidence interval means.

$b = -.024$ $\quad t = \dfrac{b - \beta_0}{s_b} = \dfrac{b}{s_b}$ so $s_b = \dfrac{b}{t} = \dfrac{-.024}{-4.35} = .0055$

for 95% df = 22 table says $t^* = 2.074$

$CI = -.024 \pm 2.074(.0055)$

$(-.035, -.013)$

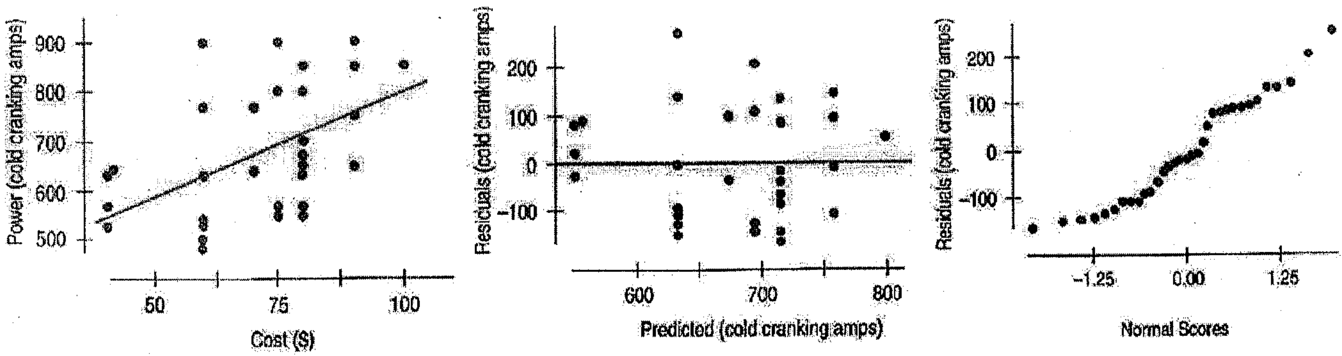
We are 95% confident that mean difference in marriage age (men – women) is between $-.035$ and $-.013$ years in age per calendar year.

25. Start the car! In October 2002, *Consumer Reports* listed the price (in dollars) and power (in cold cranking amps) of auto batteries. We want to know if more expensive batteries are generally better in terms of starting power. Here are several software displays.

Dependent variable is: Power
R-squared = 25.2%
s = 116.0 with 33 − 2 = 31 degrees of freedom

Variable	Coefficient	SE(Coeff)	t-ratio	P-value
Intercept	384.594	93.55	4.11	0.0003
Cost	4.14649	1.282	3.23	0.0029

a) How many batteries were tested? $\boxed{33}$
b) Are the conditions for inference satisfied? Explain.
c) Is there evidence of an association between the cost and cranking power of auto batteries? Test an appropriate hypothesis and state your conclusion.
d) Is the association strong? Explain.
e) What is the equation of the regression line?
f) Create a 90% confidence interval for the slope of the true line.
g) Interpret your interval in this context.



b) straight enough ✓ indep: no resid pattern ✓ fanning? no ✓ Nearly Normal res? NPP straight ✓

c) $H_0: \beta = 0$ (no linear relationship between cost & power)
$H_A: \beta \neq 0$ (linear relationship between cost & power)

$t = 3.23$, p-value $= .0029$ (2 sided) $= .00145$
with low p-value, reject H_0. There is strong evidence of a linear relationship between cost and power.

d) $r^2 = .252$ so only 25% of the variation in power is explained by cost. Even though there is a definite statistically significant slope, the predictive ability of the equation is low. Also $s = 116$ which is $\frac{116}{700}$ 17% of a typical y value. Not of much use.

e) $\widehat{power} = 384 + 4.146(\text{cost})$

f) $S_b = 1.282$ for df = 31, 90% by table $t^* = 1.697$ (df = 30)
CI = statistic ± (critval)(SE)
$= 4.146 \pm 1.697(1.282)$
$= (1.97, 6.32)$

g) We are 95% confident that mean power increases by between 1.97 and 6.32 amps for every additional $ in cost.

Chapter 27 Practice Quiz

AP Statistics Quiz A – Chapter 27 ***Name***_______________________

A college admission s counselor was interested in finding out how well high school grade point averages (HS GPA) predict first-year college GPAs (FY GPA). A random sample of data from first-year students was reviewed to obtain high school and first-year college GPAs. The data are shown below:

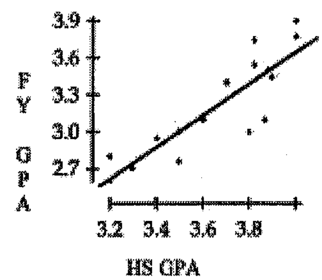
HS GPA	3.82	3.90	3.20	3.40	3.88	3.50	3.60	3.70
FY GPA	3.75	3.45	2.60	2.95	3.50	2.76	3.10	3.40

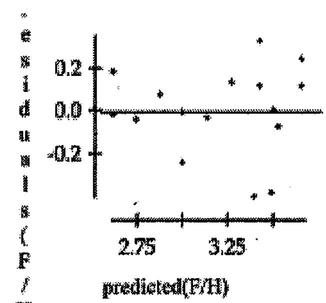
HS GPA	4.00	3.30	3.50	3.80	3.87	4.00	3.20	3.82
FY GPA	3.90	2.70	3.00	3.00	3.10	3.77	2.80	3.54

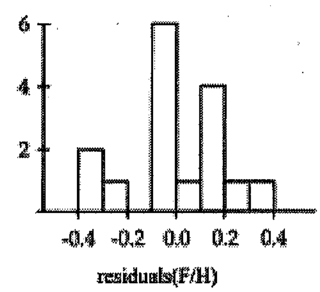
Dependent variable is: FY GPA
No Selector
R squared = 75.4% R squared (adjusted) = 73.6%
s = 0.2118 with 16 - 2 = 14 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	1.92283	1	1.92283	42.9
Residual	0.627867	14	0.044848	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-1.56410	0.7306	-2.14	0.0504
HS GPA	1.30527	0.1993	6.55	≤ 0.0001







1. Is there evidence of an association between high school and first-year college GPAs? Test an appropriate hypothesis and state your conclusion in the proper context.

$H_0: \beta = 0$ There is no association between HS and FY GPA.
$H_A: \beta \neq 0$ There is an association between HS and FY GPA.

conditions

✓ – straight enough? (scatterplot ok)
✓ – no pattern/fanning in residuals? (resid. plot ok)
✓ – residuals Nearly Normal? (resids histogram OK)

$\hat{y} = -1.564 + 1.305x$
x: HS GPA
y: FY college GPA
$r^2 = .75$

$t = 6.55$, $df = 14$
From software, p-value ≤ .0001

With $\alpha = .05$, p-value < .0001 is low, so we reject H_0.
We do have sufficient statistical evidence to conclude that there is an association between High School and First Year college GPA.

Not req'd by this question, but good to know:

$b = 1.305$
For every 1 additional GPA point in HS GPA, college FY GPA increases by 1.305 points, on average.

$r^2 = .75$
About 75% of the variation in college FY GPA is explained by the LSRL model which relates FY GPA to HS GPA.

2. **Create and interpret a 95% confidence interval for the slope of the regression line.**

no data, so we must do by hand:

CI: $b \pm t^* s_b$

From software: $b = 1.30527$

$s_b = 0.1993$

$1.30527 \pm (2.145)(0.1993)$

$(.8778, 1.7328)$

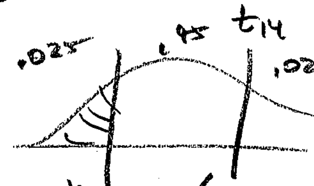
$n = 16$

$df = n - 2 = 14$

always include a text explanation of any confidence interval:

we are 95% confident that for every one additional GPA point in HSGPA, college FY GPA increases by between 0.88 and 1.73 points, on average.

t^* for 95% confidence on $df = 14$



$t^* = \text{invT}(.025, 14) = \pm 2.145$