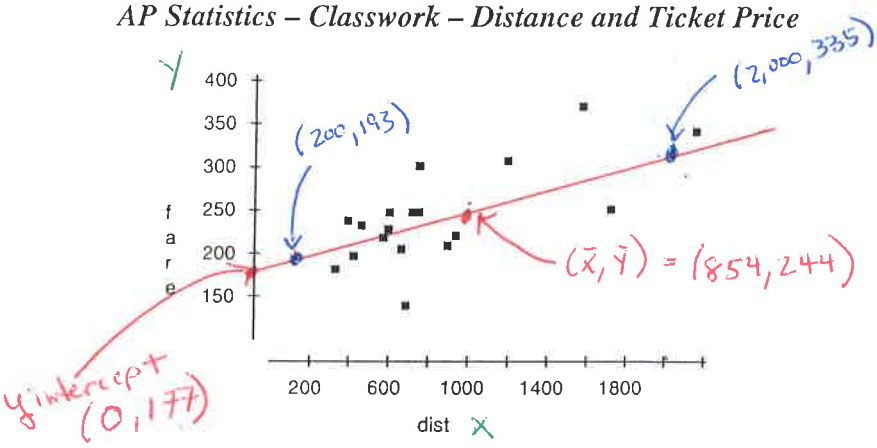
KNOW → GREEN SHEET $\hat{y} = b_0 + b_1 x$ $b_0 = \bar{y} - b_1 \bar{x}$ $b_1 = r \cdot \frac{S_y}{S_x}$

ACTIVITY 3.2B: Part 1 - Understanding AP Green Sheet Regression Formulas

Name: KEY

INSTRUCTIONS:
- Do NOT enter the data in calculator.
- Use the supplied information and AP Green Sheet formulas only.

AP Statistics – Classwork – Distance and Ticket Price



Atlanta to:	Distance (x)	Fare (y)
Baltimore	568	219
Boston	933	222
Dallas	720	249
Denver	1190	308
Detroit	602	249
Kansas City	683	141
Las Vegas	1719	252
Miami	589	229
Memphis	327	183
Minneapolis	894	209
New Orleans	419	199
NY	749	248
Okla City	749	301
Orlando	392	238
Philadelphia	657	205
St Louis	461	232
Salt Lake	1565	371
Seattle	2150	343
Summary Statistics		
Mean	$\bar{x}$ 853.7	244.33 $\bar{y}$
St Dev	S_x 497.8	56.37 S_y
Correlation	0.694	r

Plot $(\bar{x}, \bar{y})$ → (853.7, 244.33)

1. Find r^2. $r = (.694)^2$ $r^2 = .4816$
2. Explain what r^2 means in this context.

About 48.2% of the variability in air fare is accounted by the linear regression model explained by distance traveled.

3. Find the slope of the regression line.

$b_1 = r \cdot \frac{S_y}{S_x} = (.694) \cdot \frac{56.37}{497.8}$ $b_1 = .079$

4. Find the y-intercept of the regression line. Plot the y-intercept on the scatter plot.

$b_0 = \bar{y} - b_1 \bar{x} = 244.33 - .079(853.7)$ $b_0 = 176.89$

5. Write the equation of the linear model.

$\widehat{FARE} = 176.89 + 0.079(DISTANCE)$ OR $\hat{y} = 176.89 + .079x$

x = distance (miles)
y = fare ($)

TO HELP UNDERSTAND CONTEXT, USE WORDS →

6. Estimate the fare for a 200-mile flight. Plot the point on the scatter plot.

$\widehat{FARE} = 176.89 + .079(200)$ $\widehat{FARE} = \$192.69$

YOU MUST SHOW WORK FOR FULL CREDIT

Estimate the fare for a 2000-mile flight. Plot the point on the scatter plot.

$\widehat{FARE} = 176.89 + .079(2000)$ $\widehat{FARE} = \$334.89$

Part 1 (continued)

8. Using those estimates, draw the line on the scatterplot.
9. Explain what the *y*-intercept means in this context.

y intercept = a = b_0 = $176.99

The model predicted a base air fare of about $177 which might represent the fixed costs of air travel (employees, Maintenance, etc)

10. Explain what the slope means in this context.

Slope = b = b_1 = .079

For each 100 miles of distance traveled, we expect the air fare to increase about $7.90. Could say each additional mile for about 8¢ but 100 miles makes more sense.

11. The fare to fly to Los Angeles, 1719 miles from Atlanta, is $212. Find the residual.

X = 1719 miles
Y = $212
$\hat{Y}$ = $\widehat{\text{FARE}}$ = 176.89 + .079(1719) = $312.69

Residual = $Y - \hat{Y}$
= 212 − 312.69
Residual is −$100.69

12. In general, a positive residual means… $Y < \hat{Y}$

The model's predicted value was lower than the actual air fare. The model under estimates fares (IE AIR FRANCE)

13. In general, a negative residual means… $\hat{Y} > Y$

The model's predictions were higher than the actual air fare. The model over estimates fares (IE JET BLUE)

s = standard deviation of the residuals.
It measures the average size of the prediction errors (residuals) when using a regression line.

Part 2 (Understanding Computer Printouts) DO NOT LOOK AT PRIOR PAGES

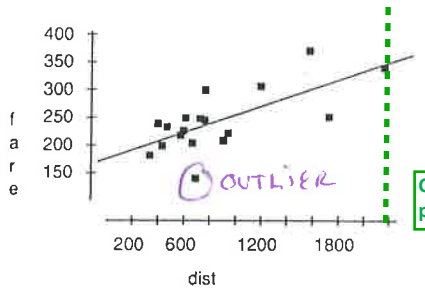
Here are computer outputs for 2 models developed from the data provided on the front page for ticket fares and distance traveled from Atlanta.

SEE NEXT PAGE FOR SSE + SST

Dependent variable is: fare ← Response Variable
No Selector
R squared = 48.2% R squared (adjusted) = 45.0%
s = 41.82 with 18 - 2 = 16 degrees of freedom
(18 = sample size)

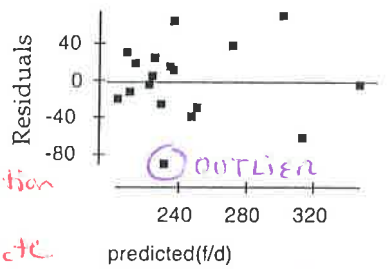
Source	Sum of Squares	df	Mean Square	F-ratio
Regression	26037.4	1	26037.4	14.9
Residual	27980.6	16	1748.79	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	177.215 (y int)	19.99	8.86	≤ 0.0001
dist	0.078619 (slope)	0.0204	3.86	0.0014



Can you use this model to predict a fare for 2,500 miles?

NO. Extrapolation beyond the data we used for are model is not reliable and should never be done.



1. Is the linear model appropriate for estimating airfare from distance flown? Why? YES
   (1) The scatter plot shows a linear association
   (2) The residual plot shows no pattern.
   Therefore, the linear model is appropriate

2. How strong is the model? Explain. for predicting air fare from distance traveled
   r^2 48.2% of the variability is accounted for by the model predicting fare based on distance traveled.

3. Identify outliers. Why are they unusual?
   $150 for a flight of about 700 miles seems low compared to other fares

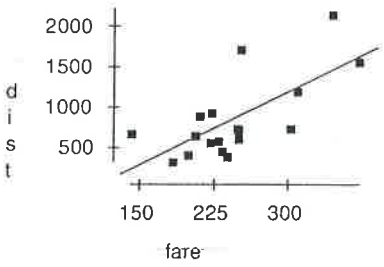
4. Write the equation of the model → FARE = 177.215 + .0786 (DISTANCE)
   It is basically the same as the LSRL found by hand but more accurate since it isn't based on rounded values.

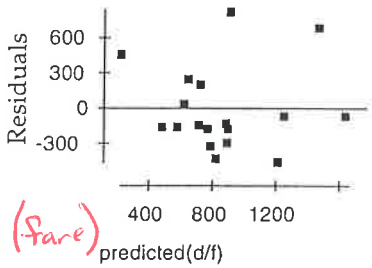
5. Predict the airfare for a 1000-mile flight
   FARE = 177.215 + .0786 (1000) → FARE = $255.83

Dependent variable is: dist ← Response Variable
Switched explanatory + response Var.
No Selector
R squared = 48.2% R squared (adjusted) = 45.0%
s = 369.3 with 18 - 2 = 16 degrees of freedom
NOTICE r^2 are the same

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	2030495	1	2030495	14.9
Residual	2182031	16	136377	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-644.287	397.9	-1.62	0.1249
fare	6.13101	1.589	3.86	0.0014





6. Write the equation of the model to estimate how far you could fly for a given price.
   distance = -644.287 + 6.13 (fare)

7. How far does this model suggest you could fly for the fare you estimated in #5?
   $255.83 FARE → distance = -644.287 + 6.13 ($255.83)
   The predicted distance is about 924 miles

DO YOU NEED TO KNOW SSE + SST? NO

SKIP EXAMPLE ON PAGE 179-181
"PACK WT AND BODY WT"

What it is saying

	SUM SQUARES
Regression	26037.4
Residual	27980.6 ← SSE
	54018 ← SST

$r^2 = 1 - \frac{SSE}{SST} = 1 - .518 = .482$

$\frac{SSE}{SST}$ → % ERROR (NOISE)