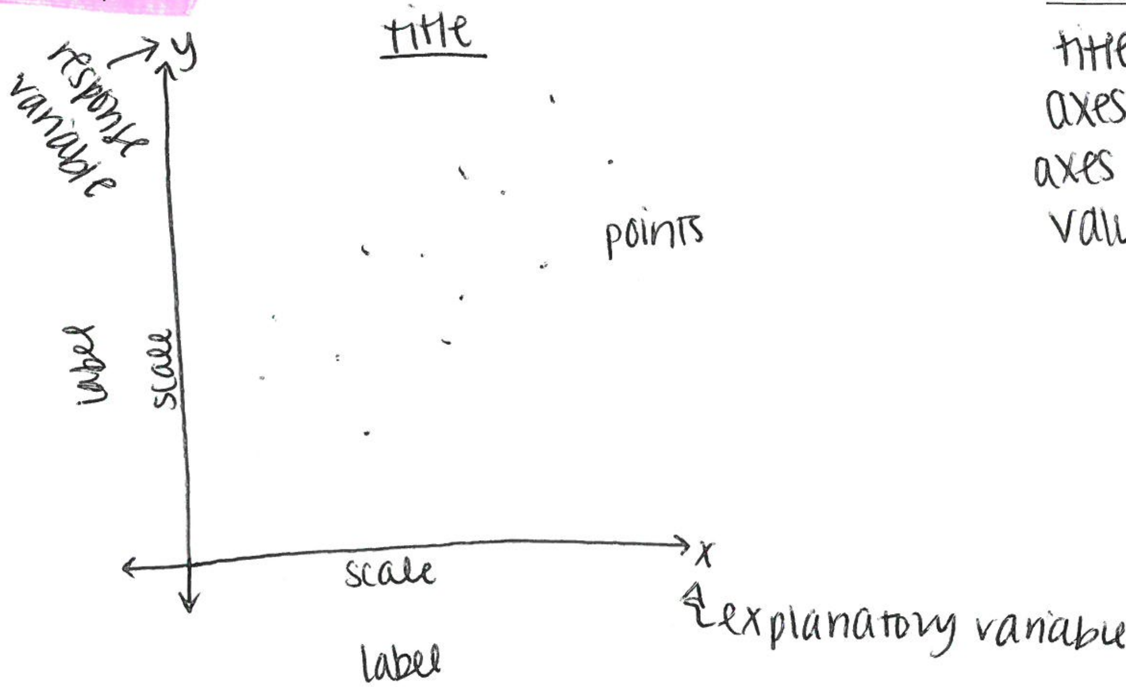
AP Statistics
Unit 02 REDO – Day 01 Notes

Name Key
Period

Scatterplots:



Include:

title
axes labels
axes scales
values (data points)

Describe using:

Form: clusters, gaps, outliers, linear or curved.

Direction: positive slope vs. negative slope

Strength: strong, moderate, weak (how tight the points are in a line/curve)

Correlation: measures the strength of the linear relationship between x and y.

Correlation Coefficient (r): same as correlation, but given as a numerical value. $-1 \le r \le +1$. as $|r|$ approaches 1, the correlation increases (becomes stronger).

Coefficient of Determination (R^2): sentence frame: the % variation in y explained by x (or the LSRL of y on x).

LSRL: least squares regression line.

the line that gets closest to most data points.

it best fits the data.

$\hat{y} = a + bx$ ← explanatory variable

predicted response variable → $\hat{y}$

a ↑ y-intercept

b ↑ slope

Three ways to find the LSRL equation:

METHOD #1: Using a list of data points:	METHOD #2: Using calculated values for mean, standard deviation, and r.	METHOD #3: Using a MiniTab Output:
1. Put data in L1 and L2 of your calculator 2. STAT > CALC > 8. LinReg(a+bx) 3. Place a and b values into your equation and be sure to write your equation with Y-HAT and IN CONTEXT.	1. Use the equations on the equation sheet to calculate a and b. 2. Place a and b values into your equation and be sure to write your equation with Y-HAT and IN CONTEXT.	1. Identify a and b. The value of a can be found in the Constant row, Coef column. The value of b can be found in the Variable row, Coef column. 2. Place a and b values into your equation and be sure to write your equation with Y-HAT and IN CONTEXT.

METHOD #1 EXAMPLE: Create the LSRL for predicting number of bags of dog food used per month based on the number of dogs a person owns. Use the data table below.

# of dogs	Bags of dog food used/month
1	3
2	6
3	8
2	7
3	10
4	11
5	15
5	15
4	12
3	9
6	17
9	25
8	24
7	20

$\widehat{\text{\# of bags of dog food}} = 0.636 + 2.792(\text{\# of dogs})$

OR

$\hat{y} = 0.636 + 2.792(x)$

$\hat{y}$ = predicted # of bags of dog food used per month

x = # of dogs owned

(if you define y instead of $\hat{y}$, omit "predicted".

b_0 on equation sheet

b_1 on equation sheet

METHOD #2 EXAMPLE: Create the LSRL for predicting number of bags of dog food used per month based on the number of dogs a person owns. Use the values below.

$\bar{x} = 4.429$
$s_x = 2.377$
$\bar{y} = 13$
$s_y = 6.668$
$r = 0.9951$

$$b = r\frac{s_y}{s_x} = 0.9951 \cdot \frac{6.668}{2.377} = 2.7915$$

$$a = \bar{y} - b\bar{x}$$

$$a = 13 - 2.7915(4.429) = 0.6366$$

$$\widehat{\text{\# of bags of dog food}} = 0.63666 + 2.7915(\text{\# of dogs})$$

or... define x and y or $\hat{y}$ and write the equation of the LSRL w/ variables (x and $\hat{y}$).

METHOD #3 EXAMPLE: Create the LSRL for predicting number of bags of dog food used per month based on the number of dogs a person owns. Use the MiniTab Output below.

this means y-int.

Predictor	Coef	SE Coef	T	P
Constant	0.6362	whatevs	whatevs	whatevs
Bags	2.7918	whatevs	whatevs	whatevs
S = whatevs	R-Sq = 0.9902		~~R-Sq(adj)~~ = whatevs	

a

b

this is the variable x so slope goes next to it

to find r, take $\sqrt{R^2}$ and use the same sign as slope (±)

we don't use this

$$\widehat{\text{\# of bags of dog food}} = 0.6362 + 2.7918(\text{\# of dogs})$$

or...

Other important things:

Calculating & Interpreting Residuals:

residual = observed value – predicted value

EXAMPLE: Calculate and interpret the residual value for a person who owns 7 dogs.

Find the actual value (from table): (7, 20)

$y = 20$ bags of dog food

7 = # of dogs, 20 = # of bags of dog food

Find the predicted value (from LSRL): (7, 20.179)

$\hat{y} = 0.636 + 2.792(7) = 20.179$

20.179 = predicted # of bags of dog food

residual = $y - \hat{y}$

$= 20 - 20.179 = -0.179$

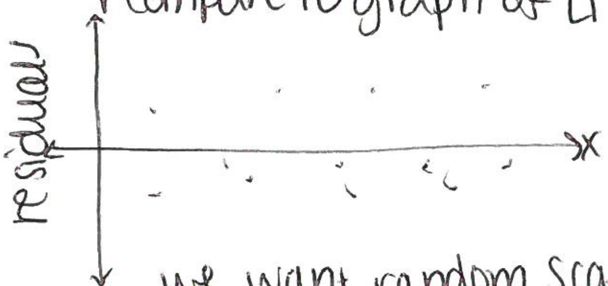
A person who owns 7 dogs is predicted to need 0.179 more bags of dog food per month than they actually need.

OR A person who owns 7 dogs needs 0.179 bags of dog food per month less than they are predicted to need.

Residual Plot:

usually x vs. residuals (for each point)

* compare to graph of L1 vs. L2



* do linreg on calc to plot residuals!

we want random scatter if the LSRL is a good fit.

Influential Point:

if removed, the LSRL drastically changes. could be an outlier (x-direction) or not!

Extrapolation:

when we use the model to predict y for values of x not in the domain used to create the model. BAD!

EXAMPLE: Would it make sense to use our model to predict the number of bags of dog food needed for a person who has 49 dogs? Why or why not?

Nope! this is extrapolation. 49 is not in the domain of x values used to create this model.

Interpreting slope and y-intercept:

Slope can be interpreted as follows: for every [one unit increase] in [x variable in context], we predict [y in context] will increase/decrease by [slope value in context w/ units].

y-int: If [x in context] is 0 (or there are none, in context), the value of [y in context] is predicted to be [y-int value w/ units in context].

EXAMPLE:

Slope: For every one additional dog owned by a person, the number of bags of dog food they need per month is predicted to increase by 2.792 bags.

y-intercept: A person who owns no dogs is expected/predicted to need 0.6366 bags of dog food per month.

Example: describe the scatterplot.

F
D
S

there is a moderately strong, positive, linear relationship between # of dogs owned by a person and # of bags of dog food needed per month.