

unit 9 - statistics - study guide



Box Plots



Using the TI-Nspire

Creating the box plot

1. Find One-Variable Statistics
2. HOME → 1 → 4
Label your data, skip = line, input all data
Remember to press enter after last data value!
3. CTRL → doc → 5 Press TAB to add your variable to the x-axis
4. MENU → 1 → 2

Finding One-Variable Statistics



Using the TI-Nspire

1. HOME → 1 → 4
Label your data, skip = line, input all data
Remember to press enter after last data value!
2. CTRL → doc → 1
3. MENU → 6 → 1 → 1
Add your data to X1 List then press OK

$$\bar{x} = \text{mean}$$

sx = sample standard deviation

σx = population standard deviation

MinX = minimum

Q_1X = quartile 1

MedianX = median

Q_3X = quartile 3

MaxX = maximum

Variation in a Data Set

- Use One-Variable Statistics

→ Standard Deviation

- Tells us how far a data point is away from the mean
- The farther apart the data, the bigger the SD
- ✎ Use σx when the data is **NOT** a sample
- ✎ Use sx when the data **IS** a sample

→ Interquartile Range (IQR)

- (quartile 3) - (quartile 1)
- Represents 50% of the data



Types of Data

One Variable Data

- Data with 1 variable
- Can be modeled using a bar graph, histogram, box plot, circle graph or dot plot

Two Variable (Bivariate) Data

- Data with 2 variables
- Typically modeled using a scatterplot
- Use line of best fit
- Casual relationship - Does one CAUSE the other to happen?

Measures of Central Tendency

- Use One-Variable Statistics
- A single number used to describe a set of data as a whole
- Most common: MEAN and MEDIAN

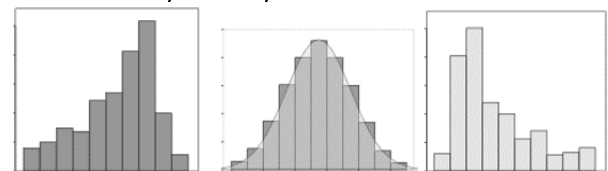
Mean - average

Median - middle value

Mode - value that appears the most

Interpreting Data

- **Skewed/Symmetry**



Left skewed/
Negative skewed No skew
Symmetrical Right skewed/
Positive skewed

- **Outliers** - a value that "lies outside" most of the other values in a set of data

Regressions/Line of Best Fit

Regression – a line that best represents the data on a scatter plot (linear, quadratic, exponential)

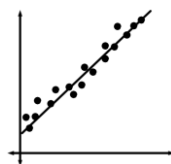


Using the TI-Nspire

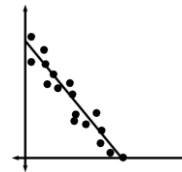
- TYPE YOUR DATA INTO LISTS**
HOME → 1 → 4
Label your data, skip = line, input all data
Remember to press enter after last data value!
- CREATE A SCATTERPLOT**
CTRL → doc → 5
TAB to “CLICK TO ADD VARIABLE” ON EACH AXIS and choose the appropriate variable for each axis. Your data will move into a scatter plot.
- SHOW LINE OF BEST FIT**
On the graph page:
menu → 4 → 6 → pick either 1, 4 or 8
- FIND A REGRESSION LINE/CURVE (AND CORRELATION COEFFICIENT)**
CTRL → doc → 1 then...
menu → 6 → 1 → pick either 3, 6 or A
add variables to x list & y list then press OK
“r” is the correlation coefficient
*substitute values into formula to get equation

Correlation Coefficient: r

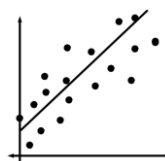
- Tells us how well the function matches the data
- ALWAYS between -1 and 1
- Use your calculator to find r



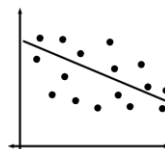
strong positive: good fit
 $r = 0.93$



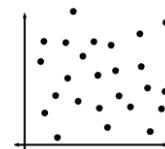
strong negative: good fit
 $r = -0.9$



weak positive
bad fit
 $r = 0.6$



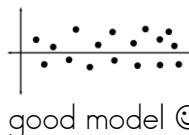
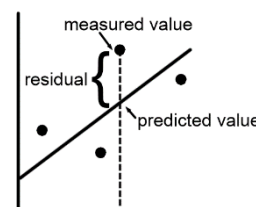
weak negative
bad fit
 $r = 0.5$



no correlation
terrible fit
 $r = 0$

Residuals

- MEASURED - PREDICTED = RESIDUAL
- GOOD: evenly spaced above/below the x-axis
- BAD: see a pattern



good model ☺



bad model ☹

Steps:

- Find the regression
- Find observed values
- Subtract: measured - predicted to get the residual value



Using the TI-Nspire

- Complete steps 1-4 of a regression
- MENU → 4 → 7 → 2 to view a residual plot
- Go back to your lists/spreadsheets page
- In the = spot in the “c” column type in $f1(x)$ to view predicted/observed y-values side by side

Two-Way Frequency Tables

EXAMPLE:

Gender	Favorite Color			TOTAL
	Blue	Red	Green	
Male	2	5	1	8
Female	3	2	8	13
TOTAL	5	7	9	21

- Always find totals if they don't give them to you
 - Be careful of how the question is worded
- What percent of females have a favorite color of red?
 $\frac{2}{13} = 0.15 = 15\%$
 - What percent of people who have a favorite color of red are females?
 $\frac{2}{7} = 0.29 = 29\%$